"It is also obvious that the individuals of the same species, though now inhabiting distant and isolated regions, must have proceeded from one spot, where their parents were first produced."

– Darwin 1859, On the Origin of Species

"Quod nullus proiceat lastadien in deep hujus civitatis, sub pena vite"

"That no one throws ballast in the deep of this town, by loss of his life"

– Wismar (Germany) statute, 1345

"Picture the mouth of the muddy, narrow River Tyne, jammed with four or five hundred keels and two or three hundred ships…"

– John Nef, 1558

Evolution in the North Atlantic: processes shaping spatial patterns of genetic diversity in introduced intertidal invertebrates

by

Anthony Leon Einfeldt

B.Sc. UBC 2008

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

Doctor of Philosophy

in the Graduate Academic Unit of Biology

Supervisor:	Jason Addison, PhD, Biology, UNB
Examining Board:	Gary Saunders, PhD, Biology, UNB Adrian Reyes-Prieto, PhD, Biology, UNB Tony Diamond, PhD, Forestry, UNB
External Examiner:	Sean Rogers, PhD, University of Calgary

This dissertation is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

March, 2018

©Anthony Einfeldt, 2018

ABSTRACT

The movement of individuals and their genes across geographic space influences a species' ecology and evolution, but it is often not possible to observe past or present movement directly. Molecular tools provide a means of inferring past movement and contemporary barriers to movement, because the known modes of mutation and inheritance underlying genetic variation provide clear predictions for patterns arising from movement and subdivision. In this thesis, I investigate how contemporary and historic patterns of movement shape evolutionary trajectories by investigating distributions of genetic variation in focal intertidal invertebrates of the North Atlantic. To determine how movement is affected by the interaction of currents with life-history traits, I sequenced mitochondrial and nuclear DNA of the intertidal amphipod *Corophium volutator* from discrete patches of mudflat habitat throughout the Northwest Atlantic. I detected patterns of genetic subdivision and gene flow concordant with hydrological patterns, demonstrating that currents shape evolution by determining dispersal pathways and cause fine-scale subdivision in marine communities. To test how *C. volutator* colonized the Northwest Atlantic coast, I investigated spatial genetic variation in populations from across its entire range using the same markers. I found that diversity in Northwest Atlantic populations was subsampled from more genetically diverse populations in the Northeast Atlantic, consistent with historic human-mediated introduction from the Northeast to the Northwest. To investigate how human-mediated dispersal affects species' evolutionary trajectories, I characterized genomic variation in *C. volutator* and a co-occurring annelid *Hediste diversicolor* in populations from the Northeast and Northwest Atlantic coasts. I found extensive genetic divergence between

the introduced and native ranges and genetic patterns consistent with historic admixture between populations within each range, providing evidence that human-mediated movement can create new allopatric lineages and erase ancestral genetic structure by promoting gene flow between otherwise isolated populations. Together, my results suggest that the increasing reach, magnitude, and frequency of global human movement will change the evolutionary trajectories of species associated with human vectors of transport. While contemporary connectivity will continue to be affected by regional processes (such as currents), uncurbed human activity will likely disrupt diversification arising from barriers at regional scales while promoting the formation of new lineages at a global scale.

DEDICATION

I dedicate this thesis to my parents, Sandra and Flemming Einfeldt, who instilled in me a deep appreciation for the natural world and a desire to understand the forces behind it. I have many fond memories of how they fostered my curiosity, creativity, and confidence. Amongst the first of these memories are the weekly walks they would take me on through the woods to find a rare species of plant, which when in bloom would produce individually wrapped chocolates. The world turns out to be full of things much like the Chocolate Tree, and I thank my parents for introducing me to the thrill of discovering their true nature.

ACKNOWLEDGEMENTS

This research grew from an initial project conceived by Jason Addison to assess connectivity amongst mudflats in the Bay of Fundy, funded by NSERC through a Strategic Project Grant to Myriam Barbeau and a Discovery Grant to Jason Addison. Additional funding for the projects in this thesis was provided by a Canadian Foundation for Innovation grant to Jason Addison, a New Brunswick Innovation Foundation grant to Jason Addison and scholarship to myself, a New Brunswick Wildlife Trust Fund grant to Jason Addison and myself, and fellowships from the Fredrik and Catherine Eaton Foundation, ACENET, and the Marguerite and Murray Vaughan Foundation awarded to myself. Many thanks to my supervisor Jason Addison and committee member Linley Jesson for their guidance and support on projects in and beyond this thesis. Thanks to my other committee member Myriam Barbeau, who gave helpful feedback on my proposals and thesis. Thanks also to Mark Forbes for direction and funding at a key point.

The broad geographic scope of this research was made possible with the help of many people. Sampling trips in North America were aided by Jeremy Doucet and Felix Zhou during their B.Sc. Honours in Jason Addison's lab, Brian Nason, Jarrett Starkey-Seto, and my parents Sandra and Flemming Einfeldt. I thank the researchers who collected and sent me *Corophium volutator* specimens, including Myriam Barbeau, David Drolet, Guy Bachelet, Anne-Laure Barillé, Jérôme Jourde, Chloé Dancié, Julia Sigwart, Rob Hughes, Martin Solan, Annelies Debacker, Sander Wijnhoven, Ines Heisterkamp, Birgit Hussel, K. Thomas Jensen, Jacob Strand, and Andreas Bick. Thanks to Jim Provan for hosting me at Queen's University Belfast, and to Julia Sigwart, Christine Maggs, Siobhan Rose Vye, Thomas Prebble, Nessa O'Connor, Lauren Sumner-

Rooney, Julia Calderwood, and many others for facilitating both scientific and social endeavours in Northern Ireland and Ireland.

Laboratory work was helped immensely by Jin-Hong Kim Jana Gruettner, who first implemented and troubleshot the genome reduction methods used in Chapter 4. Sara Edwards was a reliable sounding board (and rubber duck) for the many coding, analysis, and writing problems that I encountered. Stephen Heard and his Nerds (particularly Chandra Moffat, Julia Mlynarek, Allyson Heustis, and Mallory MacDonnell, Ken Dearborn, and Rylee Isett) were really good at Boggle, and along with Dan Quiring and Rob Johns provided stimulating weekly presentations, discussions, constructive comments, and collaboration. The mudflat ecology group (particularly David Drolet, Diana Hamilton, Kelan Kennedy, and Trevor Bringloe) provided many discussions.

This thesis was strengthened by my involvement in projects and systems not directly mentioned elsewhere in this document, which was made possible by the strong community in the UNB Biology Department. Thanks to Gary Saunders and his lab (particularly Amanda Savoie, Lesleigh Kraft, and Meghann Bruce) for ideas and discussions throughout my degree, and for spearheading a 2014 arctic sampling expedition. Thanks to Tony Diamond and his lab (particularly Kevin Kelly, Sarah Hudson, Kirsten Bowser, and Lauren Scopel) for the annual migrations to and from Machias Seal Island and the opportunities they presented. Helen Tai generously provided assistance and resources for RNA work on a project in progress that stems from the research presented here. By no means least, thank you to the graduate students at UNB and friends in Fredericton, who gave me the strength and motivation to continue this research and kept me in touch with what is truly important.

Table of Contents

ABSTRACT ii
DEDICATION iv
ACKNOWLEDGEMENTS v
Table of Contents vii
List of Tables xi
List of Figures xiii
Chapter 1 – General introduction 1
Inferring movement from distributions of variation 1
Inferring movement from molecular variation 3
Movement and evolution in North Atlantic intertidal invertebrates 6
References 10
Chapter 2 – Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the Northwest Atlantic[a] 14
Abstract 15
Introduction 16
Materials and Methods 19
Development of nDNA Markers 19
Sampling, DNA Extraction, Amplification, and Sequencing 19
Alignment, polymorphism, and phylogenetic analyses 21
Population Differentiation and Structure 22
Gene flow 24
Results 25

Polymorphism 25

Population differentiation and structure 27

Gene flow 30

Discussion 31

References 37

Chapter 3 – Anthropocene invasion of an ecosystem engineer: Resolving the history of *Corophium volutator* (Amphipoda: Corophiidae) in the North Atlantic[a] 55

Abstract 56

Introduction 57

Methods 61

Sample collection, DNA extraction, amplification, and sequencing 61

Polymorphism and phylogenetic analysis 63

Colonization model selection 64

Results 67

Polymorphism and diversity 67

Colonization model selection 68

Discussion 69

Shared genetic diversity 70

Colonization models 71

Dispersal mechanisms and introduction vectors 72

Implications for ecology 74

Acknowledgements 76

References 77

Chapter 4 – The geographic scale of diversification was reshaped by shipping in the Age of Exploration[a] 98

Abstract 99

Introduction 100

Methods 102

Sampling 102

ddRADseq library preparation 103

ddRAD-seq pre-processing 104

ddRAD-seq de novo assembly 105

Genomic diversity and structure 106

Simulated demographic scenarios 107

Scans for selection 109

Results 110

Introduced populations show reductions in genetic diversity consistent with founder effects 110

Isolation and divergence between introduced and native ranges 111

Weak genetic differentiation and isolation-by-distance between native populations 112

Phylogenetic conflict within the native range 112

Human-mediated gene flow erodes ancestral genetic structure 114

Genetic patterns are consistent with neutral evolution and preclude inference of selection by genomic scans 115

Discussion 116

Acknowledgements 121

References 122

Chapter 5 – General discussion 141
References 153

Curriculum Vitae

List of Tables

Table 2.1 Primers used for amplification of nuclear and mitochondrial DNA..... 46

Table 2.2 Sampling sites, region, number of individuals sequenced per mitochondrial locus (*n*), number of haplotypes (*H*), number of segregation sites (*S*), haplotype diversity (*h*), and mean nucleotide diversity (π) for *Corophium volutator* populations 47

Table 2.3 Nuclear DNA diversity: Sampling sites, region, number of individuals sequenced per nuclear locus (*n*), number of segregation sites (*S*), number of alleles (*H*), observed heterozygosity (H_o), expected heterozygosity (H_e), and mean nucleotide diversity (π) for *Corophium volutator* populations surveyed..... 48

Table 2.4 Isolation by distance results from IBDWS 49

Table 2.5 AMOVA for samples of *C. volutator* from oceanographic regions using mitochondrial and nuclear markers..... 50

Table 2.6 Estimates of statistical power for detecting true levels of population structure (F_{ST}) by nDNA 51

Table 3.1 Potential causes of *Corophium volutator*'s disjunct distribution in the North Atlantic and predictions of genetic consequences 92

Table 3.2 Concatenated mitochondrial and nuclear DNA diversity for *Corophium volutator* populations surveyed..... 93

Table 3.3 Observed and expected number of haplotypes and alleles in Europe (NEA) and North America (NWA) for *Corophium volutator*..... 94

Table 3.4. Comparison of models using log Bayes factors (LBF) from Bezier approximated log marginal likelihood (Bezier lmL) estimated in MIGRATE-N for nuclear and mitochondrial DNA from *Corophium volutator* 95

Table 4.1 Population genetic diversity in *Corophium volutator* and *Hediste diversicolor* 131

Table 4.2 False positive rates of selection scans on simulated data 132

Table 4.3 Parasite and algal genomes filtered from genomic sequence data..... 133

List of Figures

Figure 2.1 Sampling locations and regional groupings of *Corophium volutator* populations in North America (see Table 2.2) 52

Figure 2.2 Results from Bayesian Analyses of Population Structure using clustering of groups of populations with spatial priors 53

Figure 2.3 Estimates of gene flow between regions 54

Figure 3.1 Haplotype network for concatenated *COI*/*SUBI* mitochondrial DNA for *Corophium volutator* and sampling locations 96

Figure 3.2 Diagrams of migration models tested using concatenated mitochondrial DNA (*COI*/*SUBI*) and nuclear DNA (*CV1* and *CV2*) of *Corophium volutator* in MIGRATE-N 97

Figure 4.1 Distribution of introduced and native invertebrate populations 137

Figure 4.2 Genetic divergence between, and phylogenetic conflict within, introduced and native ranges 138

Figure 4.3 Human-mediated migration erases ancestral genetic structure. 139

Figure 4.4 Hierarchical genetic structure 140

Chapter 1 – General introduction

The world is changing, and predicting how species will respond requires an understanding of the processes that underlie their evolution. In this thesis, I investigate how evolutionary trajectories are shaped by contemporary and historic patterns of movement, using distributions of genetic variation in focal invertebrates of the North Atlantic intertidal zone. In particular, I address three main questions: 1) How does life-history interact with environmental features to shape patterns of gene flow? 2) What are the roles of natural and human-mediated dispersal in shaping species' distributions? 3) How are species' evolutionary trajectories impacted by human-mediated dispersal? To provide background to the major components of my thesis, in this introduction I first outline the tradition of inferring movement from distributions of biological variation, and how investigating spatial distributions of heritable molecules provides insight into the evolutionary aspects of movement. I then briefly summarize the motivations for each following research chapter. More detailed motivations and background information are provided in the dedicated introductions of each chapter.

Inferring movement from distributions of variation

Life on earth is extraordinarily structured across geography, and the processes that cause this structure have fascinated scientists for centuries. The spatial distributions of biological variance are themselves clues about the processes that formed them, and can be used to inform models of how taxa are created and change over time. In one of the earliest examples of inference based on species' distributions, Carolus Linnaeus

documented the distributions of species and noticed they were often restricted to certain elevations (Linnaeus 1743). His model of how biodiversity was related assumed that the individuals within each species were relatives, but that separate species were created by god and immutable (Genesis Chapters 6-9). He therefore concluded that distributions structured by elevation were evidence that species had become established at progressively lower points from an Eden-like island as the ocean, which once covered everything but this island, receded (Cox et al. 2016). Some of the assumptions underlying Linnaeus' model were flawed, but his logic of using the distributions of related individuals to infer historic events and processes has withstood the test of time. Georges-Louis Leclerc applied the same principles in his 1761 "Histoire Naturelle", but his model considered that species were mutable and related. Leclerc posited that similar fauna in Eurasia and North America were likely closely related, and concluded (in some ways correctly) that they could only have moved between the two continents if the ocean was once lower and the climate much warmer (George-Louis-Leclerc 1761). This exemplifies the general tendency for improvements to a model to in turn improve the accuracy of inference.

Nearly a century later, the model underlying relatedness received its most meaningful update. Charles Darwin argued that selection on heritable traits caused populations to change and branch apart over time, giving rise to new taxa via evolution (Darwin 1859). This model established inter- and intra-specific change as part of the same continuum, governed by common principles. However, inferences of evolutionary history from morphological data lacked an underlying mechanism and were confounded by phenotypic plasticity, convergence of morphological traits, and unpredictable rates of

change over paleontological time. In the early 20th century, Darwin's theory of evolution was reconciled with Mendel's genetic theory in the Modern Synthesis (which had many contributors, but is summarized well by Huxley, 1942), providing a mechanistic model of evolution. The emerging framework was that evolution occurs due to new variation arising from mutations in DNA, with the frequency of variants in populations changing over time due to the interplay of the genomic substrate with natural selection, random genetic drift, and gene flow via dispersal (i.e. movement). This molecular evolution provided metrics of evolutionary change with sound, testable mechanistic models underlying those changes.

Inferring movement from molecular variation

Historic and ongoing patterns of movement can be difficult to evaluate by direct means for many species. Determining historic changes to species' distributions is particularly challenging in species that do not fossilize well, or belong to morphologically conserved species complexes. Contemporary dispersal is especially difficult to observe in species that are small or inhabit environments that are difficult to access. Spatial distributions of molecular variation provide an indirect means of testing hypotheses about the processes that underlie where species came from and how dispersal connects them, because the known modes of mutation and inheritance underlying genetic variation provide clear predictions for genetic patterns arising from movement and subdivision (Slatkin 1987; Hedgecock et al. 2007). Specifically, subdivided populations are expected to differentiate from each other over time due to mutation, random genetic drift, and

selection favoring adaptations to local environmental conditions. Movement of genetic material between populations (i.e. gene flow) is predicted to oppose that differentiation.

Patterns of gene flow can be investigated using polymorphic molecules that generally adhere to a particular model of mutation and inheritance, which are commonly referred to as genetic markers. One or more markers can be used to infer the evolutionary relationships between individuals, populations, species, or higher orders of life. The popularity of different markers has risen and fallen over time according to their degree of polymorphism, number of sites across a given genome, underlying mechanism of mutation, ease of characterization, and relative cost. This has been summarized many times elsewhere (Schlötterer 2004; Avise 2012), but a few key concepts are worth noting here.

While any marker that provides a measure of variation is useful, markers provide more information if the models underlying their mutation from one form to another are well understood. For instance, allozymes – proteins with different amino-acid substitutions between individuals separated by size and charge using gel electrophoresis (Hubby and Lewontin 1966) – are highly variable but do not contain information about the DNA-level changes that underlie differences between variants. They can identify differences between individuals, but the mutational steps between those differences cannot be measured or inferred. By comparison, stretches of mitochondrial DNA with many single base-pair differences between the sequences of individuals offer both variation and information about the states of common ancestors that the variants must have descended from. Such models that use mutational steps between sequences as a proxy for time are referred to as coalescent models (Kingman 1982), and can be used to

calculate demographic properties (such as the timing or magnitude of genetic exchange between subdivided populations) by comparing genetic divergence observed between different groups of individuals.

The physical location (e.g. nuclear vs. organellar), function (e.g. autosomal vs. sex-linked; coding vs. non-coding; synonymous vs. non-synonymous), and ploidy (e.g. haploid vs. diploid vs. polyploid) of a region of DNA directly impact its mode of inheritance and the rate at which it evolves (Moore 1995; Avise 2012). This can cause the genealogies of unlinked regions of DNA to differ from each other, presenting different representations of a species' net genealogy. Understanding of organism-level processes is thus improved by consideration of genetic information from across a species' entire genetic content. For example, in flowering plants maternally inherited chloroplast DNA is typically dispersed only by seeds, while allozymes encoded by nuclear DNA are bi-parentally inherited and dispersed across a broader geographic distribution by both seeds and pollen. One of the earliest studies using both of these marker types showed that this results in chloroplast DNA having much higher genetic differentiation across geographic space than allozymes encoded by nuclear DNA in the flowering plant *Silene alba* (McCauley 1994). The two markers provide information about different processes, but both processes are important to the evolution of the species as a whole. The synthesis of information from multiple markers and different regions of the genome is therefore extraordinarily informative, and increasing the amount of genetic information analyzed can better inform predictions of how species will respond to changing environments.

Movement and evolution in North Atlantic intertidal invertebrates

Predicting evolutionary change is particularly challenging for organisms found in ocean environments. This is because not only are ocean environments rapidly changing, but also patterns of gene flow are affected by the interaction of currents with life-history traits, species' distributions are shaped by natural and human-mediated dispersal, and human-mediated dispersal impacts gene flow between otherwise isolated populations (Carlton 2003). In marine habitats, invertebrates represent the bulk of biodiversity and biomass among animals, but direct evidence of historic and contemporary movement is sparse for many species because of their small size and soft tissues that fossilize poorly. In this thesis, I address these gaps by investigating spatial distributions of genetic variation in intertidal marine invertebrates in the North Atlantic.

Major global patterns of oceanic circulation are predicted to change in the near future due to global climatic changes in salinity and temperature (Hurrell et al. 2001; Drinkwater et al. 2009). Gene flow between populations of marine organisms is typically achieved by passive dispersal of planktonic propagules in the water column via currents (Cowen and Sponaugle 2009), which facilitate admixture within basins by mixing and create barriers to gene flow where they diverge (Shanks et al. 2003; Riginos and Liggins 2013). Direct-developing invertebrate species (i.e. lacking a pelagic larval stage) are expected to be less affected by currents, as dispersal is typically achieved by the active movement of adults (Bradbury et al. 2008). However, behaviors that lead to juveniles or adults spending time in the water column may facilitate passive dispersal, albeit over smaller spatial scales than species with planktonic larvae. In Chapter 2 I explore how currents shape gene flow amongst populations of *Corophium volutator,* one of the most

abundant and ecologically important invertebrates in the sedimentary coasts of the Northwest Atlantic. The natural dispersal ability of *C. volutator* is difficult to assess by direct means; while it brood-rears its young (lacking a pelagic larval phase by which marine invertebrates typically achieve long-distance dispersal) and spends much of its adult life in constructed burrows, its vertical swimming behaviour may allow passive dispersal between habitat patches linked by currents (Drolet et al. 2012). I develop three novel nuclear markers and analyze these in combination with mitochondrial markers using frequency and coalescent-based methods to assess how life-history can interact with the physical environment to determine patterns of gene flow.

Increasing global human movement has the potential to create new ecological associations (e.g. competition, predation) by introducing species beyond the range of their natural dispersal capabilities. Despite a long history of human activity in the North Atlantic and temporal biases in observation, abundant species are often assumed to be native by default. However, there is often no evidence supporting this assumption. Resolving the origins and colonization pathways of species with unknown histories is a requisite for understanding the ecological and evolutionary context of community interactions (Elton 1958; Carlton 2001), particularly for species with large or wide-reaching ecosystem effects. For example, despite having a relatively limited dispersal ability and being distributed on both sides of the Atlantic, *C. volutator* has been assumed as native across its range. Its present distribution in the Northwest Atlantic coast was previously glaciated ~18,000 years ago (Hewitt 2000), suggesting that it may have colonized this region via natural dispersal from southern glacial refugia (Maggs et al. 2008), or from less glaciated regions of the Atlantic via human activity (Carlton 1996). In

Chapter 3, I investigate the evolutionary history of *C. volutator* in the North Atlantic using mitochondrial and nuclear sequence data. I weigh the support for recent human-mediated movement, historic human-mediated movement, and natural post-glacial movement to investigate the potential role of each in maintaining and creating ecological relationships.

Globally increasing human activity can alter dispersal pathways in species associated with human vectors of transport (Lacoursière-Roussel et al. 2012; Hudson et al. 2016), but little is known about how creating new populations and moving individuals amongst populations in their native ranges alters species' evolutionary trajectories. This is in part because understanding of how human movement shapes evolution comes largely from observations of recent introductions that have not been separated from their sources for very many generations. To determine how gene flow from human transport affects the evolutionary trajectories of species in their native and introduced ranges, in Chapter 4 I investigate genomic patterns in the historically introduced invertebrates, *C. volutator* and *Hediste diversicolor*, across the North Atlantic. These species co-occur, have similar life-histories, and have similar histories of introduction from the Northeast to the Northwest Atlantic[1] (Einfeldt et al. 2014). Their inherent traits suggest strong

[1] In his B.Sc.Honours thesis, Jeremy Doucet investigated the relationship between genetic structure and hydrology in *Hediste diversicolor*. I generated additional data and combined these with published sequences to resolve this species' introduction history. Because the associated publication overlaps with

association with transport via semi-dry ballast and weak association with transport via modern vectors, which provides limits for how long introduced populations have been isolated from their sources. I use the genetic divergence accumulated between introduced and native ranges, isolated from each other for ~150-400 years, to compare patterns of divergence in the species' native range and evaluate the effect of human-mediated dispersal across both native and introduced populations.

The research in Chapters 2-3 has been published before the submission of this thesis (Einfeldt and Addison 2013; Einfeldt and Addison 2015). The research in Chapter 4 is currently in review.

Jeremy Doucet's B.Sc.Honours thesis and demonstrates nearly perfect parallels to my research on *C. volutator* and thus contributes no new conceptual material, it is not included in this thesis.

References

Avise JC. 2012. Molecular markers, natural history and evolution. Springer Science & Business Media.

Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, Campana SE. 2008. Global patterns in marine dispersal estimates: the influence of geography, taxonomic category, and life history. Proc R Soc Biol 275: 1803–1809.

Carlton JT. 1996. Biological invasions and cryptogenic species. Ecology. 77: 1653–1655.

Carlton JT. 2001. The scale and ecological consequences of biological invasions in the World's oceans. Invasive species and biodiversity management. 24: 195.

Carlton JT. 2003. Community assembly and historical biogeography in the North Atlantic Ocean: the potential role of human-mediated dispersal vectors. Hydrobiologia. 503: 1-8.

Cowen RK and Sponaugle S. 2009. Larval dispersal and marine population connectivity. Annual Reviews. 1: 443-466.

Cox CB, Ladle R, and Moore PD. 2016. Biogeography: an ecological and evolutionary approach. John Wiley & Sons.

Darwin C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: Murray.

Drinkwater KF, Mueter F, Friedland KD, Taylor M, Hunt GL, Hare J, Melle W. 2009. Recent climate forcing and physical oceanographic changes in Northern Hemisphere

regions: A review and comparison of four marine ecosystems. Progress in Oceanography. 81: 10-28.

Drolet, D., Bringloe, T. T., Coffin, M. R., Barbeau, M. A., & Hamilton, D. J. 2012. Potential for between-mudflat movement and metapopulation dynamics in an intertidal burrowing amphipod. Marine Ecology Progress Series, 449, 197-209.

Einfeldt AL and Addison JA. 2013. Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the northwest Atlantic. Marine biology. 160: 1015-1027.

Einfeldt AL, Doucet JR, Addison JA. 2014. Phylogeography and cryptic introduction of the ragworm *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. Invertebrate biology. 133: 232-241.

Einfeldt AL and Addison JA. 2015. Anthropocene invasion of an ecosystem engineer: resolving the history of *Corophium volutator* (Amphipoda: Corophiidae) in the North Atlantic. Biological journal of the Linnean Society. 115: 288-304.

Georges-Louis-Leclerc, B. 1761. Animaux communs aux deux continents, t. IX de l'Histoire naturelle générale et particulière avec la description du cabinet du Roy, édité par Jean Piteveau, Œuvres philosophiques de Buffon, PUF.

Hedgecock D, Barber PH, and Edmands S. 2007. Genetic approaches to measuring connectivity. Oceanography. 20: 70-79.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

Hubby JL and Lewontin RC. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics. 54: 203–215.

Hudson J, Viard F, Roby C, Rius M. 2016. Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. Biology letters. 12: 20160620.

Hurrell JW, Kushnir Y, and Visbeck M. 2001. The north Atlantic oscillation. Science. 291: 603-605.

Huxley J. 1942. Evolution the modern synthesis. George Allen and Unwin.

Kingman JFC. 1982. The coalescent. Stochastic processes and their applications. 13: 235-248.

Lacoursière-Roussel ANAÏS, Bock DG, Cristescu ME, Guichard F, Girard P, Legendre P, McKindsey CW. 2012. Disentangling invasion processes in a dynamic shipping–boating network. Molecular ecology. 21: 4227-4241.

Von Linnaeus, C. 1743. On the growth of the habitable earth. Select Dissertations from the Amoenitates Academicae.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Vainola R, Viard F, Wares J. 2008. Evaluating signature of glacial refugia for North Atlantic Benthic Marine Taxa. Ecology 89: S108–S122.

McCauley DE. 1994. Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. Proceedings of the National Academy of Sciences. 91: 8127-8131.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution. 49: 718-726.

Riginos C and Liggins L. 2013. Seascape genetics: populations, individuals, and genes marooned and adrift. Geography compass. 7: 197-216.

Shanks AL, Grantham BA, and Carr MH. 2003. Propagule dispersal distance and the size and spacing of marine reserves. Ecological applications, S159-S169.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science. 236: 787-792.

Chapter 2 – Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the Northwest Atlantic[a]

Einfeldt AL[1] (corresponding author), Addison JA[2]

[a]*Published in Marine Biology, January 2013*

[1]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3. email: tony.einfeldt@gmail.com.

[2]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3

Abstract

The mechanisms driving genetic structure in marine systems are elusive due to the difficulty of identifying temporal and spatial barriers to dispersal. By studying marine invertebrate species with limited dispersal potential, genetic structure can be directly related to physical and biological factors restricting connectivity. In the northwest Atlantic, the benthic brood-rearing amphipod *Corophium volutator* is distributed across basins with distinct circulation patterns, and has the potential to disperse passively during its adult stage. We analyzed spatial genetic variation and migration rates across *C. volutator*'s North American range with sequence data for mitochondrial DNA and three novel nuclear markers using frequency and coalescent-based methods. We found low genetic differentiation within basins, but strong subdivision within the Bay of Fundy and a striking biogeographic break between the Bay of Fundy and Gulf of Maine, suggesting that genetic drift may act on populations in which connectivity is restricted due to the limitation of passive dispersal by hydrological patterns.

Introduction

Population connectivity affects the contemporary ecosystem dynamics through demographic changes via dispersal (Wright 1931) and the evolutionary trajectories of species through population genetic changes caused by gene flow (Watson et al. 2012). In marine environments, barriers that limit connectivity between populations can be subtle, and dependent on the intrinsic dispersal ability of a species (Bohonak 1999; Bradbury and Snelgrove 2001; Bradbury et al. 2008), local oceanographic processes (Lee et al. 1992), and demographic history (Avise 2000). Molecular tools provide an indirect method of investigating barriers to gene flow and the likely boundaries of demographic connectivity, elucidating the role of contemporary processes in generating distributions of genetic variation (Lowe and Allendorf 2010).

Mode of larval development is often considered the most important factor in determining the scale of population connectivity and degree of genetic subdivision in marine invertebrates, as species lacking passive larval dispersal generally exhibit stronger patterns of population divergence than those with planktonic larvae (Palumbi 1994; Hellberg 1996; Matthaeis et al. 2000; Collin 2001; Teske et al. 2006; Janson 2008). The passive dispersal of planktonic larvae provides a mechanism for connectivity at large spatial scales, often causing patterns of population subdivision that correspond to currents that facilitate larval dispersal via advection (Teske et al. 2008; Jennings 2009; White et al. 2010). In contrast, species with direct development or crawl-away larvae are intrinsically less influenced by currents and are expected to exhibit isolation by distance over a smaller geographic scale, with genetic similarity correlating more to the degree of

geographic separation between populations than to hydrological patterns (Goldson et al. 2001; Grantham et al. 2003; Cowen and Sponaugle 2009; Weersing and Toonen 2009; Kelly and Palumbi 2010). However, the generalization that dispersal potential is reduced in taxa lacking a planktonic larval stage assumes that individuals in their adult stage remain localized, overlooking other potential dispersal mechanisms such as adult swimming behaviour.

The marine amphipod *Corophium volutator* is an abundant prey item for many species of fish and migratory shorebirds throughout soft sediment environments in the Bay of Fundy and Gulf of Maine (Shoemaker 1947; Peer and Linkletter 1986), and contributes to sediment stability through the construction of burrows (Meadows et al. 1990; Mouritsen et al. 1998). Like all amphipods, *C. volutator* lacks a planktonic larval dispersal stage, and is expected to exhibit genetic isolation by distance between populations due to the absence of a mechanism for larval dispersal. However, *C. volutator* is known to actively enter the water column during flood tides (Hughes 1988; Lawrie and Raffaelli 1998; Drolet et al. 2009), and can survive for days in drift ice moving tens of kilometers before being deposited on mudflats elsewhere (Macfarlane et al. 2009). These passive dispersal mechanisms may facilitate high gene flow through advection within regions where currents mix (e.g. Teske et al. 2006; White et al. 2010), contrary to predictions based on assumptions of limited dispersal inferred from life history alone. The Bay of Fundy is home to the highest tidal range in the world (Garrett 1972; Desplanque and Mossman 2001), exposing intertidal habitat to exceptionally strong hydrological forces and causing characteristically strong currents throughout the bay. These forces may facilitate the homogenization of genetic variation within regional

water masses via passive adult dispersal, which would result in patterns of population genetic subdivision that correspond to local hydrology rather than the geographic distance between populations.

To investigate the roles of dispersal ability and local hydrology in generating patterns of connectivity in soft sediment environments, we studied the population genetic structure of *C. volutator* across its North American range using two mitochondrial and three novel nuclear markers. This range encompasses two well-defined regions within the northwest Atlantic that are hydrologically segregated by splitting gyres: the Bay of Fundy and the Gulf of Maine (Fig. 2.1). The Bay of Fundy is further subdivided into a main trunk, which experiences high tidal velocities, and two distinct basins that experience extreme tidal velocities: Minas Basin and Chignecto Bay. If strong tidal currents are capable of moving adult *C. volutator* between mudflats, we predict that connectivity will be high within the admixing water masses of the Bay of Fundy and the Gulf of Maine, and restricted between them. As the Bay of Fundy itself is subdivided by two major currents, we predict that genetic connectivity will be restricted between the main trunk and separate basins within the Bay of Fundy, but to a lesser extent than between the Bay of Fundy and Gulf of Maine due to greater current velocities occurring deeper within the Bay of Fundy.

Materials and Methods

Development of nDNA Markers

We developed nuclear markers using an amplification, cloning, and sequencing strategy similar to Naqvi and Chatoo (1996). Products from randomly amplified polymorphic DNA (RAPD) were screened for consistent amplification from twelve individuals using the RAPD primers UBC367 (UBC), OPA9 (Operon Technologies), and M13 (Pharmacia). RAPD-PCR was performed in 25 μL reactions with 2 ng DNA template, 1 X ThermoPol reaction buffer (New England Biolabs; NEB), 0.2 mM dNTPs (NEB), 2 mM $MgSO_4$, 0.5 μM primers, and 1 U *Taq* polymerase (NEB) with thermal cycling conditions of 94° (2:00); 40 cycles of 94° (1:00), 35° (1:00), and 72° (2:00); and 72° (5:00). Bands were separated by electrophoresis on 2% agarose gels, visualized under blue light with SybrSafe, excised, and purified using QIAGen Gel Purification Kit. DNA products were cloned and sequenced using a CloneJET™ kit (Fermentas) following a standard protocol**.** These sequences were then used to engineer primer pairs (Table 2.1) using Primer3 (Rozen and Skaletsky 2000).

Sampling, DNA Extraction, Amplification, and Sequencing

We collected adult *C. volutator* from mudflats and soft sediment in bays, river estuaries, and salt marshes from Cape Cod, Massachusetts to as far north as the Iles de la Madeleine, Québec during extensive field surveys between June 2009 and August 2011. In total, 24 geographic locations (hereafter referred to as populations) were collected in the Bay of Fundy and Gulf of Maine, encompassing the species' present extant range in

North America (Fig. 2.1 and Table 2.2). All specimens were preserved in 95% ethanol and stored at -20°C.

DNA was extracted using a cetyltrimethylammonium bromide (CTAB) protocol (Grosberg et al. 1996), with remaining tissue stored in 95% ethanol at -80°C as a voucher. Eggs and larvae were excised from the brood-pouches of gravid females prior to DNA extraction to avoid amplification of nuclear DNA from multiple individuals.

Preliminary sequences using primers for cytochrome *c* oxidase subunit 1 (*COI*; Folmer et al. 1994) were predominantly monomorphic, and provided little resolution of population genetic structure. Thus we targeted a second mtDNA fragment using primers SUBIF from Davolos and Maclean (2005) and COII-CROZ from Crozier et al. (1989) to amplify and sequence a region spanning a portion of *COI*, an intergenic spacer, and *COII*. From these sequences, we engineered the specific reverse primer SUBIRA, used with SUBIF to provide an additional mitochondrial DNA marker for population genetic analyses (hereafter *SUBI*).

PCR reactions for both mtDNA and nDNA loci were performed in a 20 μL reaction volume with 20 ng DNA template, 1 X ThermoPol reaction buffer (NEB), 0.2 mM dNTPs (NEB), 2.5 mM $MgSO_4$, 0.5 μM forward and reverse primers (Table 2.1), and 0.8 U *Taq* polymerase (NEB). Thermal cycling conditions were 95° (3:00); 34 cycles of 95° (0:30), 45° (0:45), and 72° (1:00); and 72° (1:00). Amplicons were checked by electrophoresis on 1% agarose gels and direct sequenced using forward, reverse, or both primers at the McGill University and Genome Québec Innovation Centre (Montréal, Québec).

Alignment, polymorphism, and phylogenetic analyses

All sequences were aligned, edited, and trimmed to standard length in Sequencher 5.0 (Gene Codes). Maximum likelihood estimated transition/transversion bias for analyses in MIGRATE was calculated for mitochondrial sequences under the Kimura (1980) 2-parameter model using MEGA v4.0 (Tamura et al. 2007). Mitochondrial sequences were collapsed into haplotypes and mapped into a haplotype network using statistical parsimony as implemented in TCS v1.21 (Clement et al. 2000). The allelic phase of nuclear sequences for heterozygous individuals was inferred using PHASE v2.1.1 (Stephens et al. 2001; Stephens & Scheet 2005) for 10,000 iterations and a burn-in of 1000, with all haplotype reconstructions included in the data analyses having a posterior probability $P>0.95$ consistently over 3 runs. All sequences are available from Genbank (Accessions #JX991308-JX992554).

Haplotype and nucleotide diversities were calculated for each population sample at mitochondrial and nuclear loci using DNASP v5.1 (Librado and Rozas 2009). We calculated expected and observed heterozygosities and fixation indices using ARLEQUIN v3.11 (Excoffier et al. 2005). We used ARLEQUIN to test for departures from neutrality based on allelic states or segregating sites by calculating Fu and Li's *F* and Tajima's *D*, respectively. The significance of these neutrality tests was assessed by performing 10,000 bootstrap replicates. Nuclear markers were confirmed to be unlinked using GENEPOP v1.2 (Raymond and Rousset 1995; Rousset 2008).

Population Differentiation and Structure

We assessed the statistical power of our nuclear data for three loci to detect genetic differentiation between populations given differences in allelic diversity and sample sizes with the program POWSIM v.4.1 (Ryman and Palm 2006). We assessed power for two models: 24 populations, reflecting the level of sampling in our study; and 3 populations, reflecting a coarse regional division of our study area. Default parameter values were used for memorizations (1000), batches (100), and iterations per batch (1000) over 1000 repeated simulations for each value of F_{ST}. Statistical power was determined as average of three repeated analyses estimating the proportion of simulations with significant differences (P<0.05) from Fisher's exact test.

We computed global F_{ST} and tested for pairwise genetic differences between mudflats (F_{ST}) using the distance based approach (Kimura 2-parameter) implemented in ARLEQUIN, with significance assessed using 10,000 permutations of the data. We examined the relationship between genetic isolation and geographic distances between mudflats using Mantel tests implemented in IBDWS (Wright 1943; Jensen et al. 2005) to test for correlation between Slatkin's genetic similarity index (Slatkin 1993) and log geographic distances, which is recommended for two-dimensional models (Rousset 1997). Pairwise geographic distances were measured according to the shortest possible marine pathway between GPS site coordinates for each population. Isolation by distance was analyzed for *SUBI and CV1+CV2+CV3* (*COI* lacked sufficient diversity for analyses to be informative) across the study range and within regions.

To investigate the role of hydrological patterns in structuring populations, we used indices of differentiation (Φ) with 10,000 permutations in a hierarchical analysis of

molecular variance (AMOVA) implemented in ARLEQUIN to test *a priori* expectations of population structure derived from contemporary current patterns using a Kimura (1980) 2-parameter distance model (Fig. 2.1). Differences in genetic variation among regions were explored for the major divisions between the inner Bay of Fundy (two separate basins; East and West), outer Bay of Fundy (a region of reciprocating currents; Middle), and Gulf of Maine (separated from Fundy by segregating gyres; Outer). To account for current vectors and the coarse scale at which information was available, we explored the assignment of populations near the borders of group subdivisions (DGS, DGN, MAC, LBS, SMS, SMN; see Fig. 2.1) by including them in either adjacent region in subsequent analyses.

To explore our data for underlying genetic structure we implemented the clustering method described by Pritchard et al. (2000) in STRUCTURE v2.3.3 to identify genotype groupings by assigning individual genotypes to *k* groups, minimizing Hardy-Weinberg and linkage disequilibrium in an admixture model with correlated allele frequencies between populations and no population priors. For each value of *k* up to *k*=10, 20 independent runs of a Bayesian MCMC search of 100,000 steps with a burn-in of 50,000 were performed for combined mtDNA and nDNA, nDNA only, and mtDNA only (mtDNA coded as a single phase with the other missing in combined analyses). To find the value of *k* that best estimates the number of clusters for the combined and independent mitochondrial and nuclear data sets, we indexed the rate of change in the log probability of data between successive values of *k* using the method of Evanno et al. (2005).

To incorporate spatial information into our investigation of population structure, we explored the clustering of genetically similar populations with spatial coordinates of sample sites as priors for mitochondrial and nuclear sequence data in BAPS v5.4 (Corander et al. 2006; Corander et al. 2008), which uses a stochastic optimization algorithm to analyze Bayesian models of population structure. Ten replicates were run with *k* (maximum number of clusters) set to 24 in all analyses, including mitochondrial data alone, nuclear data alone, and both with and without spatial priors.

Gene flow

To explore the relationship between gene flow and hydrology, we estimated the vectors of gene flow between contiguous geographic regions using coalescent analyses of migration rates (M) and effective population sizes (N_e) for mtDNA and nDNA sequence data separately using MIGRATE v3.2.16 (Beerli and Felsenstein 1999; Beerli 2006). We used population groupings consistent with results from analyses using AMOVA and BAPS, estimating asymmetric rates of gene flow between four groups of populations for mtDNA (East, West, Middle, and Outer) and three groups of populations for nDNA (East+West, Middle, and Outer; Table 2.2). We implemented a stepping-stone model of gene flow that fixed migration rates between non-adjacent population groupings to zero to reflect the limited passive dispersal potential of *C. volutator*.

A Bayesian MCMC search based on a chain of 200,000,000 steps sampled every 100 steps for a total of 2,000,000 samples, with a burn-in of 50,000 steps and four additional heated chains. Exponential prior distributions were used for both migration

rate (M=2,000,000 for mtDNA; 110,000 for nDNA) and effective population size (θ=0.015 for mtDNA, 0.006 for nDNA), which were tested for potential bias by repeating the simulation process with uniform priors. The results from these search strategies were consistent with maximum likelihood MCMC searches with five replicate runs of 20 short chains of 5,000,000 steps sampled every 50 steps for a total of 100,000 samples.

Results

Polymorphism

Nuclear markers *CV1* and *CV2* exhibited similar genetic diversity. Sequences from *CV1* nDNA primers were the most diverse, with 1-5 alleles per population totaling 9 alleles in 252 individuals. Sequences from *CV2* primers yielded 2-5 alleles per population totaling 7 alleles in 191 individuals (Table 2.3). The least diverse nuclear marker was *CV3*, with only 1-2 alleles per population totaling 2 alleles in 200 individuals; these low levels of polymorphism were uninformative, and did not contribute to detection of population genetic patterns. Haplotype diversities for nuclear markers ranged from 0.0000-0.8333, and nucleotide diversities ranged from 0.0000-0.0111. Values of global and population level F_{IS} did not differ significantly from those expected under Hardy-Weinberg equilibrium for any of the nuclear markers.

Of the 152 samples sequenced at the *COI* locus, 149 were monomorphic, and three were unique haplotypes differing from the common haplotype each by one single nucleotide polymorphism. Due to the absence of diversity, *COI* data were not included in any further analyses.

At the *SUBI* mitochondrial locus, we detected 7 haplotypes in 452 individuals, three of which were common (n=268, n=91, n=89) and 4 of which were singletons (Fig. 2.1). When populations were pooled into regions, haplotype diversity for *SUBI* was highest in the Middle region (h=0.5185) and lowest in the Outer region (h=0.0503), where it was also lowest for the three nuclear markers. Negative values of Tajima's D were exhibited by three populations (MN, -1.4950; PC, -1.7295; LCN, -1.4296; P<0.05), reflecting a slight excess of low-frequency mutations and potentially suggesting population expansion.

Maximum parsimony and neighbor-joining analyses for *SUBI* show that the two most genetically distant common haplotypes differ by 3 base pair mutations (n=268 and n=91) and co-occur in the East, West, and Middle regions, but are absent from the Outer region (Fig. 2.1). The third most common haplotype (n=89) was phylogenetically intermediate to the two most common haplotypes, differing from them by 2 and 1 base pair substitutions, respectively. This third haplotype was absent from the East and West groups, present in two populations of the Middle group (SJS and POC), and predominant in the Outer region, where all populations except LCN were monomorphic. Singletons differing from the most common haplotype each by one base pair substitution were found in LCN, MN and NB. SJS was the only population in which all three common haplotypes were found.

Population differentiation and structure

Estimates of statistical power indicated that nDNA data for three loci were capable of identifying a true F_{ST} of 0.0200 with a probability of >97% for a model of 3 populations, and a true F_{ST} of 0.0250 with a probability of >95% for a model of 24 populations (Table 2.6). We found a <5% probability of detecting genetic structure when true F_{ST}=0 (Type I error) for both the 3 population and 24 population models.

Overall values of F_{ST} were highest for the SUBI marker (F_{ST}=0.7167; P<0.0001), with less differentiation between populations found for nDNA (F_{ST}=0.2136; P<0.0001) Pairwise comparisons of populations showed significant genetic differentiation (F_{ST}) between 182 of 276 comparisons (66%) with *SUB1* and 194 of 276 comparisons (70%) with nDNA *SUBI* showed few significant differences in populations compared within the four separate geographic regions of West, East, Middle, and Outer, but many significant differences in populations compared between these regions. In the Middle region, SJS and LBS showed significant differentiation from all populations except each other, and POC was significantly differentiated from all populations at *SUBI*.

No significant correlation between genetic distance and geographic distance was found using either mtDNA or nDNA at the scale of the entire study range or within regions (Table 2.4). For *SUBI*, pairwise comparisons resulted in a disproportionate number of pairs that were either highly similar or highly differentiated. The 95% confidence intervals for slope encompassed zero in all analyses; the hypothesis that geographic distance is a predictor of genetic distance was not supported by mitochondrial or nuclear data.

Values of Φ_{CT} obtained with AMOVA from the nDNA data were considerably lower than with mtDNA (Table 2.5), but produced similar results. Significant subdivision was detected using nDNA between the groups East+West, Middle, and Outer (Φ_{CT}=0.0328, P<0.05). For *SUBI*, the highest values of Φ_{CT} were found by comparing the groups East+West+Middle vs. Outer, and significant values of Φ_{CT} were obtained by further separating the Bay of Fundy according to major currents into the groupings East, West, and Middle. Significant genetic differentiation was detected between all groups (P<0.025), with the largest differentiation between West vs. Outer and East vs. Outer regions, moderate differentiation between West vs. Middle and Outer vs. Middle regions, and the least differentiation between East vs. Middle and East vs. West regions. Although AMOVA indicated a significant genetic subdivision between East and West (P=0.0040), this pairing had a lower value of Φ_{CT} (=0.1436) than all other comparisons among the four groups.

Analysis of replicated outputs from STRUCTURE using Evanno's method weakly identified five clusters from mtDNA data, two clusters from nDNA, and two clusters from combined data. For *SUBI*, the highest peak of ΔK occurred at k=5, and the inability to identify a mixture model when using only a single locus resulted in equal individual assignment probabilities to each cluster. For both the nDNA and the combined nDNA+mtDNA data sets, the highest peak of ΔK occurred at k=2, with values of ΔK within the same order of magnitude for all analyses. The population assignment probability of individuals for nDNA at k=2 did not exceed 0.803, and for combined nDNA+mtDNA did not exceed 0.898, with some individuals having assignment probabilities of 0.500 for each group with both data sets. STRUCTURE does not perform

well with a low number of loci or when migration between regions is high (Evanno et al. 2005), explaining the weakness of the signal detected using this method.

To overcome the ambiguity of individual assignment tests when investigating a small number of loci, we incorporated information on which population individuals belong to as priors in BAPS. Clustering of groups of individuals with and without spatial priors reflected our *a priori* geographic subdivisions based on hydrology (Fig. 2.2). For the *SUBI* locus, we detected 4 clusters: Outer; West+OBN; East; and LBS+SJS+POC. POC's grouping was ambiguous; clustering POC as a fifth cluster or with SJS+LBS did not change measures of variance or log marginal likelihoods (0.0), and the consolidation of LBS+SJS+POC to create a total of 4 clusters was equally as likely an arrangement as the 5-cluster scenario. Due to the central geographic location of LBS, POC, and SJS within the study range, we favored the 4-cluster result as a more parsimonious solution. OBN's placement with distant Chignecto Bay populations despite being geographically situated between LBS and POC was the only exception to a pattern of geographically contiguous groups. To maintain distinct geographic boundaries between clusters for our analysis of gene flow, we placed OBN with the three other central populations. This did not have a strong effect on the clustering log marginal likelihood (-3.3), as compared to moving it to the Outer group (-37.3). Combined nuclear and mitochondrial data without spatial priors resulted in 5 clusters: Outer; East+SMS-MN; West+MN+OBN; LBS+SJS; and POC. Inclusion of POC with LBS+SJS did not have a strong effect on the clustering log marginal likelihood (-2.8) as compared to moving it to any other group. Adding spatial priors to combined nuclear and mitochondrial data resulted in three clusters: East+West+OBN; Outer; and SJS+POC+LBS. This spatial structure was similar to the

analysis with *SUBI* data alone, with the exception of East+West being grouped together. Analysis using only nuclear markers resulted in two clusters: East+West+SJS+OBN+MAC and Outer+POC+LBS. The assignment of SJS+OBN+MAC and POC+LBS to spatially non-contiguous groups suggests the presence of a contact zone between the two major clusters, which was grouped as a separate cluster to maintain distinct geographic boundaries for our analysis of gene flow and allow comparison between patterns shown by mtDNA and nDNA.

Gene flow

Coalescent analyses of gene flow were used to investigate the relationship between hydrological patterns and connectivity between four groups of populations for mtDNA and three groups of populations for nDNA, assigned based on previous analyses of genetic similarity, structure, and geographic proximity. Analyses were performed separately for nuclear and mitochondrial markers, with the groups East and West collapsed in analyses of nuclear data in order to avoid estimating non-informative migration parameters between groups identified as a single cluster in analyses of genetic structure.

The coalescent analysis of gene flow in MIGRATE with *SUBI* identified asymmetric migration rates between regions that varied by over four orders of magnitude (Fig. 2.3). Gene flow (*M*) vectors were greater into East from each adjacent region than out of East, higher out of West to each adjacent region than into West, and higher from Outer to Middle than from Middle to Outer. Migration rates between Middle and East

exhibited the least directional asymmetry, but *M* from Middle to East was still 2.5x as great as in the opposing direction. Rates of gene flow between the three Bay of Fundy regions were relatively high compared to rates between the Bay of Fundy and the Gulf of Maine (Middle and Outer), with the exception of *M* from Middle to West, where rates were comparable to those from Outer to Middle. Patterns of gene flow estimated using nDNA showed similar asymmetry to analysis using mtDNA. Estimates of gene flow from Middle to both Outer and East+West had 95% confidence intervals that included zero, with gene flow into Middle from Outer and East+West being several orders of magnitude greater; consistent with the treatment of Middle as an interface region. This striking restriction of gene flow detected using both mitochondrial and nuclear markers at the segregating Bay of Fundy and Gulf of Maine gyres (Fig. 2.1) characterizes a strong biogeographic break; an expected pattern if major currents play a role in structuring populations of *C. volutator*.

Discussion

Our sequence data from throughout the North American range of *C. volutator* reveal a striking biogeographic break between the Bay of Fundy and the Gulf of Maine, with further genetic subdivision between regions within the Bay of Fundy. Patterns of genetic divergence among populations are inconsistent with a model of isolation by distance, and instead correspond to the regional boundaries defined by major hydrological patterns separating the Bay of Fundy from the Gulf of Maine and further subdividing basins within the Bay of Fundy. Our results suggest that hydrology may

influence gene flow via the limitation of dispersal, enabling genetic drift to act on populations with reduced connectivity.

Mode of dispersal has been implied as a strong determining factor of genetic patterns in marine invertebrates, with species lacking a planktonic larval stage typically having a much higher degree of genetic differentiation amongst populations than species with planktonic larvae over similar geographic scales (Levin 2006; Weersing and Toonen 2009). While *C. volutator* showed a high degree of genetic differentiation (*SUBI* F_{ST}=0.7167, P<0.0001; nDNA F_{ST}=0.2136, P<0.0001), consistent with our expectations for a benthic brood-rearing amphipods with sediment-associated burrowing and feeding (e.g. Ashton et al. 2008; Hogg et al. 2006; Kelly et al. 2006), we failed to detect isolation by distance with mtDNA or nDNA. Isolation by distance is most evident in marine invertebrates between populations separated by two to five times the species' mean dispersal distance (Palumbi 2003), and the surprising lack of correlation between genetic and geographic distance in *C. volutator* suggests a far greater dispersal potential than expected for a brood-rearing amphipod; possibly due to regional currents driving dispersal of adults that swim into the water column, facilitating increased gene flow within basins and causing restricted gene flow across regions where currents diverge.

A phylogeographic break between the Bay of Fundy and Gulf of Maine was characterized by statistical analyses of population structure (AMOVA and BAPS) and coalescent analysis in MIGRATE, indicating a strong barrier to gene flow occurring between the Bay of Fundy and Gulf of Maine (Fig. 2.3). None of the mitochondrial haplotypes present in the Bay of Fundy were found in the Gulf of Maine, which may be a result of asymmetrical gene flow from the Gulf of Maine to the Bay of Fundy. Nuclear

data supported the division between the Bay of Fundy and Gulf of Maine, but did not increase the resolution of population structure within the Bay of Fundy. The identification of the major phylogeographic break using both nDNA and mtDNA but fine-scale subdivisions being identifiable only with mtDNA demonstrates the importance of using multiple independent loci in population genetic analyses, and is consistent with smaller effective population sizes for mtDNA resulting in more rapid fixation of alleles in isolated populations as a result of genetic drift (Addison et al. 2008; Slatkin 1987, Palumbi et al. 2001).

The phylogeographic break between the Bay of Fundy and Gulf of Maine coincides with the splitting of major water masses (Fig. 2.1), where predominant currents south of this break continue into the Gulf of Maine gyre, and predominant currents north of this break carry into the Bay of Fundy gyre and follow a reciprocating pattern further up the Bay of Fundy (Pettigrew et al. 2005; Aretxabaleta et al. 2008; Manning et al. 2009). The magnitude of gene flow revealed by mtDNA is greatest between regions in the inner Bay of Fundy, where tidal currents have the greatest velocity due to the focusing of large masses of water into narrow topographies. The division of water masses between the groups East and West also corresponds with significant genetic subdivision between these two regions, whereas populations within each respective region exhibit a high degree of genetic similarity to each other. Nuclear data showed similar patterns of gene flow between the Bay of Fundy and Gulf of Maine, with the greatest gene flow occurring from the Gulf of Maine and Bay of Fundy into a contact zone (Middle). The agreement of hydrological patterns with barriers to gene flow, coupled with the failure of isolation by distance to explain patterns of genetic subdivision despite a high degree of

differentiation between populations, suggests that hydrology influences population genetic structure in *C. volutator* by promoting passive adult dispersal within basins, resulting in hydrology-defined regions of genetic homogeneity.

The concordance of genetic subdivision with hydrological patterns suggests that restricted gene flow between regions may provide an opportunity for the development of localized subpopulations to occur in response to varying environments and an absence of the homogenizing effects of gene flow. Previous investigations have found demography (Barbeau et al. 2009) and genetic variation using RAPDs (Wilson et al. 1997) in *C. volutator* to be highly variable between mudflats in the Bay of Fundy, suggesting that spatially discrete mudflats are the unit of highest interest for ecological and conservational consideration. However, the potential for between-mudflat movement facilitated by high tidal velocities (Drolet et al. 2012) suggests that larger isolated regions may be more important to the evolution and ecology of *C. volutator*, with population connectivity having a large influence on population genetic patterns following population bottlenecks occurring at different spatial scales.

Several processes may contribute to genetic bottlenecks in *C. volutator*, at the scale of both individual mudflats and over larger regions. The interannual variation and periodic population crashes experienced by some mudflats (Hicklin and Smith 1984; Peer and Linkletter 1986; Shepherd et al. 1995; Sprague et al. 2008) undoubtedly have the potential to cause genetic bottlenecks at the scale of individual mudflats. Chung et al. (2011) tracked genetic diversity at *COI* in the amphipod *Melita plumulosa* following chemical-spill induced die-off over three years, and found the initially reduced genetic diversity at affected sites to recover to similar levels of diversity as unaffected sites by

the end of the study period. The unusually monomorphic genetic composition of OBN, DGS, and DGN despite their proximity to mudflats with greater genetic diversity may then be similarly indicative of these populations having experienced reductions in genetic diversity following recent demographic bottlenecks.

On a larger scale, the central-marginal hypothesis predicts that a species inhabiting a geographical gradient of environmental conditions should harbor greater genetic diversity where population abundance and growth is highest, and suffer decreasing variation and abundance as conditions depart from this optimum (Hengeveld and Haeck 1982; Brown 1984). Population densities, abundance, and the amount of suitable habitat are all greatest in the vast mudflats of the upper Bay of Fundy, whereas the Gulf of Maine contains smaller habitats and a larger proportion of rocky shoreline that is inhospitable to *C. volutator*. This increased separation between populations in the Gulf of Maine may allow the effects of genetic drift to be more pronounced by restricting dispersal between regions and reducing effective population sizes; supported by the (near) fixation of a single allele for *SUBI* in this region. In the Bay of Fundy, vast mudflats with stable populations are much more common and widely distributed (Barbeau et al. 2009; Drolet et al. 2012), which may account for higher observed genetic diversities in this region. The presence of all three common mtDNA haplotypes in Middle may at first appear to support this central region acting as a source of genetic variation; however, the smaller and rockier habitats in this region when compared to the inner Bay of Fundy (West and East) suggest that this is not a biologically reasonable explanation. Coalescent analyses in MIGRATE support a more parsimonious solution to this pattern: Middle is a contact zone between the Bay of Fundy and Gulf of Maine, with the

population genetic composition of populations determined by incoming gene flow from both regions and stochastic processes that limit genetic diversity.

In conclusion, the population genetic patterns evidenced by mtDNA and nDNA in this study suggest that connectivity is influenced by hydrology throughout the North American range of *C. volutator*. There is no evidence of isolation by distance within basins or across the study range, and genetic subdivision is evident between regions defined by segregating currents. These patterns are likely the result of dispersal ability enabled by swimming behaviour, facilitating high gene flow within hydrologically admixing regions, restricted gene flow between basins, and ultimately enabling the accumulation of divergence through genetic drift.

Acknowledgements. This work could not have been accomplished without samples collected by Travis Gerwing and David Drolet (University of New Brunswick). The authors wish to thank Linley Jesson (University of New Brunswick) for valuable discussion and feedback, the associate editor Cynthia Riginos, and four anonymous reviewers for their comments, which helped improve the manuscript. This work was supported by funds to the Addison laboratory from the Natural Sciences and Engineering Research Council of Canada, the Canadian Foundation for Innovation, the New Brunswick Innovation Foundation, and the University of New Brunswick.

References

Addison JA, Ort BS, Mesa KA, Pogson GH (2008) Range-wide genetic homogeneity in the California sea mussel (*Mytilus californianus*): a comparison of allozymes, nuclear DNA markers, and mitochondrial DNA sequences. Mol Ecol 17: 4222-4232

Ashton GV, Stevens MI, Hart MC, Green DH, Burrows MT, Cook EJ, Willis KJ (2008) Mitochondrial DNA reveals multiple Northern Hemisphere introductions of *Caprella mutica* (Crustacea, Amphipoda). Mol Ecol 17: 1293-1303

Aretxabaleta AL, McGillicuddy DJ, Smith KW, Lynch DR (2008) Model simulations of the Bay of Fundy Gyre: 1. Climatological results. J Geophys Res. doi: 10.1029/2007JC004480

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Barbeau MA, Grecian LA, Arnold EE, Sheahan DC, Hamilton DJ (2009) Spatial and temporal variation in the population dynamics of the intertidal amphipod Corophium volutator in the upper Bay of Fundy, Canada. J Crustac Biol 29: 491-506

Beerli, P (2006) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. Bioinformatics 22(3): 341–345

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152(2): 763–73

Bohonak AJ (1999) Dispersal, gene flow, and population structure. Quarterly Rev Biol 74:21-45

Bradbury IR, Snelgrove PVR (2001) Contrasting larval transport in demersal fish and benthic invertebrates: the roles of behaviour and advective processes in determining spatial pattern. Can J Fish Aquat Sci 58: 811-823

Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, and Campana SE (2008) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category, and life history. Proc R Soc Biol 275: 1803-1809

Brown JH (1984) On the relationship between abundance and distribution of species. Am Nat 124: 255 –279

Chung PP, Hyne RV, Mann RM, Ballard JWO (2011) Temporal and geographical genetic variation in the amphipod *Melita plumulosa* (Crustacea: Melitidae): Link of a localized change in haplotype frequencies to a chemical spill. Chemosphere 82: 1050-1055

Clement M, Posada D and Crandall K (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1660

Collin R (2001) The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Calyptraeidae). Mol Ecol 10: 2249-2262

Corander J, Marttinen P, Mäntyniemi S (2006) A Bayesian identification of stock mixtures from molecular marker data. Fishery Bulletin 104: 550-558

Corander J, Sirén J, Arjas E (2008) Bayesian Spatial Modelling of Genetic Population Structure. Computation Stat 23: 111-129

Cowen RK, Sponaugle S (2009) Larval dispersal and population connectivity. Annu Rev Mar Sci 1: 433-466

Crozier RH, Crozier YC, Mackinlay AG (1989) The COI and COII region of honeybee mitochondrial DANN: evidence for variation in insect mitochondrial evolution rates. Mol Biol Evol 6: 399-411

Davolos D, Maclean N (2005) Mitochondrial COI-NC-COII sequences in talitrid amphipods (Crustacea). Heredity 94: 81-86

Desplanque C and Mossman DJ (2001) Bay of Fundy Tides. Geosci Canada 28: 1-11

Drolet D and Barbeau MA (2009) Diel and semi-lunar cycles in the swimming activity of the intertidal, benthic amphipod *Corophium volutator* in the Upper Bay of Fundy, Canada. J Crust Biol 29: 51-56

Drolet D, Bringloe TT, Coffin MRS, Barbeau MA, Hamilton DJ (2012) Potential for between-mudflat movement and metapopulations dynamics in an intertidal burrowing amphipod. MEPS 449: 197-209

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN, v3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50

Folmer O, Black M, Hoeh W, Lutz R, and 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

Garrett C (1972) Tidal resonance in the Bay of Fundy and Gulf of Maine. Nature 238: 441-443

Goldson AJ, Hughes RN, Gliddon CJ (2001) Population genetic consequences of larval dispersal mode and hydrography: a case study with bryozoans. Mar Biol 138: 1037-1042

Grantham BA, Eckert GL, Shanks AL (2003) Dispersal potential of marine invertebrates in diverse habitats. Ecol Appl 13: 108-116

Grosberg RK, Levitan DR, Cameron BB (1996) Characterization of genetic structure and genealogies using RAPD-PCR markers: a random primer for the novice and nervous. In JD Ferraris and SR Palumbi (eds) Molecular zoology: Advances, strategies, and protocols. Wiley-Liss, New York, pp 67-100

Hellberg ME (1996) Dependence of gene flow on geographic distance in two solitary corals with different larval dispersal capabilities. Evolution 50: 1167-1175

Hengeveld R, Haeck J (1982) The distribution of abundance. 1. Measurements. J Biogeogr 9, 303–316

Hicklin PW, Smith PC (1984) Selection of foraging sites and invertebrate prey by migrant semipalmated sandpipers, *Calidris pusilla* (Pallas), in Minas Basin, Bay of Fundy. Can J Zool 62: 2201-2210

Hogg ID, Stevens MI, Schnabel KE, Chapman MA (2006) Deeply divergent lineages of the widespread New Zealand amphipod *Paracalliope fluviatilis* revealed using allozyme and mtDNA analyses. Freshwater Biol 51: 236-248

Hughes RG (1988) Dispersal by benthic invertebrates: the in situ swimming behaviour of the amphipod *Corophium volutator*. J Mar Biol Ass UK 68: 565-579

Janson K (2008) Allozyme and shell variation in two marine snails (*Littorina*, *Prosobranchia*) with different dispersal abilities. Biol J Linnean Soc 30: 245-256

Jennings RM, Shank TM, Mullineaux LS, Halanych KM (2009) Assessment of the Cape Cod phylogeographic break using the bamboo worm *Clymenella torquata* reveals the role of regional water masses in dispersal. J Heredity 100: 86-96

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

Kelly DW, MacIsaac HJ, Heath DD (2006) Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. Evolution 60: 257-267

Kelly RP, Palumbi SR (2010) Genetic structure among 50 species of the Northeastern Pacific rocky intertidal community. PLoS ONE 5: e8594

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120

Lawrie SM, Raffaelli DG (1998) In situ swimming behaviour of the amphipod *Corophium volutator* (Pallas). J Exp Mar Biol Ecol 224: 237-251

Levin LA (2006) Recent progress in understanding larval dispersal: new directions and digressions. Integr Comp Biol 46: 282-297

Librado P and Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452

Lee TN, Rooth C, Williams E, McGowan M, Szmant AF, Clarke ME (1992) Influence of Florida Current, gyres, and wind driven circulation on transport of larvae and recruitment in the Florida Keys coral reefs. Cont Shelf Res 12: 971-1002

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? Mol Ecol 19: 3038-3051

Macfarlane CB, Drolet D, Barbeau MA, Hamilton DJ, Ollerhead J (2009) Movement of drift ice as a vector for transport of sediment and invertebrates on intertidal mudflats of the Bay of Fundy. Proc CGU HS Comm River Ice Process Environ. St. John's, Newfoundland and Labrador 238-248

Manning JP, McGillicuddy DJ, Pettigrew NR, Churchill JH, Incze LS (2009) Drifter observations of the Gulf of Maine Coastal Current. Cont Shelf Res 29: 835-845

Matthaeis ED, Davolos D, Cobolli M, Ketmaier V (2000) Isolation by distance in equilibrium and nonequilibrium populations of four Talitrid species in the Mediterranean Sea. Evolution 54: 1606-1613

Meadows PS, Tait J, Hussain SA (1990) Effects of estuarine infauna on sediment stability and particle sedimentation. Hydrobiologia 190: 263-266

Mouritsen KN, Mouritsen LT, Jensen KT (1998) Change of topography and sediment characteristics on an intertidal mud-flat following mass-mortality of the amphipod *Corophium volutator*. J Mar Biol Ass UK 78: 1167-1180

Naqvi NI, Chattoo BB (1996) Development of a sequence characterized amplified region (SCAR) based indirect selection method for a dominant blas-resistance gene in rice. Genome 39: 26-30

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Proc Natl Acad Sci 70: 3321-3323

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Ann Rev Ecol Syst 25: 547-572

Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial DNA: the three-times rule. Evolution 55: 859-868

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecol Appl 13: S146-S158

Peer DL, Linkletter LE, Hicklin PW (1986) Life history and reproductive biology of *Corophium volutator* (Crustacea: Amphipoda) and the influence of shorebird predation on population structure in Chignecto Bay, Bay of Fundy, Canada. Neth J Sea Res 20: 359-373

Pettigrew NR, Churchill JH, Janzen CD, Mangum LJ, Signell RP, Thomas AC, Townsend DW, Wallinga JP, Xue H (2005) The kinematic and hydrographic structure of the Golf of Maine Coastal Current. Deep-Sea Res II 52: 2369-2391

Pritchard JK, Stephens M, and Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 945-959

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J Heredity 86: 248–249

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

Rousset F (2008) Genepop'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Rozen S and Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In S Krawertz and S Misener (eds) Bioinformatics methods and protocols: methods in molecular ecology. Humana Press, Totowa NJ, pp 365-386

Ryman N and Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology 6:600-602

Shepherd PCF, Partridge VA, Hicklin PW (1995) Changes in sediment types and invertebrate fauna in the intertidal mudflats of the Bay of Fundy between 1977 and 1994. Can Wildl Serv Technical Rep 237. Canadian Wildlife Service, Environment Canada, Ottawa.

Shoemaker CR (1947) Further notes on the amphipod genus *Corophium* from the east coast of North America. Proceedings of the Washington Academy of Sciences 37: 47-63

Slatkin M (1987) Gene flow and the geographical structure of populations. Science 236: 787-792

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Sprague AJ, Hamilton DJ, Diamond AW (2008) Site safety and food affect movements of semipalmated sandpipers (*Calidris pusilla*) migrating through the upper Bay of Fundy. Avian Conservation Ecology 3: 4

Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978-989

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76: 449-462

Tamura K., Dudley J., Nei M., and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599

Teske PR, McQuaid CD, Froneman PW, Barker NP (2006) Impacts of marine biogeographic boundaries on phylogeographic patterns of three South African estuarine crustaceans. Mar Ecol Progr Ser 314: 283-293

Teske PR, Papadopoulos I, Newman BK, Dworschak PC, McQuaid CD, Barker NP (2008) Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn. BMC Evol Biol 8: 341-355

Watson JR, Kendall BE, Siegel DA, Mitarai S (2012) Changing seascapes, stochastic connectivity, and marine metapopulations dynamics. Am Nat 180: 99-112

Weersing K and Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. Mar Ecol Prog Ser 393: 1-12

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean currents help explain population genetic structure. Proc R Soc Biol Sci 277: 1685-1694

Wilson AB, Boates JS, Snyder M (1997) Genetic isolation of populations of the gammaridean amphipod, *Corophium volutator*, in the Bay of Fundy, Canada. Mol Ecol 6: 917-923

Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97-159

Wright S (1943) Isolation by distance. Genetics 28:114–138

Table 2.1 Primers used for amplification of nuclear and mitochondrial DNA.

	Primer sequence (5' to 3')	Aligned sequence length (bp)
Mitochondrial primers		
SUBIF[3]	AAGAGGCATACCTCGACGATACTC	
COII-CROZ[2]	CCACAAATTTCTGAACATTGACC	~800
SUBIRA[1]	GCACCGTTCAAGAATGGATAA	576
LCO1490[4]	GGTCAACAAATCATAAAGATATTGG	
HCO2198[4]	TAAACTTCAGGGTGACCAAAAAATCA	496
Nuclear primers		
CV1F[1]	AGTATGCCGCTCCATACACC	
CV1R[1]	ACCTAAGGCTGACGGTTCAA	279
CV2F[1]	TTGCCAATCTTGTCTGAGCA	
CV2R[1]	CCAGATGTTAGTTCAGGGAGTG	201
CV3F[1]	GCAGATGACAAGCGGTTTCT	
CV3R[1]	GGCTTTCCTGTGACAGATGG	358
RAPD primers		
UBC-367[5]	ACCTTTGGCT	
OPA9[6]	GGGTAACGCC	
M13[7]	GTAAAACGACGGCCAGT	

[1] This study
[2] Crozier et al. 1989
[3] Davolos and Maclean 2005
[4] Folmer et al. 1994
[5] UBC Biotechnology Laboratory
[6] Operon Technologies
[7] Pharmacia

Table 2.2 Sampling sites, region, number of individuals sequenced per mitochondrial locus (*n*), number of haplotypes (*H*), number of segregation sites (*S*), haplotype diversity (*h*), and mean nucleotide diversity (π) for *Corophium volutator* populations.

				SUBI					*COI*		
Site	**Population**	**Abbr.**	**Region**	***n***	***S***	***H***	***h***	***π***	***n***	***S***	***H***
1	Avonport	AV	East	19	3	2	0.4094	0.0021	14	0	1
2	Blomidon	BL	East	34	3	2	0.4866	0.0025			
3	Moose Cove	MC	East	20	3	2	0.1895	0.0010			
4	Noel Bay	NB	East	29	4	3	0.4310	0.0021			
5	Starr's Point	SP	East	27	3	2	0.3134	0.0016	11	0	1
6	Daniel's Flat	DF	West	30	3	2	0.1862	0.0010			
7	Grande Anse	GA	West	23	3	2	0.1660	0.0009	15	1	2
8	Minudie	MN	West	37	4	3	0.1562	0.0006			
9	Mary's Point	MP	West	34	0	1	0.0000	0.0000			
10	Peck's Cove	PC	West	32	3	2	0.0625	0.0003			
11	Saint Martin's North	SMN	West	6	0	1	0.0000	0.0000	6	0	1
12	Saint Martin's South	SMS	West	4	0	1	0.0000	0.0000	4	0	1
13	Lubec	LBS	Middle	23	0	1	0.0000	0.0000	10	0	1
14	Oak Bay	OBN	Middle	13	0	1	0.0000	0.0000	6	0	1
15	Pocologan	POC	Middle	23	1	2	0.5217	0.0009	11	0	1
16	Saint John	SJS	Middle	19	3	3	0.2924	0.0012	6	0	1
17	Digby North	DGN	Outer	6	0	1	0.0000	0.0000	4	0	1
18	Digby South	DGS	Outer	6	0	1	0.0000	0.0000	4	0	1
19	Falmouth	FMS	Outer	12	0	1	0.0000	0.0000	12	1	2
20	Jonesport	JNP	Outer	9	0	1	0.0000	0.0000	9	0	1
21	Lowe's Cove	LCN	Outer	11	2	3	0.3455	0.0006	10	1	2
22	Machiasport	MAC	Outer	11	0	1	0.0000	0.0000	10	0	1
23	Milbridge	MIL	Outer	12	0	1	0.0000	0.0000	9	0	1
24	Waldoboro	WAL	Outer	12	0	1	0.0000	0.0000	11	0	1
	Total			452	7	7	0.5704	0.0023	152	3	4

Table 2.3 Nuclear DNA diversity: Sampling sites, region, number of individuals sequenced per nuclear locus (*n*), number of segregation sites (*S*), number of alleles (*H*), observed heterozygosity (H_o), expected heterozygosity (H_e), and mean nucleotide diversity (π) for *Corophium volutator* populations surveyed.

			CV1						*CV2*						*CV3*					
Population	**Abbr**	**Region**	***n***	***S***	***H***	***Ho***	***He***	***π***	***n***	***S***	***H***	***Ho***	***He***	***π***	***n***	***S***	***H***	***Ho***	***He***	***π***
Avonport	AV	East	14	4	3	0.3571	0.4418	0.0046	12	4	3	0.5000	0.6812	0.0082	14	0	1	0.0000	0.0000	0.0000
Blomidon	BL	East	15	4	4	0.5333	0.6230	0.0057	7	4	4	0.5714	0.5824	0.0089	8	0	1	0.0000	0.0000	0.0000
Moose Cove	MC	East	14	4	3	0.7143	0.6058	0.0062	8	4	3	0.6250	0.6250	0.0076	11	0	1	0.0000	0.0000	0.0000
Noel Bay	NB	East	20	4	4	0.4000	0.5500	0.0055	17	4	3	0.6471	0.6560	0.0095	20	1	2	0.0500	0.0500	0.0002
Starr's Point	SP	East	11	4	4	0.4545	0.6104	0.0060	10	4	3	0.6000	0.6368	0.0066	10	0	1	0.0000	0.0000	0.0000
Daniel's Flat	DF	West	18	4	3	0.4444	0.6095	0.0062	14	4	4	0.6429	0.6111	0.0072	15	0	1	0.0000	0.0000	0.0000
Grande Anse	GA	West	14	4	5	0.4286	0.4788	0.0044							9	0	1	0.0000	0.0000	0.0000
Minudie	MN	West	17	4	5	0.5882	0.4706	0.0044	12	4	5	0.8333	0.6667	0.0090	16	0	1	0.0000	0.0000	0.0000
Mary's Point	MP	West	13	4	4	0.5385	0.6800	0.0064	12	4	3	0.9167	0.6703	0.0093	9	0	1	0.0000	0.0000	0.0000
Peck's Cove	PC	West	18	4	3	0.3889	0.5095	0.0054	20	4	3	0.4500	0.6167	0.0086	19	0	1	0.0000	0.0000	0.0000
St Martin's N	SMN	West	3	4	3	0.6667	0.7333	0.0072	5	4	3	0.4000	0.5111	0.0048						
St Martin's S	SMS	West							2	1	2	0.5000	0.5000	0.0025						
Lubec	LBS	Middle	8	2	2	0.1250	0.1250	0.0009	8	4	4	1.0000	0.7250	0.0102	6	0	1	0.0000	0.0000	0.0000
Oak Bay	OBN	Middle	8	4	5	0.5000	0.7000	0.0054												
Pocologan	POC	Middle	15	0	1	0.0000	0.0000	0.0000	18	4	5	0.5556	0.5921	0.0060	17	0	1	0.0000	0.0000	0.0000
Saint John	SJS	Middle	14	4	3	0.5714	0.5000	0.0052	8	4	5	0.6250	0.7333	0.0094	5	0	1	0.0000	0.0000	0.0000
Digby North	DGN	Outer	4	0	1	0.0000	0.0000	0.0000	2	4	3	0.5000	0.8333	0.0100	4	0	1	0.0000	0.0000	0.0000
Digby South	DGS	Outer	2	0	1	0.0000	0.0000	0.0000							1	0	1	0.0000	0.0000	0.0000
Falmouth	FMS	Outer	7	2	2	0.1429	0.1429	0.0010	12	4	3	0.8333	0.6486	0.0068	11	0	1	0.0000	0.0000	0.0000
Jonesport	JNP	Outer	5	4	4	0.6000	0.5333	0.0040	9	4	4	0.7778	0.6797	0.0071	7	0	1	0.0000	0.0000	0.0000
Lowe's Cove	LCN	Outer	9	3	2	0.3333	0.2941	0.0032	9	4	3	0.6667	0.6209	0.0058						
Machiasport	MAC	Outer	7	4	3	0.5714	0.6703	0.0061												
Milbridge	MIL	Outer	8	3	2	0.0000	0.2333	0.0025	6	4	3	0.5000	0.5909	0.0087	8	0	1	0.0000	0.0000	0.0000
Waldoboro	WAL	Outer	8	4	3	0.2500	0.2417	0.0022							10	0	1	0.0000	0.0000	0.0000
Total			252	5	9			0.0055	191	4	7			.0084	200	1	2			0.0000

Table 2.4 Isolation by distance results from IBDWS. P-values for Mantel tests of correlation between Slatkin's similar index ($\hat{M}$) and log geographic distance for mtDNA and nDNA.

Marker	Spearman's rank correlation index (r)	P-value	Linear regression coefficient (R^2)
SUBI Global	-0.3008	0.9997	0.0905
CV1+CV2+CV3 Global	-0.0843	0.9136	0.0071
SUBI East	0.1375	0.6750	0.0189
CV1+CV2+CV3 East	0.2474	0.6440	0.0612
SUBI West	0.4048	0.9590	0.1639
CV1+CV2+CV3 West	-0.2077	0.7710	0.0432
SUBI Middle	0.5990	0.8690	0.3589
CV1+CV2+CV3 Middle	-0.4542	0.7120	0.2063
SUBI Outer	0.2384	0.9660	0.0568
CV1+CV2+CV3 Outer	-0.7122	0.9880	0.5072

Table 2.5 AMOVA for samples of *C. volutator* from oceanographic regions using mitochondrial and nuclear markers (P* < 0.0001, ***P* < 0.005, ****P* < 0.05).**

Marker	ΦSC	ΦST	ΦCT
SUBI			
E vs W+M vs O	0.4906*	0.7238*	0.4578*
E+W+M vs O	0.4742*	0.8192*	0.6561*
E vs W vs O vs M	0.2403*	0.7167*	0.6271*
E vs W	0.0114**	0.1534**	0.1436**
CV1+CV2+CV3			
E+W+M vs O	0.2096*	0.2203*	0.0135
E+W vs M vs O	0.1974*	0.2237*	0.0328***
CV1+CV2			
E+W+M vs O	0.1614*	0.1847*	0.0278
E+W vs M vs O	0.1527*	0.1808*	0.0333***

Table 2.6 Estimates of statistical power for detecting true levels of population structure (F_{ST}) by nDNA.

True F_{ST}	$k = 3$	$k = 24$
0.0000	0.0280	0.0300
0.0010	0.0843	0.0463
0.0025	0.1890	0.0820
0.0050	0.4403	0.1747
0.0100	0.8130	0.4427
0.0200	0.9720	0.8810
0.0250	0.9867	0.9583
0.0500	0.9993	0.9997
0.1000	1.0000	1.0000
0.2000	1.0000	1.0000

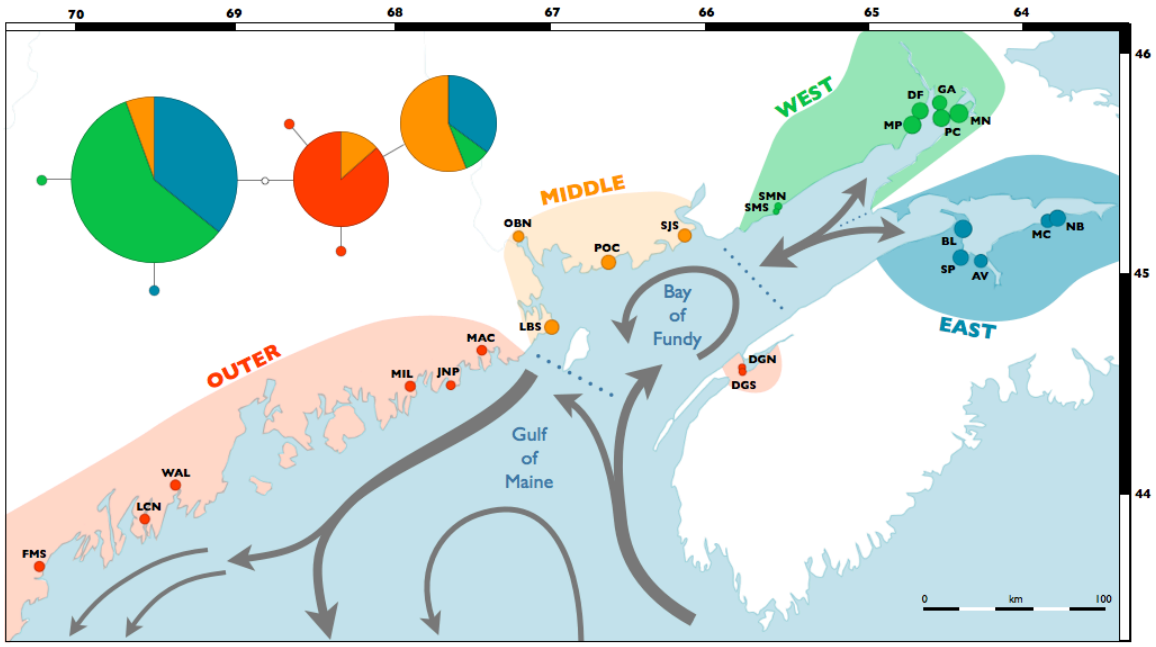


Figure 2.1 Sampling locations and regional groupings of *Corophium volutator* populations in North America (see Table 2.2). Circle area is proportional to number of individuals sampled at each location. Arrows depict circulation diagrams for the Bay of Fundy (adapted from Aretxabaleta et al. 2008) and Gulf of Maine (adapted from Pettigrew et al. 2005). Haplotype network for *Corophium volutator* mtDNA marker *SUBI.* Each symbol represents a unique sequence. Lines indicate single substitution differences between haplotypes (open symbol indicates inferred haplotype). Color code corresponds to geographic regions in which each haplotype was found. Pie chart area is proportional to haplotype frequency.

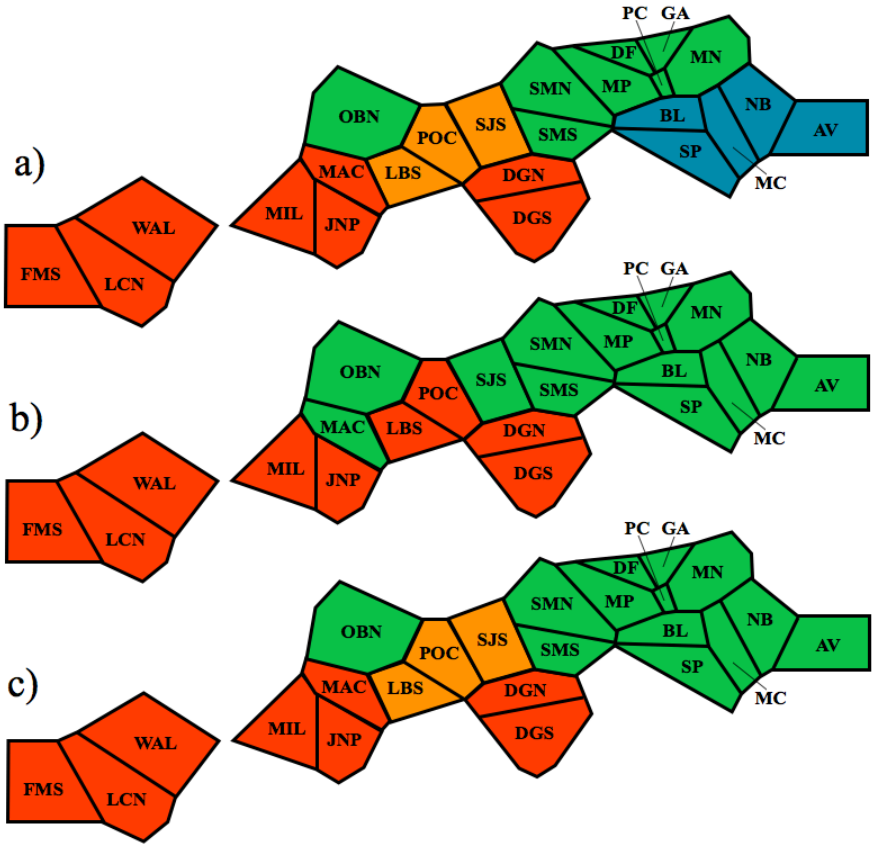


Figure 2.2 Results from Bayesian Analyses of Population Structure using clustering of groups of populations with spatial priors for a) *SUBI* mtDNA b) *CV1+CV2+CV3* nDNA c) *SUBI+CV1+CV2+CV3* combined data.

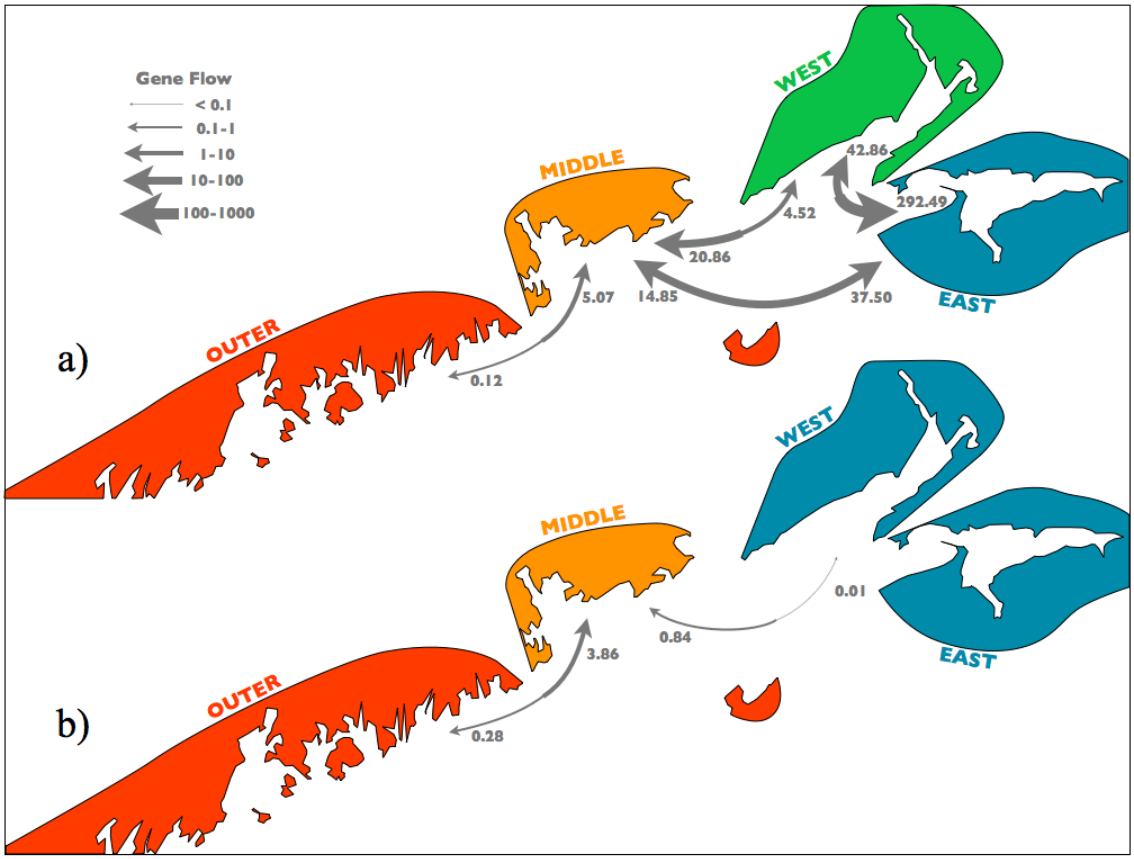


Figure 2.3 Estimates of gene flow between regions ($M=\theta m/\mu$ x 10^5) for a) *SUBI* mtDNA and b) *CV1+CV2* nDNA generated in MIGRATE. Arrow thickness corresponds to order of magnitude. Due to the high genetic similarity of East and West detected for analyses of population genetic differentiation with nDNA data, these regions were grouped for analyses of gene flow using nDNA to minimize the number of estimated parameters.

Chapter 3 – Anthropocene invasion of an ecosystem engineer: Resolving the history of *Corophium volutator* (Amphipoda: Corophiidae) in the North Atlantic[a]

Einfeldt AL[1] (corresponding author), Addison JA[2]

[a]*Published in Biological Journal of the Linnean Society, January 2015*

[1]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3. email: tony.einfeldt@gmail.com.

[2]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3

Abstract

Resolving the natural histories of species is important to interpretation of ecological patterns, as it provides evolutionary context for interactions between organisms and their environment. Despite playing an integral role on the intertidal mudflats of the North Atlantic as an abundant food source for predators and as an ecosystem engineer that alters the soft sediment environment, no previous studies have provided empirical evidence to determine the biogeographic origin of the amphipod *Corophium volutator*. To resolve its status as introduced or indigenous in Europe and North America, we analyzed sequence data for two mitochondrial loci and two nuclear markers to determine whether the present range *C. volutator* is the result of unresolved taxonomy, persistence in glacial refugia, natural trans-Atlantic dispersal, or human-mediated introduction. Our results demonstrate reduced genetic diversity in North American populations that is a subsample of diversity in European populations, with coalescent analysis of mitochondrial and nuclear DNA supporting different models of multiple introductions from Europe to the Bay of Fundy and Gulf of Maine in North America. These results suggest that *C. volutator* was introduced to North America prior to the first surveys of local biota in the 20th century, which has broad implications for interpretations of community and ecosystem interactions in the North Atlantic intertidal.

Introduction

Biological introductions are a significant driver of human-caused environmental change that can disrupt ecosystem processes such as primary productivity, decomposition, hydrology, geomorphology, nutrient cycling, and disturbance regimes (Vitousek *et al*., 1997; Ehrenfeld, 2010; Simberloff 2011; Vila *et al*. 2011). To manage and predict the future impact of introductions requires an understanding of ecosystems' natural states; however, our knowledge of what is natural is biased by the lack of both the quantity and quality of baseline data (Ruiz *et al*., 2000). This is particularly problematic in environments where primary studies of local biota lag behind introduction vectors associated with global exploration and trade, making the extent of historic introductions unknown and potentially underestimated. The common assumption that species with no documented introduction history are native has been challenged by Carlton (1996), who suggests that species which can not be reliably assigned a status of native or introduced by observational data may have been influenced by historic anthropogenic dispersal. In many cases, molecular evidence has revealed species once thought to be widespread or cosmopolitan to be the product of historic introductions or unresolved taxonomy (Geller *et al*., 2010). Resolving the biogeographic origins of species with uncertain histories is critical to understanding the impact of human activity on community ecology, natural diversity, and evolution (Grosholz, 2002; Carlton, 2003; Carlton 2009).

Marine life in the North Atlantic has a particularly suspect natural history, as extensive exploration and trade between Europe and the new world predated biological surveys by centuries (Carlton, 2003). Species that are indigenous to the North American

Atlantic had to either persist in glacial refugia during Pleistocene glaciations ~20,000 years ago or colonize the North American coast via long distance dispersal from Europe, where glaciations were less severe (Hewitt, 2000; Wares & Cunningham, 2001). Although natural long-distance dispersal across the Atlantic is hindered by vast expanses of open water, mid-Atlantic islands such as Iceland and Greenland provided stepping-stones between Europe and North America for species capable of larval dispersal or rafting via ocean currents (reviewed in Ingolfsson, 1992). A disjunct range (present in Europe and America but not in between) is inconsistent with the stepping-stone pattern expected to result from natural dispersal, and can be explained by one of four models put forth by Haydar (2012): *a)* persistence in glacial refugia on both coasts with no post-glacial colonization of intermediate islands; *b)* unresolved taxonomy with morphologically similar but genetically distinct species on each coast; *c)* natural trans-oceanic dispersal without the colonization of intermediate islands; and *d)* human-mediated introduction. To our knowledge, no studies have yet found evidence for models *a* or *c* in species with a disjunct distribution in the North Atlantic, suggesting that species with similar distributions may be either introduced or phylogenetically unresolved.

In the absence of direct paleontological, archaeological, or observational evidence, molecular tools can be used to explicitly test models of population history. Within species, genetic data may be sufficient to rule out one or more hypotheses, but do not always provide the ability to distinguish between competing models (e.g. Haydar *et al*., 2011). If a disjunct distribution is the result of persistence on both coasts during the Pleistocene glaciations, each coast should harbor one or more monophyletic lineages as a result of restricted gene flow following vicariance (Haydar, 2012). In this circumstance,

colonization from southern glacial refugia to northern regions is expected to cause serial founder events that result in decreasing genetic diversity as a function of distance from refugial populations (e.g. Vainio & Vainola, 2003; Maggs *et al*., 2008; Provan & Bennett, 2008), although the level of diversity on either coast could vary depending on the severity of bottlenecks. Natural trans-Alantic dispersal could result in either a pattern of uniform outward range expansion if dispersal arose from a rare event or admixture amongst colonized populations in the presence of ongoing gene flow. While both natural models predict a signal of uniform expansion from the source populations, anthropogenic dispersal has the potential to facilitate multiple introductions followed by the admixture of genetically distinct lineages (e.g. Brawley *et al*., 2009; Blakeslee *et al*., 2010). Introduced species tend to undergo more severe bottlenecks than species that colonize new habitat naturally (Nei *et al*., 1975; Sakai *et al*., 2001; Hanfling *et al*., 2002), and different introduction processes can lead to a variety of genetic patterns in introduced populations (Voisin *et al*., 2005). Using molecular tools to differentiate between these patterns can provide insight into evolutionary relationships in communities and ecosystems in which introduction vectors predate historical observations.

While many introduced species in the Northwest Atlantic intertidal have altered community dynamics through direct interactions (e.g. *Styela clava, Botrylloides diegensis,* and *Membranipora membranacea*, Berman *et al*., 1992; *Mytilopsis leucophaeata* and *Ostrea edulis*, Carlton, 1999; *Carcinus maenas,* Carlton & Cohen, 2003; *Littorina littorea* and *Fucus serratus*, Brawley *et al*., 2009), some introduced species that alter the physical environment (e.g. *Hediste diversicolor,* Einfeldt *et al*. 2014) have the potential to cause change to the entire ecosystem. The littoral amphipod

Corophium volutator (Pallas 1766) is an ecosystem engineer that affects environmental conditions for all other members of the soft sediment community by constructing and moving water through U-shaped burrows, altering substrate stability, permeability, water content, nutrient cycling, and oxidation rates (Meadows & Tait, 1989; Meadows *et al*., 1990; Mouritsen *et al*., 1998). While important as an abundant and nutritious prey item for many species of economically valuable fish (McCurdy *et al*., 2005), migratory shorebird (Peer *et al*., 1986), and other invertebrates (Pihl & Rosenberg, 1984), its non-trophic impact may overshadow these interactions by influencing ecosystem structure and function via niche construction (Widdows & Brinsley, 2002).

Despite *C. volutator*'s central role in the intertidal community, there is no direct observational or fossil evidence to indicate its biogeographic origin in the North Atlantic. Although listed as amphi-Atlantic by Bousfield (1973), no studies have found *C. volutator* in Iceland or Greenland, the stepping-stone landmasses between Europe and North America. This disjunct distribution is not likely the result of natural dispersal; due to a brood-rearing life history, *C. volutator* has a limited dispersal potential that is further restricted by current boundaries, causing strong patterns of genetic differentiation along coastlines at small spatial scales (Wilson *et al.*, 1997). In North America, hydrological patterns that create barriers to dispersal have been attributed to maintaining strong genetic subdivision of *C. volutator* between the Bay of Fundy and the Gulf of Maine (Einfeldt & Addison, 2013). A similar pattern of genetic subdivision is exhibited by the polychaete *Hediste diversicolor,* which was introduced to North America from European populations and commonly occurs with *C. volutator* on both sides of the Atlantic (Einfeldt *et al*., 2014). Despite suggestions that *C. volutator* may have been introduced to North America

(Chapman, 1999; Carlton and Ruiz, 2005) that conflict with Wilson *et al.*'s (1997) statement that preliminary (and unpublished) genetic investigations show a long period of separation between the Bay of Fundy and European populations, the hypothesis that *C. volutator* was introduced to North America has not been empirically tested. Other contentions over *C. volutator'*s biogeography have been resolved as more detailed morphological information within the genus became more readily available, reducing chances of misidentification. *C. volutator* was thought to be introduced in the Mediterranean (Crawford, 1937; Segerstrale, 1959), but since Bellan-Santini *et al.* (1982) called for confirmation of its presence, no reliable records of *C. volutator* have been presented from this area.

Here, we use DNA sequence data from mitochondrial and nuclear markers to test models of population history for the ecosystem engineer *C. volutator*. To determine whether its disjunct distribution in the North Atlantic is the result of unresolved taxonomy, glacial persistence, natural trans-oceanic colonization, or anthropogenic introduction, we compare patterns of diversity between North American and European populations and test whether North American populations result from single or multiple colonization events.

Methods

Sample collection, DNA extraction, amplification, and sequencing

Samples of *C. volutator* were collected from 16 sites spanning the species' range on the Atlantic coast of Europe and shipped in 24% ethanol to Canada. Specimens were stored in 95% ethanol at -20°C prior to DNA extraction and sequencing.

DNA was extracted from specimens of *C. volutator* using a cetyltrimethylammonium bromide (CTAB) protocol (Doyle & Doyle, 1987), with remaining tissue stored in 95% ethanol at -80 °C as a voucher. Eggs and larvae were excised from the brood pouches of gravid females prior to DNA extraction to avoid amplification of nuclear DNA from multiple individuals.

To investigate phylogeographic patterns in *C. volutator* and compare them to those found in species with known introduction histories and published mitochondrial sequence data, we amplified two mitochondrial DNA fragments: the 5' end of cytochrome *c* oxidase subunit I (*COI*) and a fragment spanning a 143 bp portion of the 3' end of *COI,* an intergenic spacer, and 376 bp part of the 5' end of *COII* (*SUBI*). As inference based on a single matrilineal gene tree can potentially provide a biased view of an organism's evolutionary history, we also sequenced the anonymous unlinked nuclear loci *CV1* and *CV2*. Amplification for all markers was done using primers and protocols described in Einfeldt & Addison (2013).

Amplicons were checked on 1% agarose gels, sequenced on Applied Biosystem's 3730xl DNA Analyzer technology using forward, reverse, or both primers at Genome Québec Innovation Centre (Montréal, Québec), and aligned, edited, and trimmed to standard length in Sequencher 5.0 (Gene Codes). We combined these new European data with the North American data set analyzed by Einfeldt & Addison (2013; GenBank

accessions #JX991308-JX992554).

Polymorphism and phylogenetic analysis

The disjunct range of *C. volutator* in the North Atlantic can be explained by one of four hypotheses, which have different predictions for overall levels of polymorphism and phylogenetic relationships between alleles in the Northwest Atlantic (NWA) and Northeast Atlantic (NEA) (Table 3.1). If this range is the result of a taxonomic artefact, divergence at the *COI* marker between sequences from populations in the NWA and the NEA should be greater than 0.16 substitutions per site, which is indicative of species level differentiation at this locus in crustaceans (Lefébure *et al.,* 2006). Reciprocal monophyly with high diversity at all loci in the NWA and the NEA would suggest that *C. volutator* persisted in glacial refugia on both coasts. Intermediate or high diversity in the NWA at all loci with alleles identical to or descended from those found in the NEA would indicate natural trans-oceanic dispersal from the NEA to the NWA. In contrast, a pattern of low diversity in the NWA with alleles being a subsample of those found in the NEA but non-monophyletic would suggest a recent and likely human-mediated introduction.

Allelic phases of nDNA were inferred using PHASE v2.1.1 (Stephens *et al*., 2001; Stephens & Scheet, 2005) for 10,000 iterations after a burn-in of 1,000, with all haplotype reconstructions included in analyses having a posterior probability $P>0.95$ consistently over 3 runs. Statistical parsimony implemented in TCS v1.21 (Clement *et al.,* 2000) was used to map haplotype networks of mtDNA sequences. We calculated measures of genetic diversity in European populations to compare to data published for

North American populations (Einfeldt & Addison, 2013). Haplotype and nucleotide diversity was calculated using DNASP v5.1 (Librado & Rozas, 2009), and nuclear markers were confirmed to be unlinked using GENEPOP v1.2 (Raymond & Rousset, 1995; Rousset, 2008). Observed and expected heterozygosities and fixation indices were calculated using ARLEQUIN v3.11 (Excoffier *et al*., 2005), with Fu and Li's *F* and Tajima's *D* calculated to test for departures from neutrality based on allelic states or segregating sites, respectively. The significance of these neutrality tests was assessed by performing 10,000 bootstrap replicates.

Because the effects of sampling may lead to an underestimation of genetic diversity, we used rarefaction curves to estimate the total diversity present in the NWA and the NEA. ESTIMATE-S (Colwell *et al.*, 2012) was used to generate haplotype estimation curves through Monte Carlo resampling and randomization of sample order over 10,000 replicates. To account for the effects of rare haplotypes on total diversity, we employed the robust non-parametric Chao-2 indicator to estimate the number of haplotypes or alleles present in the NEA and NWA for each marker (Walther & Morand, 1998; Gotelli & Colwell, 2001; Foggo *et al*., 2003).

Colonization model selection

We found that *C. volutator's* genetic diversity in the NWA is a subsample of diversity found in the NEA, suggesting that populations in the NWA were colonized from the NEA. However, it is not obvious whether colonization of the three genetically subdivided regions in the NWA described by Einfeldt & Addison (2013) resulted from multiple introductions from the NEA or from dispersal along the coast following trans-

Atlantic colonization to one region. Einfeldt & Addison (2013) report higher genetic diversity in the central region of the NWA, suggesting it may be a contact zone between them. We tested six models of connectivity between the NEA and the NWA that reflect simple population genetic hypotheses and competing introduction histories consistent with either natural or anthropogenic dispersal (Figure 3.2):

1) 'Single introduction pathway': asymmetric migration from the NEA to the middle region of the NWA with symmetric migration between adjacent regions in the NWA.

2) 'Two introduction pathways': asymmetric migration from the NEA to the northern and southern regions of the NWA with symmetric migration between adjacent regions in the NWA.

3) 'Three introduction pathways': asymmetric migration from the NEA to all regions of the NWA with symmetric migration between adjacent regions in the NWA.

4) 'Two populations': all samples in the NWA are of the same population, with asymmetric migration from the NEA to the NWA.

5) 'Open': symmetric migration between the NEA and the three NWA regions.

6) 'Panmixia': all samples are of the same population.

Models 5 and 6 provide simple models for comparison with other colonization models. Models 1and 4 are consistent with either natural or anthropogenic trans-Atlantic dispersal, and thus provide little information regarding when colonization occurred. Models 2 and 3 reflect multiple introductions, which would require successful colonization of different areas within a suitably narrow timeframe for trans-Atlantic gene flow to outweigh local recruitment; a pattern that is unlikely to occur via natural trans-Atlantic dispersal, and is thus indicative of anthropogenic introduction.

To test which colonization models have the highest probability given the genetic data from *C. volutator* in the north Atlantic, we employed a model selection approach (Johnson & Omland, 2004; Beerli & Palczewski, 2010) to calculate the relative posterior probability of the six models using Bayesian inference under a coalescent framework with the program MIGRATE-N v3.4.4 (Beerli & Felsenstein, 2001). We translated the six models into migration matrices and estimated posterior values of migration and theta using a Bayesian MCMC search strategy using initial parameter values generated from an F_{ST} calculation. We examined effective sample sizes for each parameter to assess stationarity of the Markov chains. Following preliminary trials, we applied a static heating scheme with 6 different temperatures for 10 concurrent (replicate) long chains totaling 40,000,000 steps, with each chain having 40,000 steps recorded over intervals of 100 steps after a burn-in of 10,000 generations. Each model was rerun for a total of 10 replicates to check for consistency of parameter estimates and model marginal log-likelihood (lmL). To evaluate support for each model, we calculated the mean Bezier thermodynamic approximation of lmL of each model, from which we calculated the natural log Bayes factors (LBF) with respect to the model with the highest mean Bezier lmL using the method presented by Beerli & Palczewski (2010). Model choice probabilities were calculated following Kass & Raftery (1995). To check for consistency of parameter estimates, the best model for each data set was rerun for three times more replicates (30) and five times more recorded steps (20,000).

Results

Polymorphism and diversity

We found 25 variable positions (24 transitions and 1 transversion) over 496 sequenced nucleotides of *COI* (GenBank accessions: KP683379 - KP683563), of which 80% (20/25) were synonymous. In *SUBI* (GenBank accessions: KP683564 - KP683770), we found 29 variable positions (24 transitions and 5 transversions) over 576 sequenced nucleotides, with 93% (13/14) of substitutions in the *COI* region being synonymous and 87% (13/15) of substitutions in the *COII* region being synonymous. In *CV1* (GenBank accessions: KP683771 - KP683901), we found 16 variable positions (7 transitions and 9 transversions) over 279 nucleotides. In *CV2* (GenBank accessions: KP683902 - KP684029), we found 12 variable positions (4 transitions and 8 transversions) in 201 nucleotides. Tajima's *D* was not significantly different from zero for *COI* (0.2199, $p > 0.1$), *SUBI* (1.17273, $p > 0.1$), *CV1* (-0.8868, $p > 0.1$), and *CV2* (-0.7734, $p > 0.1$). Fu and Li's *F* was not significantly different from zero for *COI* (0.5241, $p > 0.1$), *SUBI* (0.5657, $p > 0.1$), *CV1* (0.3470, $p > 0.1$), and *CV2* (-1.4484, $p > 0.1$). Mitochondrial DNA sequence data were concatenated into the marker *COI/SUBI* for all further analyses.

Corophium volutator exhibited lower genetic diversity in North America than in Europe, with common mtDNA (*COI/SUBI*) haplotypes and nDNA (*CV1* and *CV2*) alleles in the NWA representing a non-monophyletic subsample of those found in the NEA (Figure 3.1, Table 3.2). We observed a total of 39 haplotypes for concatenated *COI*/*SUBI* mtDNA with 3 common haplotypes in the NWA shared with the NEA, 5 singleton haplotypes unique to the NWA (present at the North and South range edges), and 26 haplotypes unique to the NEA. The NWA had significantly fewer haplotypes than the NEA ($\chi^2 = 16.1$, d.f. = 1, $p < 0.001$). The nuclear marker *CV1* had 34 alleles, with 7 NWA alleles shared with the NEA, 2 singleton alleles unique to the NWA, and 25 alleles unique to the NEA. The nuclear marker *CV2* had a total of 16 alleles, with 4 NWA alleles shared with the NEA, 1 allele unique to the NWA, and 11 alleles unique to the NEA. Both *CV1* and *CV2* had significantly fewer alleles in the NWA compared to the NEA (*CV1*: $\chi^2 = 12.9$, d.f. = 1, $p < 0.001$; *CV2*: $\chi^2 = 37.7$, d.f. = 1, $p < 0.001$).

To evaluate the thoroughness of our sampling, we estimated the expected number of haplotypes in the NEA and the NWA using the Chao2 parameter (Table 3.3). For concatenated mtDNA, we found that the expected maximum number of haplotypes in the NWA was 17 (95% CI – 10-57); an order of magnitude less than the 766 (95% CI – 79-767) expected in the NEA. For *CV1*, the expected maximum number of haplotypes in the NWA was 19 (95% CI – 11-59); an order of magnitude less than the 297 (95% CI – 79-1523) expected in the NEA. For *CV2*, the expected maximum number of haplotypes in the NWA was 5 (95% CI did not deviate from this value); considerably less than the 18 (95% CI – 15-34) expected in the NEA. These results suggest that our sampling was sufficient to capture most genetic diversity in the NWA, but additional sampling of populations in the NEA would reveal considerably more genetic variation for all three loci.

Colonization model selection

Parameter estimates of MIGRATE-N analyses using mtDNA and nDNA reached convergence, with little to no variation shown when replicating analyses for 3 times more replicates or 5 times more recorded steps. In every run, effective sample sizes for each parameter were greater that 1000. Independent colonization model testing using Bayes factors calculated from marginal log-likelihood outputs for both data sets supported models of multiple colonization pathways, but the models with the highest probability were different for each marker type (Table 3.4). Analyses of mtDNA (*COI*/*SUBI*) supported a model of two colonization pathways to the NWA from the NEA (LBF = 0.0; probability = 0.8122) over a single colonization model (LBF = -4.5; probability = 0.1779) and all other models (probabilities < 0.01). Analyses of nDNA (*CV1* and *CV2*) supported a model of three introduction pathways from the NEA (LBF = 0.0; probability = 0.9277) over a model of symmetric migration between the three regions of the NWA and the NEA (LBF = -5.1; probability = 0.0720) and all other models (probabilities < 0.01). For both data sets, the panmixia model was the poorest fit, and the model of two populations with asymmetric migration to the NWA from the NEA was the second least supported.

The rank order of models 1, 2, 3, and 5 differed between data sets; although the top 3 models for mtDNA represent colonization from the NEA to the NWA, this was not the case for nDNA. While the faster rate of evolution of mtDNA can lead to geographically structured phylogenies without corroboration of nuclear gene sequences in populations with relatively short periods of isolation (Hare, 2001; Zink & Barrowclough, 2008), this does not fully explain the disagreement between our analyses. These differences highlight the importance of considering multiple markers and marker types in reconstructing population histories, and suggest further investigation using additional loci may better resolve colonization pathways followed by *C. volutator* from the NEA to the NWA.

Discussion

The reduction of genetic diversity and pattern of shared ancestry between disjunct populations of *Corophium volutator* in the north Atlantic suggest this species was recently introduced from Europe to North America. Genetic diversity of both nuclear and mitochondrial markers was significantly reduced in the NWA, and common alleles in the NWA represented a nested subset of diversity found in the NEA, indicating that populations in the NWA are descended from ancestors in the NEA. Our coalescent analysis of nuclear DNA sequences strongly support a model of unidirectional gene flow following multiple colonization pathways from the NEA the NWA, supporting the hypothesis that populations of *C. volutator* in the Bay of Fundy and the Gulf of Maine were independently established from European populations. This pattern of multiple introductions is unlikely to have occurred via natural dispersal, suggesting that shipping practices during the Age of Discovery played a key role in establishing populations in North America. This conclusion challenges assumptions about community and ecosystem level interactions in the soft sediment intertidal of the NWA, as the relationships between *C. volutator,* its predators, competitors, and the environment are likely the product of a much more recent evolutionary history than previously thought.

Shared genetic diversity

Introduced populations typically experience strong genetic bottlenecks, with introduced diversity representing a subsample of that found in the species' native range (e.g. Bastrop *et al*., 1998; Slade & Moritz, 1998; Ashton *et al*., 2008; Audzijonyte *et al*., 2008; Darling *et al*., 2008; Dlugosch & Parker, 2008; Lejeusne *et al*., 2011; Slothouber Galbreath *et al*., 2010). While introductions from multiple sources can alleviate these bottlenecks, these mixed populations still exhibit a pattern of subsampling, with alleles identical to or closely descended from those found in more phylogenetically diverse indigenous populations (Geller *et al*., 1997; Frankham, 2004; Hauber *et al*., 2011; Lacoursière-Roussel *et al*., 2012; Cahill & Viard, 2014). In contrast, natural post-glacial trans-Atlantic dispersal is expected to result in a distribution including the intermediate landmasses of Iceland and Greenland and one or more monophyletic lineages on each side of the Atlantic (e.g. Krebes *et al.,* 2011; Panova *et al*., 2011; Palsson *et al*., 2014). If *C. volutator* was recently introduced to the NWA from the NEA, we expect genetic diversity in the NWA to be a subsample from more phylogenetically diverse populations in the NEA. Consistent with this prediction, we found that all genetic diversity in the NWA was identical to or recently descended from that found in the NEA (Figure 3.1). All concatenated mtDNA haplotypes (*COI/SUBI*) and nDNA alleles (*CV1* and *CV2*) for which more than a single copy was found in the NWA were shared with the NEA, with the exception of a single *CV2* allele for which 6 copies were detected in the NWA only. The 5 concatenated mtDNA haplotypes and 2 *CV1* alleles detected only in the NWA were found in only one instance. Our rarefaction analyses show that the expected number of alleles in the NEA is more than an order of magnitude greater than in the NWA for concatenated mtDNA and *CV1*, and 3.5x greater for *CV2* (Table 3.3), suggesting that we have sampled the NWA adequately enough to capture all common variants but that a considerable amount of unsampled diversity exists in the NEA. The genetic diversity found in the NWA populations of *C. volutator* thus represents a subset of diversity in the NEA, with a small number of exceptions that could be an artifact of the limitations of sampling diverse indigenous populations. These findings are consistent with genetic studies of species with documented introductions; due to the practical limits of

completely sampling diverse indigenous populations, it is common for unique diversity to be found in introduced ranges (e.g. Dawson *et al.,* 2005; Miura *et al*., 2006; Roman, 2006; Blakeslee *et al*., 2008).

Colonization models

Analyses of gene flow using mitochondrial and nuclear DNA support different models, but both marker types provide evidence of unidirectional colonization following multiple colonization pathways from the NEA to the NWA. Natural trans-Atlantic dispersal is expected to facilitate bilateral gene flow between both coasts (e.g. *Asterias rubens*: Harper *et al.*, 2007); however, the model of symmetric gene flow between populations was not the highest ranked by analyses using either marker type. Although natural dispersal could potentially result in unidirectional colonization if driven by predominant currents, southern habitat that first became available during glacial recession would likely be colonized first, with serial founder events from northwards range expansion along the NWA predicted to cause a pattern of decreasing genetic diversity in more northern populations. Contrary to this expectation, analyses using both concatenated mtDNA and nDNA show no evidence of southern refugia and rank models consistent with natural dispersal via a single colonization pathway lower than one of two models reflecting multiple colonization pathways.

While each marker type supports a different colonization model, the models ranked highest by analyses of nDNA and mtDNA are consistent with multiple introduction pathways from the NEA to the NWA. Analyses using mtDNA support a model of two introduction pathways, to the Bay of Fundy and to the Gulf of Maine, with a contact zone between the two regions. In contrast, analyses using nDNA support a model of separate introduction pathways to the Bay of Fundy, the Gulf of Maine, and the region where they meet. For natural colonization to produce a pattern of multiple introduction pathways, colonization to each region in the NWA would have to occur within relatively similar time frames for migration across the Atlantic to outweigh along-shore dispersal rates of early colonists, which would favour the success of the first

lineages to colonize the NWA. While a scenario of multiple colonization events is not impossible via natural trans-Atlantic dispersal, it is much more likely to occur via human-mediated dispersal (Geller *et al*., 2010).

Dispersal mechanisms and introduction vectors

Natural long-distance dispersal by marine invertebrates can occur via planktonic larval dispersal (Thorson, 1950; Scheltema, 1971) and rafting on ice, seaweed, or other debris (Highsmith, 1985). The brood-rearing life history of *C. volutator* rules out planktonic dispersal as a potential mechanism of dispersal. *C. volutator* lives in close association with sediment, and has not been observed rafting on seaweed (Ingolfsson, 1995), but can survive freezing in ice blocks that facilitate rafting (Macfarlane *et al.,* 2013). However, this mechanism would have directed dispersal from Europe to North America along predominant currents that pass the intermediate landmasses of Iceland and Greenland. Despite soft-sediment invertebrates that commonly co-occur with *C. volutator* in Europe being found in Iceland and Greenland (Ingolfsson, 1996), *C. volutator* is not found on either intermediate land mass. *C. volutator* experiences linear population decreases during winter and higher rates of mortality in colder temperatures, likely due to the high energetic costs maintaining a low freezing point of body fluids (Drolet *et al*., 2013), making ice rafting an unlikely vector for natural dispersal of *C. volutator* from the NEA to the NWA. Known mechanisms of natural trans-Atlantic dispersal thus do not adequately explain the disjunct distribution of *C. volutator* across ~5000km of open ocean.

The pattern of multiple colonization events from the NEA to the NWA is consistent with human-mediated transport during early European exploration and trade in the New World. Although subject to more trade and exploration by Europeans much earlier than the Pacific Coast of North America, the NWA has only half as many documented invasions, which is partially attributable to temporal biases in data availability (reviewed by Ruiz *et al*., 2000; Carlton, 2003). Shipping pressure from European explorers, traders, and colonists intensified following the establishment of

settlements during the first decade of the 1600s in the Gulf of Maine and Bay of Fundy, but the first recorded biological surveys began in the late 1600s (Denys, 1672) and focused mainly on conspicuous terrestrial vertebrate and plant species. *C. volutator* was not first described until 1766 by Pallas on the coast of Norway, and was first recorded in North America by Huntsman & Sparks (1924), who noted its high abundance and unknown distribution in the NWA. This delay between initial shipping pressure and scientific surveys that took note of invertebrates in littoral soft sediment habitat may have allowed *C. volutator* to be introduced and proliferate in the NWA without these events being observed directly.

Shipping is the most common introduction vector in marine systems, and has been responsible for the movement of organisms mainly via ballast water or fouling on hulls and anchors (reviewed in Ruiz *et al*., 2000). Despite their tendency to dwell in burrows, adult *C. volutator* swim during flood tide (Drolet & Barbeau, 2009) and were found in a ballast tank in Norway (Gollasch *et al.,* 2002). However, it is not known how common this occurrence is; a survey of macro-invertebrates in 62 boats arriving in North American ports found no *C. volutator* in any of 67 ballast tanks (Briski *et al*., 2012). Although ballast water is a potential vector for anthropogenic dispersal, it was likely not adopted early enough to account for the high abundances of *C. volutator* observed in North America during the early 1900s. Prior to the adoption of ballast tanks in 1880, ships controlled buoyancy with semi-dry ballast: rock and sediment loaded into damp holds that was collected from heaps in the intertidal at different tidal heights, enabling global introductions of benthic marine invertebrates (Bax *et al*., 2003; Minchin *et al.*, 2009). *C. volutator'*s preferred habitat in the littoral soft sediment of estuaries and river mouths that provided safe harbor for vessels in the Atlantic would likely have enabled inadvertent movement of many individuals along with this substrate. While modern ballast tanks enable contemporary transport of *C. volutator*, semi-dry ballast is a more likely vector for the initial introduction of *C. volutator* from the NEA to the NWA, and may have had a greater influence on the soft sediment community than previously suspected.

Implications for ecology

Whether a species is introduced or indigenous across its range has repercussions for studies of local population dynamics, trophic and community interactions, and evolution. Introduced species are often considered threats to marine biodiversity and as harmful to indigenous populations (Elton, 1958; Vitousek *et al*., 1997; Ruiz *et al*., 1999; Chapin *et al*., 2000; Bax *et al*., 2003); particularly when these invaders are ecosystem engineers (Crooks, 2002; Cuddington & Hastings, 2004). In the Bay of Fundy, *C. volutator* has been viewed as an integral member of the soft sediment ecosystem and populations are used as an index of ecological health (Hawkins, 1985; Murdoch *et al*., 1986); a perspective that is at odds with how invasive ecosystem engineers are commonly viewed. A history of anthropogenic introduction challenges the notion that the stability of this species alone is informative to overall ecosystem function and health, and calls to question previous interpretations of ecosystem dynamics. Our results suggest that contemporary ecological studies of the mudflats in the Bay of Fundy and Gulf of Maine should consider that this system might still be recovering from or adapting to a shift in community structure following the introduction of *C. volutator*.

Resolving the history of seemingly important prey items provides valuable evolutionary context to ecological studies, as using abundance as a metric of value overlooks the potential flexibility of predators' responses to a changing community structure. While it is not known whether *C. volutator* displaced more common invertebrates or provided a new energetic contribution to the soft sediment ecosystem, there is some evidence that *C. volutator* is a preferred prey item simply due to its widespread availability. The migratory semipalmated sandpiper *Calidris pusilla* (Linnaeus 1766) was thought to be dependent on *C. volutator* to provide sufficient energy for annual migrations by doubling its weight while feeding on mudflats in the Bay of Fundy (Hamilton *et al*., 2003). Contrary to the assumption that *C. volutator* is a necessary component of semipalmated sandpipers' diet, a more recent study by the same research group suggests that these birds may consume different prey items based on availability (MacDonald *et al*., 2012). Contrasting findings of prey preference are also found for the Atlantic sturgeon *Acipenser oxyrhinchus oxyrhinchus* (Mitchill 1815) in the Bay of

Fundy. Pearson *et al*. (2007) report targeted feeding on aggregations of *C. volutator*, while Maclean *et al*. (2013) found that sturgeon have a strong preference for polychaetes. Both the semipalmated sandpiper and sturgeon have been present in the NWA longer than we suggest *C. volutator* has been, implying that any apparent reliance on *C. volutator* may be a response to changes in prey availability following *C. volutator*'s arrival. These shifts in diet may have consequences for predators beyond nutritive value; elevated concentrations of mercury in semipalmated sandpipers associated with feeding in the Bay of Fundy (Didyk *et al*., 2005) is thought to be influenced mostly by polychaete worms that are at a higher trophic level and have greater bioaccumulation factors than *C. volutator* (Sizmur *et al*., 2013).

In addition to direct interactions, the introduction of an ecosystem engineer can indirectly affect indigenous biota by physically altering the environment (reviewed by Wallentinus & Nyberg, 2007; Mermillod-Blondin & Rosenberg, 2011). Environmental alterations may be beneficial for some species and detrimental to others, making it difficult to assess the impact of introduced species on local biota. *C. volutator* significantly alters the physical environment by reducing soft-sediment erosion (Mouritsen *et al*., 1998) and sedimentation of re-suspended particles through secretions used to construct burrows (Meadows *et al*., 1990), increasing turbidity through the construction of these burrows (De Deckere *et al*., 2000; Biles *et al*., 2002), and increasing CO_2/O_2 fluxes and denitrification through burrow ventilation (Pelegri & Blackburn, 1994). Bioturbation caused by *C. volutator* results in dramatic increases in the aqueous concentration of sediment-bound contaminants, which become subsequently accumulated by filter-feeders (Ciarelli *et al*., 1999); a potential concern for shellfish fisheries. *C. volutator* negatively impacts the successional development of salt marsh vegetation by preventing the establishment of seedlings (Gerdol & Hughes, 1993), decreasing the area covered by salt marshes and expanding soft-sediment habitat. Within mudflats, the presence of *C. volutator* can negatively impact the densities of other infaunal species (Commito, 1982; Hughes & Gerdol, 1997) and increase diatom species richness by reducing biomass of dominant taxa in biofilm (Hagerthey *et al*., 2002). These significant impacts on the soft sediment ecosystem fundamentally alter habitat qualities, and may

have potentially led to a restructuring of the soft sediment infauna when *C. volutator* became established in the NWA.

Acknowledgements

We thank the researchers who provided samples from the NEA: G Bachelet, M Barbeau, A Barille, J Jourde, C Dancie, J Sigwart, R Hughes, M Solan, A Debacker, S Wijnhoven, I Heisterkamp, B Hussel, KT Jensen, J Strand, and A Bick. We thank the associate editor JA Allen, JT Carlton, and an anonymous reviewer for their comments, which helped strengthen the manuscript. This research was supported by grants from the Fredrik and Catherine Eaton Fellowship (to ALE); the Marguerite and Murray Vaughan Fellowship (to ALE); the New Brunswick Wildlife Trust Fund (to JAA and ALE); and grants to JAA from the Canadian Foundation for Innovation, the New Brunswick Innovation Foundation, and the Natural Science and Engineering Research Council Discovery Grants Program.

References

Ashton GV, Stevens MI, Hart MC, Green DH, Burrows MT, Cook EJ, Willis KJ. 2008. Mitochondrial DNA reveals multiple Northern Hemisphere introductions of *Caprella mutica* (Crustacea, Amphipoda). *Molecular Ecology* **17**: 1293–1303.

Audzijonyte A, Ovcarenko I, Bastrop R, Väinölä R. 2008. Two cryptic species of the *Hediste diversicolor* group (Polychaeta, Nereididae) in the Baltic Sea, with mitochondrial signatures of different population histories. *Marine Biology* **155**: 599–612.

Bax N, Williamson A, Aguero M, Gonzalez E, Geeves W. 2003. Marine invasive alien species: a threat to global biodiversity. *Marine Policy* **27**: 313-323.

Bastrop R, Jürss K, Sturmbauer C. 1998. Cryptic species in marine polychaete and their independent introduction from North America to Europe. *Molecular Biology and Evolution* **15**: 97–103.

Beerli P, Palczewski M. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**: 313-326.

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA* **98**: 4563-4568.

Bellan-Santini D, Karaman G, Krapp-Schickel G, Ledoyer M, Myers A, Ruffo S, Schiecke U. 1982. Part I. Acanthonotozomatidae to Gammaridae. In: Ruffo S, ed. *The amphipoda of the Mediterranean*. Monaco: Mém inst océanogr, 13: 1-364.

Berman J, Harris L, Lambert W, Buttrick M, Dufresne M. 1992. Recent invasion of the Gulf of Maine: three contrasting ecological histories. *Conservation Biology* **6**: 435-441.

Biles CL, Paterson DM, Ford RB, Solan M, Raffaelli DG. 2002. Bioturbation, ecosystem functioning, and community structure. *Hydrology and Earth System Sciences,* **6**: 999-1005.

Blakeslee AMH, Byers JE, Lesser MP. 2008. Solving cryptogenic histories using host and parasite molecular genetics: the resolution of *Littorina littorea* 's North American origin. *Molecular Ecology* **17**: 3684–3696.

Blakeslee AMH, McKenzie CH, Darling JA, Byers JE, Pringle JM, Roman J. 2010. A hitchhiker's guide to the Maritimes: anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland. *Diversity and Distributions* **16**: 879–891.

Bousfield EL. 1973. Shallow-water Gammaridean Amphipoda of New England. Ithaca, NY: Cornell University Press.

Brawley SH, Coyer JA, Blakeslee AMH, Hoarau G, Johnson LE, Byers JE, Stam WT, Olsen JL. 2009. Historical invasions of the intertidal zone of Atlantic North America associated with distinctive patterns of trade and emigration. *Proceedings of the National Academy of Sciences, USA* **106**: 8239–8244.

Briski E, Ghabooli S, Bailey SA, MacIsaac HJ. 2012. Invasion risk posed by macroinvertebrates transported in ships' ballast tanks. *Biological Invasions* **14**: 1843–1850.

Cahill AE, Viard F. 2014. Genetic structure in native and non-native populations of the direct-developing gastropod *Crepidula convexa*. *Marine Biology* **161**: 2433–2443.

Carlton JT. 1996. Biological invasions and cryptogenic species. *Ecology* **77**: 1653–1655.

Carlton JT. 1999. Molluscan invasions in marine and estuarine communities. *Malacologia* **41**: 439-454.

Carlton JT. 2003. Community assembly and historical biogeography in the North Atlantic Ocean: the potential role of human-mediated dispersal vectors. *Hydrobiologia* **503**: 1–8.

Carlton JT, Cohen AN. 2003. Episodic global dispersal in shallow water marine organisms: the case history of the European shore crabs *Carcinus maenas* and *C. aestuarii. Journal of Biogeography* **30**: 1809-1820.

Carlton JT, Ruiz GM. 2005. The magnitude and consequences of bioinvasions in marine ecosystems: implications for conservation biology. In: Norse EA, Crowder LB, eds. *Marine conservation biology: the science of maintaining the sea's biodiversity*. Washington: Island Press, 123-148.

Carlton JT. 2009. Deep invasion ecology and the assembly of communities in historical time. In: Rilov G, Crooks JA, eds. *Biological invasions in marine ecosystems.* Berlin: Springer-Verlag, 13-56.

Chapin FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OS, Hobbie SE, Mack MC, Diaz S. 2000. Consequences of changing biodiversity. *Nature* **405**: 234-242.

Chapman JW. 1999. Climate effects on the geography of nonindigenous Pericaridan Crustacean introductions in estuaries. In: J Pederson, ed. *Marine bioinvasions: proceedings of the first national conference, January 24-27, Massachusetts Institute of Technology Sea Grant Program.* Cambridge, MA, 66-80.

Ciarelli S, van Straalen NM, Klap VA, van Wezel AP. 1999. Effects of sediment bioturbation by the estuarine amphipod *Corophium volutator* on fluoroanthene resuspension and transfer into the mussel (*Mytilus edulis*). *Environmental Toxicology and Chemistry* **18**: 318-328.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657-1660.

Colwell RK, Chao A, Gotelli NJ, Lin SY, Mao CX, Chazdon RL, Longino JT. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. *Journal of Plant Ecology* **5**: 3-21.

Commito JA. 1982. Importance of predation by infaunal polychaetes in controlling the structure of a soft-bottom community in Maine, USA. *Marine Biology* **68**: 77-81.

Crawford GI. 1937. A review of the amphipod genus *Corophium*, with notes on the British species. *Journal of the Marine Biological Association of the United Kingdom* **21**: 589-630.

Crooks JA. 2002. Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. *Oikos* **97**: 153-166.

Cuddington K, Hastings A. 2004. Invasive engineers. *Ecological modelling* **178**: 335-347.

Dawson MN, Gupta AS, England MH. 2005. Coupled biophysical global ocean model and molecular genetic analyses identify multiple introductions of cryptogenic species. *Proceedings of the National Academy Sciences*, *USA* **102**: 11968-11973.

Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB. 2008. Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology* **17**: 4992–5007.

De Deckere EMGT, van de Koppel J, Heip CHR. 2000. The influence of *Corophium volutator* abundance on resuspension. *Hydrobiologia* **426**: 37-42.

Denys N. 1672. The description and natural history of the coasts of North America (Acadia). Translated by William Ganong (1908).

Didyk AS, Bourgeois NA, Arp PA, Bourque J, Braune B, Cox GK, Ritchie C, Wells PG, Burt MDB. 2005. The roles of *Corophium volutator* and mercury in compromising shorebird migration. In: 6th Bay of Fundy Science Workshop, Cornwallis, Nova Scotia.

Dlugosch KM, Parker IM. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**: 431–449.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Drolet D, Barbeau MA. 2009. Diel and semi-lunar cycles in the swimming activity of the intertidal, benthic amphipod *Corophium volutator* in the upper Bay of Fundy. *Journal of Crustacean Biology* **29**: 51-56.

Drolet D, Kennedy K, Barbeau MA. 2013. Winter population dynamics and survival strategies of the intertidal mudflat amphipod *Corophium volutator* (Pallas). *Journal of Experimental Marine Biology and Ecology* **441**: 126-137.

Ehrenfeld JG. 2010. Ecosystem Consequences of Biological Invasions. *Annual Review of Ecology, Evolution, and Systematics* **41**: 59–80.

Einfeldt AL, Addison JA. 2013. Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the northwest Atlantic. *Marine Biology* **160**: 1015–1027.

Einfeldt AL, Doucet JR, Addison JA. 2014. Phylogeography and cryptic introduction of the ragworm *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. *Invertebrate Biology* **133**: 232-241.

Elton CS. 1958. The ecology of invasions by animals and plants. Methuen, London.

Excoffier L, Laval G, Schneider S. 2005. ARLEQUIN, v3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**: 47-50.

Foggo A, Attrill MJ, Frost MT, Rowden AA. 2003. Estimating marine species richness: an evaluation of extrapolative techniques. *Marine Ecology Progress Series* **248**: 15-26.

Frankham R. 2004. Resolving the genetic paradox in invasive species. *Heredity* **94**: 385–385.

Geller JB, Walton ED, Grosholz ED, Ruiz GM. 1997. Cryptic invasions of the crab *Carcinus* detected by phylogeography. *Molecular Ecology* **6**: 901-906.

Geller JB, Darling JA, Carlton JT. 2010. Genetic Perspectives on Marine Biological Invasions. *Annual Review of Marine Science* **2**: 367–393.

Gerdol V, Hughes RG. 1993. Effect of the amphipod *Corophium volutator* on the colonisation of mud by the halophyte *Salicornia europaea*. *Marine Ecology Progress Series* **97**: 61-69.

Gollasch S, MacDonald E, Belson S, Botnen H, Christensen J, Hamer J, Houvenaghel G, Jelmert A, Lucas I, Masson D, McCollin T, Olenin S, Persson A, Wallentinus I, Wetsteyn L, Wittling T. 2002. Life in ballast tanks. In: Leppakoski *et al.*, eds. *Invasive aquatic species of Europe*. Netherlands: Kluwer Academic Publishers, 317-231.

Gotelli NJ, Colwell RK. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**: 379-391.

Grosholz E. 2002. Ecological and evolutionary consequences of coastal invasions. *Trends in Ecology and Evolution* **17**: 22–27.

Hagerthey SE, Defew EC, Paterson DM. 2002. Influence of Corophium volutator and Hydrobia ulvae on intertidal benthic diatom assemblages under different nutrient and temperature regimes. *Marine Ecology Progress Series* **245**: 47-59.

Hamilton DJ, Barbeau MA, Diamond AW. 2003. Shorebirds, mud snails, and *Corophium volutator* in the upper Bay of Fundy, Canada: predicting bird activity on intertidal mud flats. *Canadian Journal of Zoology* **81**: 1358-1366.

Hanfling B, Kollmann J. 2002. An evolutionary perspective on biological invasions. *Trends in Ecology and Evolution* **17**: 545-546.

Hare MP. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**: 700-706.

Harper FM, Addison JA, Hart MW. 2007. Introgression versus immigration in hybridizing high-dispersal echinoderms. *Evolution* **61**: 2410-2418.

Hauber DP, Saltonstall K, White DA, Hood CS. 2011. Genetic variation in the common reed, *Phragmites australis*, in the Mississippi river delta marshes: evidence for multiple introductions. *Estuaries and Coasts* **34**: 851-862.

Hawkins CM. 1985. Population carbon budgets and the importance of the amphipod *Corophium volutator* in the carbon transfer on a Cumberland Basin mudflat, upper Bay of Fundy, Canada. *Netherlands Journal of Sea Research* **19**: 165-176.

Haydar D, Hoarau G, Olsen JL, Stam WT, Wolff WJ. 2011. Introduced or glacial relict? Phylogeography of the cryptogenic tunicate *Molgula manhattensis* (Ascidiacea, Pleurogona). *Diversity and Distributions* **17**: 68–80.

Haydar D. 2012. What is natural? The scale of cryptogenesis in the North Atlantic Ocean. *Diversity and Distributions* **18**: 101–110.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

Highsmith RC. 1985. Floating and algal rafting as potential dispersal mechanisms in brooding invertebrates. *Marine Ecology Progress Series* **25**: 169-179.

Huntsman AG, Sparks MI. 1924. No. 6: Limiting factors for marine animals. 3. Relative resistance to high temperatures. *Contributions to Canadian Biology and Fisheries*, *Nova Scotia* **2**: 97-114.

Hughes RG, Gerdol V. 1997. Factors affecting the distribution of the amphipod *Corophium volutator* in two estuaries in south-east England. *Estuarine, Coastal and Shelf Science* **44**: 621-627.

Ingolfsson A. 1992. The origin of the rocky shore fauna of Iceland and the Canadian Maritimes. *Journal of Biogeography* **19**: 705-712.

Ingolfsson A. 1995. Floating clumps of seaweed around Iceland: natural microcosms and a natural means of dispersal for shore fauna. *Marine Biology* **122**: 13-21.

Ingolfsson A. 1996. The distribution of intertidal macrofauna on the coasts of Iceland in relation to temperature. *Sarsia* **81**: 29-44.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* **19**: 101-108.

Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* **90**: 773-795.

Krebes L, Blank M, Bastrop R. 2011. Phylogeography, historical demography and postglacial colonization routes of two amphi-Atlantic distributed amphipods. *Systematics and Biodiversity* **9**: 259-273.

Lacoursière-Roussel A, Bock DG, Cristescu ME, Guichard F, Girard P, Legendre P, McKindsey CW. 2012. Disentangling invasion processes in a dynamic shipping-boating network. *Molecular Ecology* **21**: 4227–4241.

Lefébure T, Douady CJ, Gouy M, Gibert J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* **40**: 435-447.

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

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

MacDonald EC, Ginn M, Hamilton DJ. 2012. Variability in foraging behaviour and implications for diet breadth among semipalmated sandpipers staging in the upper Bay of Fundy. *The Condor* **114**: 135-144.

Macfarlane CBA, Drolet D, Barbeau MA, Hamilton DJ, Ollerhead J. 2013. Dispersal of marine benthic invertebrates through ice rafting. *Ecology* **94**: 250-256.

Maclean MF, Dadswell MJ, Stokesbury MJW. 2013. Feeding ecology of the Atlantic sturgeon, *Acipenser oxyrinchus oxyrinchus* (Mitchill, 1815) on the infauna of Minas Basin, Bay of Fundy. *Journal of Applied Ichthyology* **29**: 503-509.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Vainola R, Viard F, Wares J. 2008. Evaluating signature of glacial refugia for North Atlantic Benthic Marine Taxa. *Ecology* **89**: S108–S122.

McCurdy DG, Forbes MR, Logan SP, Lancaster D, Mautner SI. 2005. Foraging and impacts by benthic fish on the intertidal amphipod *Corophium volutator*. *Journal of Crustacean Biology* **25**: 558-564.

Meadows PS, Tait J. 1989. Modification of sediment permeability and shear strength by two burrowing invertebrates. *Marine Biology* **101**: 75-82.

Meadows PS, Tait J, Hussain SA. 1990. Effects of estuarine infauna on sediment stability and particle sedimentation. *Hydrobiologia* **190**: 263-266.

Mermillod-Blondin F, Rosenberg R. 2011. The functional significance of bioturbation and biodeposition on biogeochemical processes at the water-sediment interface in

freshwater and marine ecosystems. *Journal of the North American Benthological Society* **30**: 770-778.

Minchin D, Gollasch S, Cohen AN, Hewitt CL, Olenin S. 2009. Characterizing vectors of marine invasion. *Biological invasions in marine ecosystems* (ed. by G. Rilov and J.A. Crooks), pp. 109-116, Springer Berlin Heidelberg.

Miura O, Torchin ME, Kuris AM, Hechinger RF, Chiba S. 2006. Introduced cryptic species of parasites exhibit different invasion pathways. *Proceedings of the National Academy of Sciences, USA* **103**: 19818–19823.

Mouritsen KN, Mouritsen LT, Jensen KT. 1998. Change of topography and sediment characteristics on an intertidal mud-flat following mass-mortality of the amphipod *Corophium volutator*. *Journal of the Marine Biological Association of the UK* **78**: 1167-1180.

Murdoch MH, Barlocher F, Laltoo ML. 1986. Population dynamics and nutrition of *Corophium volutator* (Pallas) in the Cumberland Basin (Bay of Fundy). *Journal of Experimental Marine Biology and Ecology* **103**: 235-249.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1-10.

Palsson S, Magnusdottir H, Reynisdottir S, Jonsson ZO, Ornolfsdottir EB. 2014. Divergence and molecular variation in common whelk *Buccinum undatum* (Gastropoda: Buccinidae) in Iceland: a trans-Atlantic comparison. *Biological Journal of the Linnean Society* **111**: 145-159.

Panova M, Blakeslee AMH, Miller AW, Makinen T, Ruiz GM, Johannesson K, Andre C. Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. *PLoS ONE* **6**: e17511.

Pearson NJ, Gingras MK, Armitage IA, Pemberton SG. 2007. Significance of Atlantic sturgeon feeding excavations, Mary's Point, Bay of Fundy, New Brunswick, Canada. *Palaios* **22**: 5-12.

Peer DL, Linkletter LE, Hicklin PW. 1986. Life history and reproductive biology of *Corophium volutator* (Crustacea: Amphipoda) and the influence of shorebird predation on population structure in Chignecto Bay, Bay of Fundy, Canada. *Netherlands Journal of Sea Research* **20**: 359-373.

Pelegri SP, Blackburn TH. 1994. Bioturbation effects of the amphipod *Corophium volutator* on microbial nitrogen transformations in marine sediments. *Marine Biology* **121**: 253-258.

Pihl L, Rosenberg R. 1984. Food selection and consumption of the shrimp *Crangon crangon* in some shallow marine areas in western Sweden. *Marine Ecology Progress Series* **15**: 159-168.

Provan J, Bennett K. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution* **23**: 564–571.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Roman J. 2006. Diluting the founder effect: cryptic invasions expand a marine invader's range. *Proceedings of the Royal Society Biological Sciences* **273**: 2453–2459.

Rousset F. 2008. Genepop'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**: 103–106.

Ruiz GM, Fofonoff PW, Hines AH. 1999. Non-indigenous species as stressors in estuarine and marine communities: assessing invasion impacts and interactions. *Limnology and Oceanography* **44**: 950-972.

Ruiz GM, Fofonoff PW, Carlton JT, Wonham MJ, Hines AH. 2000. Invasion of coastal marine communities in North America: apparent patterns, processes, and biases. *Annual Review of Ecology, Evolution, and Systematics* **31**: 481–531.

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG. 2001. The population biology of invasive species. *Ecology Evolution and Systematics* **32**: 305-332.

Scheltema RS. 1971. Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. *Biological Bulletin* **140**: 284-322.

Simberloff D. 2011. How common are invasion-induced ecosystem impacts? *Biological invasions* **13**: 1255-1268.

Sizmur T, Canario J, Gerwing TG, Mallory ML, O'Driscoll NJ. 2013. Mercury and methylmercury bioaccumulation by polychaete worms is governed by both feeding ecology and mercury bioavailability in coastal mudflats. *Environmental Pollution* **176**: 18-25.

Slade RW, Moritz C. 1998. Phylogeography of *Bufo marinus* from its natural and introduced ranges. *Proceedings of the Royal Society Biological Sciences* **265**: 769-777.

Slothouber Galbreath JGM, Smith JE, Becnel JJ, Butlin RK, Dunn AM. 2010. Reduction in post-invasion genetic diversity in *Crangonyx pseudogracilis* (Amphipoda: Crustacea): a genetic bottleneck or the work of hitchhiking vertically transmitted microparasites? *Biological Invasions* **12**: 191–209.

Stephens M, Smith N, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978-989.

Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics* **76**: 449-462.

Thorson G. 1950. Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews* **25**: 1-45.

Vainio JK, Vainola R. 2003. Refugial races and postglacial colonization history of the freshwater amphipod *Gammarus lacustris* in Northern Europe. *Biological Journal of the Linnean Society* **79**: 523-542.

Vila M, Espinar JL, Hejda M, Hulme PE, Jarosik V, Maron JL, Pergl J, Schaffner U, Sun Y, Pysek P. 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters* **14**: 702-708.

Vitousek PM, D'Antonio CM, Loope LL, Rejmanek M, Westbrooks R. 1997. Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology* **21**: 1–16.

Voisin M, Engel CR, Viard F. 2005. Differential shuffling of native genetic diversity across introduced regions in a brown alga: aquaculture vs. maritime traffic effects. *Proceedings of the National Academy of Sciences, USA* **102**: 5432-5437.

Wallentinus I, Nyberg CD. 2007. Introduced marine organisms as habitat modifiers. *Marine Pollution Bulletin* **55**: 323-332.

Walther BA, Morand S. 1998. Comparative performance of species richness estimation methods. *Parasitology* **116**: 395-405.

Widdows J, Brinsley M. 2002. Impact of biotic and abiotic processes on sediment dynamics and the consequences to the structure and functioning of the intertidal zone. *Journal of Sea Research* **48**: 143-156.

Wilson AB, Boates JS, Snyder M. 1997. Genetic isolation of populations of the gammaridean amphipod, *Corophium volutator,* in the Bay of Fundy, Canada. *Molecular Ecology* **6**: 917-923.

Wares JP, Cunningham CW. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**: 2455–2469.

Zink RM, Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**: 2107-2121.

Table 3.1 Potential causes of *Corophium volutator*'s disjunct distribution in the North Atlantic and predictions of genetic consequences.

Cause of disjunct range	Diversity in North America	Genetic pattern
Taxonomic artefact	High	Reciprocal monophyly; divergence > molecular threshold for species
Persistence in glacial refugia	High	Reciprocal monophyly; divergence < molecular threshold for species
Natural trans-oceanic dispersal	High/ Intermediate	No reciprocal monophyly, alleles descended from source diversity
Human-mediated introduction	Low	Alleles are subsample of source diversity

Table 3.2 Concatenated mitochondrial and nuclear DNA diversity for *Corophium volutator* populations surveyed: sampling sites, number of individuals sequenced per locus (*n*), number of segregation sites (*S*), number of haplotypes or alleles (*H*), haplotype diversity (*h*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and mean nucleotide diversity (π).

			COI/SUBI					*CV1*						*CV2*					
Site	**Population**	**Abbr.**	***n***	***S***	***H***	***h***	π	***n***	***S***	***H***	***Ho***	***He***	π	***n***	***S***	***H***	***Ho***	***He***	π
1	Nieuwpoort	**BG**	6	19	3	0.6000	0.0093	28	5	5	0.5714	0.5688	0.0034	16	5	5	0.5000	0.5333	0.0042
2	Warnow	**DEB**	10	0	1	0.0000	0.0000	12	2	3	0.6667	0.6212	0.0039	12	4	4	0.3333	0.4545	0.0054
3	Sylt	**DEN**	11	20	5	0.7636	0.0088	14	5	5	0.5714	0.7582	0.0074	10	5	4	0.6000	0.6444	0.0063
4	Salzhaff	**DES**	6	18	2	0.5333	0.0090	14	6	9	0.8571	0.8791	0.0075	10	6	4	0.6000	0.5333	0.0060
5	Dorum-Neufeld	**DEW**	6	18	3	0.6000	0.0088	6	5	5	1.0000	0.9333	0.0086	6	5	5	0.6667	0.9333	0.0096
6	Roskilde	**DK**	9	19	2	0.2222	0.0040	14	2	3	0.5714	0.7912	0.0025	12	4	4	0.6667	0.7576	0.0075
7	Skallingen	**DKS**	8	19	4	0.7500	0.0089	12	4	5	0.8333	0.7879	0.0064	12	6	4	0.5000	0.5606	0.0070
8	St.-Estephe	**FGI**	6	17	2	0.5333	0.0085	10	3	3	0.2000	0.6444	0.0046	12	3	3	0.6667	0.5303	0.0065
9	Loire	**FLO**	21	25	3	0.4667	0.0067	16	5	7	0.4444	0.8758	0.0059	36	5	5	0.1667	0.2603	0.0024
10	Le Platin	**FLR**	17	18	7	0.7206	0.0073	20	4	7	1.0000	0.8737	0.0063	8	4	4	0.6000	0.6444	0.0059
11	Guillec	**FRG**	12	0	1	0.0000	0.0000	20	4	5	0.7000	0.7368	0.0043	22	3	4	0.3636	0.4632	0.0025
12	Seine	**FRS**	11	19	5	0.6182	0.0076	24	6	7	0.8333	0.8732	0.0059	18	1	2	0.1111	0.1111	0.0006
13	Ireland	**IRL**	12	15	4	0.7576	0.0070	16	3	4	0.3750	0.6167	0.0039	16	5	4	0.7500	0.6917	0.0082
14	Netherlands	**NL**	18	18	3	0.2157	0.0034	26	8	11	0.8462	0.8831	0.0076	12	5	5	0.6250	0.8083	0.0048
15	Scotland	**SY**	12	20	3	0.6212	0.0096	14	3	4	0.7143	0.7582	0.0037	16	6	5	0.8750	0.6833	0.0056
16	United Kingdom	**UKP**	7	17	2	0.4762	0.0076	12	5	7	0.7143	0.9121	0.0064	22	5	7	0.7273	0.6840	0.0056
	Europe		**172**	**52**	**34**	**0.9254**	**0.0108**	**258**	**16**	**31**			**0.8525**	**240**	**9**	**15**			**0.0060**
	North America		**138**	**8**	**8**	**0.6452**	**0.0011**	**252**	**5**	**9**			**0.0055**	**191**	**4**	**7**			**0.0084**

Table 3.3 Observed and expected number of haplotypes and alleles in Europe (NEA) and North America (NWA) for *Corophium volutator*. Non-overlapping confidence intervals indicate higher genetic diversity in Europe than in North America, suggesting unique North American haplotypes and alleles are likely among the many remaining unsampled in Europe.

Marker	Region	Populations sampled	H_{obs}	$H_{max\text{-}exp}$	Lower 95%	Upper 95%
COI/SUBI	Europe	16	34	216.3	79.4	765.7
	Bay of Fundy	9	4	4.9	4.1	16
	Gulf of Maine	8	5	10.3	5.9	37.5
	North America	17	8	17.4	9.8	56.9
CV1	Europe	16	31	296.5	78.9	1522.6
	Bay of Fundy	16	8	13.6	8.9	43.3
	Gulf of Maine	8	4	4.0	4.0	5.48
	North America	24	9	18.6	10.8	58.9
CV2	Europe	16	15	17.7	15.4	33.9
	Bay of Fundy	14	5	5.0	5.0	5.8
	Gulf of Maine	5	3	3.0	3.0	3.2
	North America	19	7	5.0	5.0	5.0

Table 3.4. Comparison of models using log Bayes factors (LBF) from Bezier approximated log marginal likelihood (Bezier lmL) estimated in MIGRATE-N for nuclear and mitochondrial DNA from *Corophium volutator*.

Model	Description of model	*COI/SUBI* lmL (10 run average)	LBF	Probability	Rank	*CV1+CV2* lmL (10 run average)	LBF	Probability	Rank
1	'Single introduction pathway': asymmetric migration from the NEA to the middle region of the NWA, symmetric migration between adjacent regions in the NWA	-3738.51	-3.0	0.1779	2	-2893.71	-29.7	0.0000	4
2	'Two introduction pathways': asymmetric migration from the NEA to the northern and southern regions of the NWA, symmetric migration between adjacent regions in the NWA	-3736.99	0.0	0.8122	best	-2886.79	-15.8	0.0003	3
3	'Three introduction pathways': asymmetric migration from the NEA to all regions of the NWA, symmetric migration between adjacent regions in the NWA	-3741.41	-8.8	0.0098	3	-2878.88	0.0	0.9277	best
4	'Two populations': all samples in the NWA are of the same population, asymmetric migration from the NEA to the NWA	-3812.32	-150.7	0.0000	5	-2964.72	-171.7	0.0000	5
5	'Open': symmetric migration between the NEA and the three NWA regions	-3748.86	-23.7	0.0000	4	-2881.44	-5.1	0.0720	2
6	'Panmixia': all samples are of the same population	-4197.46	-920.9	0.0000	6	-3177.69	-597.6	0.0000	6

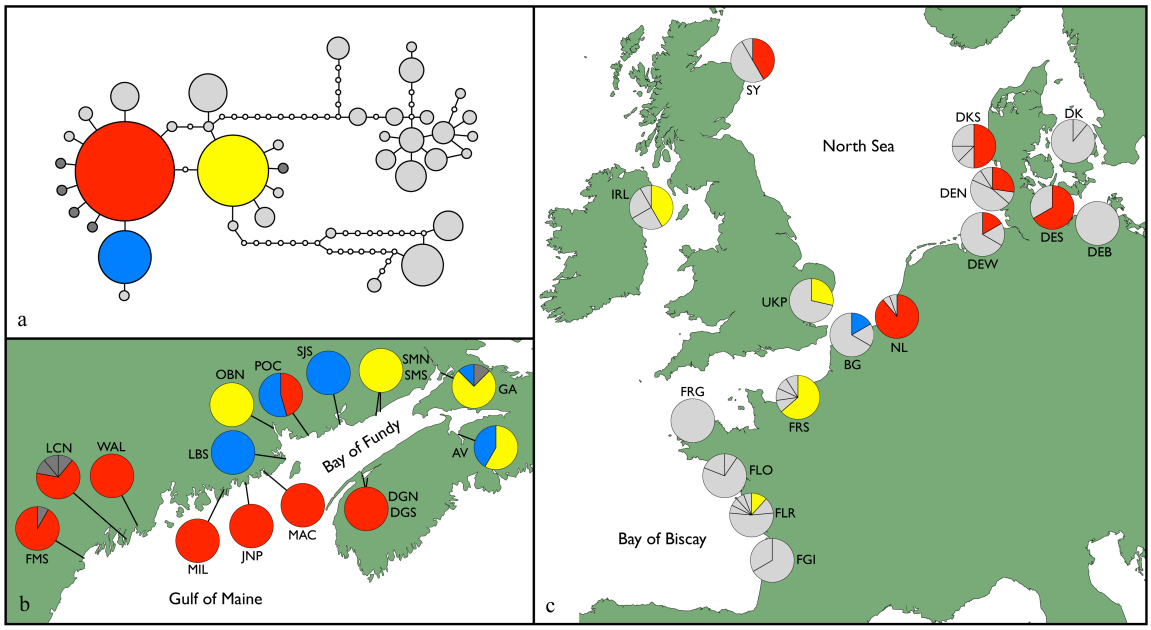


Figure 3.1 a) Haplotype network for concatenated *COI*/*SUBI* mitochondrial DNA for *Corophium volutator*. Circle area is proportional to the number of haplotypes sequenced. b) Sampling locations and haplotype distribution of *COI*/*SUBI* for *Corophium volutator* in the NWA, from Einfeldt & Addison (2013). c) Sampling locations and haplotype distribution of *COI*/*SUBI* for *Corophium volutator* in the NEA. Red, blue, and yellow haplotypes are shared between the NWA and the NEA; dark shaded haplotypes were found once, only in the NWA; light shaded haplotypes were found only in the NEA.

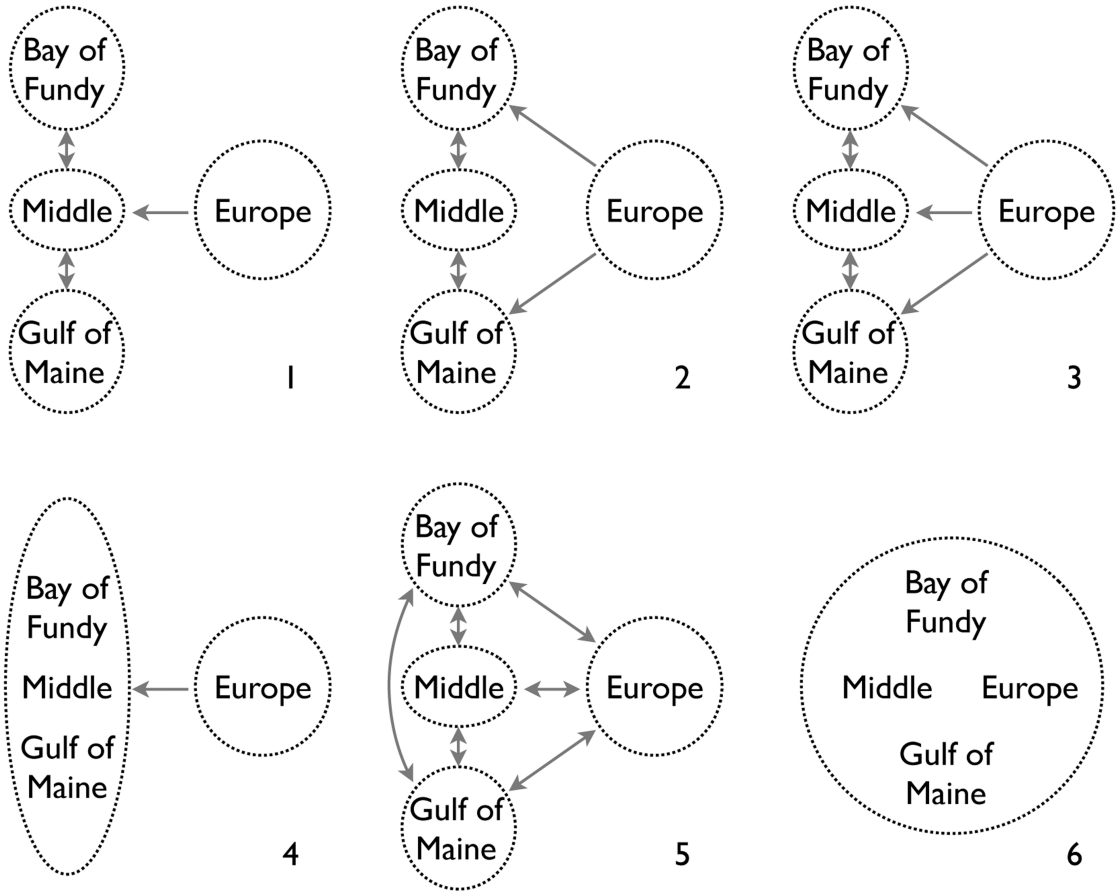


Figure 3.2 Diagrams of migration models tested using concatenated mitochondrial DNA (*COI*/*SUBI*) and nuclear DNA (*CV1* and *CV2*) of *Corophium volutator* in MIGRATE-N. Numeric labels correspond to models described in the introduction and Table 3.4. Dotted lines depict population groupings of four biogeographic regions: the Bay of Fundy, the Gulf of Maine, a Middle region in North America, and Europe. Arrows depict directionality of gene flow among populations.

Chapter 4 – The geographic scale of diversification was reshaped by shipping in the Age of Exploration[a]

Einfeldt AL[1] (corresponding author), Jesson LK[2], Addison JA[3]

[a]*In review, Nature Communications Biology, February 2018*

[1]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3. email: tony.einfeldt@gmail.com.

[2] The New Zealand Institute for Plant & Food Research Limited.

[3]University of New Brunswick. Biology Department, PO 4400. Fredericton, New Brunswick, Canada, E3B5A3

Abstract

Human activity can alter natural patterns of gene flow in species associated with human vectors of transport. To investigate how recent human movement shapes evolution in species associated with human transport, we exploit a system where changes in shipping practices over the last millennium created a temporary window of human-mediated dispersal for two sympatric species native to the Old World and introduced to New World. We present evidence of past mixture among native populations and strong vicariance between native and introduced populations. We find similar genetic patterns in species from different phyla, demonstrating that the evolutionary effects of human-mediated dispersal are repeatable across a wide variety of taxa. Our results support predictions that human-mediated dispersal erases ancestral genetic structure arising from natural evolutionary processes, but also creates new populations that evolve independently from their sources. Together, these processes shift the geographic scale at which diversification occurs from regional to global. These findings provide valuable insight into how contemporary human activity is likely reshaping global patterns of biodiversity across many systems.

Introduction

Humans are moving species beyond the limits of their natural dispersal abilities at unprecedented rates (Seebens et al. 2017), but how this shapes the evolutionary trajectories of introduced species is contentious. Human-mediated dispersal promotes the formation of new lineages when introduced populations are geographically isolated from their sources (Lee 2002; Montesinos et al. 2012). However, the same vectors can also negate ancestral divergence between native populations by disrupting natural barriers to gene flow (Alvarez et al. 2007; Vonlanthen et al. 2012; Hudson et al. 2016). Comparing these two processes is challenging due to the short time scales at which most documented introductions occur (Vellend et al. 2007). Thus, the net effect of human-mediated dispersal on diversification is uncertain.

Barriers to dispersal are expected to promote the accumulation of genetic differences among populations via genetic drift and selection (Coyne and Orr 2004; Cowie and Holland 2006). Geographically limited natural dispersal can lead to isolation by distance (IBD) that shapes patterns of genetic diversity within native ranges (Wright 1943), and leads to the prediction that introduced populations should be most genetically similar to populations closest to the source of their founders (Estoup and Guillemaud 2010). Human-mediated transport is expected to disrupt these patterns by artificially increasing gene flow among populations that would not otherwise be directly connected, resulting in the loss of ancestral patterns of genetic structure. Because human vectors of transport are typically more active at regional than global scales (Seebens et al. 2013), populations established via human activity are predicted to be less connected to their

sources than native populations are to each other, shifting the geographic scale of diversification from regional to global. These changes could preclude application of phylogeographic principles to identify source populations and colonization pathways, influencing our understanding of the mechanisms driving evolutionary change in species associated with human activity.

To investigate how human activity reshapes evolutionary patterns, we exploit a system of historic changes to shipping technology that resulted in a temporary period of human-mediated movement of estuarine invertebrates that have limited natural capacities for dispersal. In the 10th Century, the adoption of wooden ships that used rocks and sediments dredged from the intertidal and near-shore (semi-dry ballast) for stability enabled long-distance exchange of abiotic and biotic materials in the European Atlantic (Steel 1832; McGrail 1989; Ansorge et al. 2011). During the 11th-15th Centuries the magnitude and reach of these practices increased with the rise of maritime trade around the North and Baltic Seas, and extended in geographic reach during the mid 15th Century with European exploration and trade between Europe and the New World. This exchange abruptly declined, or indeed likely ceased entirely, in the late 19th Century when wooden vessels were replaced by steel-hulled ships that use seawater as ballast (Carlton 2003).

Sediment-dwelling species that lack a pelagic life history stage are expected to associate with semi-dry ballast but not water ballast (Hewitt et al. 2009), and newly founded populations introduced via semi-dry ballast are thus expected to have been isolated from their sources for over a century. In the absence of gene flow, these introduced populations are free to diverge from their sources. This provides a benchmark to which patterns of genetic differentiation between native populations can be compared.

If human-mediated dispersal changes the geographic scale at which diversification occurs, genetic differentiation between native populations is expected to be similar to or weaker than that observed between populations in both ranges.

Here, we investigate the influence of human-mediated transport on intraspecific genetic diversification in the co-occurring amphipod *Corophium volutator* (Pallas 1766) and annelid *Hediste diversicolor* (Müller 1776) introduced from Europe to North America (Einfeldt et al. 2014; Einfeldt and Addison 2015). Both species are brood-rearing (i.e. lack pelagic larval dispersal) and reside in shallow constructed burrows during their adult stages (Peer et al. 1986; Scaps 2002), which is expected to impart a poor capacity for natural long-distance dispersal, and a higher probability of being incorporated into semi-dry ballast than water ballast. We use genome-wide SNPs from native European and naturalized North American populations to compare patterns of divergence within and between ranges, providing insight into the relative importance of the homogenizing and diversifying effects of human-mediated dispersal to evolution.

Methods

Sampling

We sampled intertidal mudflats for *C. volutator* and *H. diversicolor* in Europe between July-September 2013 and in North America between July-August 2014, preserving individual specimens in 95% ethanol and freezing at -20°C before further processing. From our collections, we selected 121 individuals of *Corophium volutator* from 21 sites and 104 individuals of *Hediste diversicolor* from 20 sites (Table 4.1; Figure 4.1) from across the geographic distributions of both species. Some samples of *H.*

diversicolor were identified by mitochondrial DNA to belong to cryptic species not originating from the North Atlantic (Figure 4.1; Virgilio et al. 2009), and were excluded from analyses in order to focus on intraspecific patterns of genetic variation.

ddRADseq library preparation

We extracted genomic DNA from whole adults of *C. volutator* (removing eggs and juveniles from females to avoid extraction of DNA from fertilized eggs) and tissue samples of *H. diversicolor* using DNeasy blood and tissue kits (Qiagen) following the manufacturers protocol. We prepared double digest restriction-site associated DNA (ddRADseq) (Peterson et al. 2012) libraries for each individual, which were then combined for a total of 5 lanes of sequencing on the Illumina HiSeq platform. We digested 100ng of DNA from each individual with Sau96I (a CpG methylation sensitive infrequent cutter) and MlucI (a methylation non-sensitive frequent cutter) for 2.5 hours at 37°C. To allow identification of sequences from each individual, we ligated 1 of 57 adapters with a unique 4-9bp barcode at the Sau96I cut-site remnant for each individual sample per library. To reduce PCR amplification of fragments with a MlucI (frequent) cut-site on both ends, we ligated 1 of 4 universal Y-adapters to the MlucI cut-site remnant. We pooled 50ng of ligated DNA from each individual to their respective libraries and purified these pooled samples using QIAquick PCR purification kits (Qiagen) following the manufacturer's protocol. We amplified each pool with 16 PCR cycles of 98°C, 65°C, and 72°C. To purify the PCR products and size select for fragments >100bp and <300bp, we performed gel extractions on each library using

QIAquick Gel Extraction kits (Qiagen) following the manufacturer's protocol. We quantified the concentration of DNA in each library using a Qubit fluorometer and sequenced 100bp paired-end reads with the Illumina HiSeq platform at Génome Québec.

ddRAD-seq pre-processing

We demultiplexed raw DNA sequences using *fastq-multx* (Aronesty 2013), keeping only exact matches. We trimmed barcodes and cut-site remnants with *fastx_trimmer*. To filter for read quality and remove read-through resulting from sequencing short DNA fragments, we used *TrimmomaticPE* (Bolger et al. 2014) based on recommended filters (Del Fabbro et al. 2013) of a quality threshold of *phred*=22, a sliding window of 4bp with a quality threshold of *phred*=28, and sequence lists to be clipped for all combined Illumina adapters and barcodes. As short reads result in overlapping paired-end reads, we merged reads using *PEAR* (Zhang et al. 2013). To make all sequences a standard length for downstream analyses, we trimmed merged and non-merged reads to 85bp, which maximized the overall number of nucleotides kept in the total data set. Our genome reduction approach on whole-organism DNA results in DNA contamination by parasites and common sympatric micro-organisms, which cannot be filtered by aligning to a genome due to the lack of available genomic resources for *C. volutator* and *H. diversicolor*. To reduce sequence contamination, we filtered reads using Kraken (Wood and Salzberg 2014) against the standard human database, the standard Minikraken microbial contaminant database, and custom databases built from publicly available genomes for 13 trematodes, 57 nematodes, 4 myxozoans, 20 microsporidians, 9

basidiomycetes, 41 apicomplexans, 36 oomycetes, and 26 algae (Table 4.3). Contaminating species are not likely to be present in every single location sampled, so we further reduced sequence contamination by filtering for missing data at the population level, as described below.

ddRAD-seq de novo assembly

We assembled a *de novo* reference and called SNPs using STACKS (Catchen et al. 2013) on the ACEnet's Fundy computing resource cluster. We compared results for several different combinations of parameters, with the number of nucleotide mismatches within an individual allowed to group stacks of identical reads into a locus (*M*) set to 2, 4, 6, or 8, and the number of nucleotide mismatches among stacks among individuals to form catalog loci (*n*) set to 2, 4, 6, or 8. Minimum stack depth was kept constant at 3. For each of these combinations of *M* and *n*, we ran the STACKS program *populations* with a filter requiring each population (*r*) to have 50% or 100% of individuals with data to retain a locus, and a filter requiring the proportion of populations that have data for a locus (*p*) to be 50%, 75%, or 100%. We ran STACKS twice for each of these combinations of parameters, once keeping all SNPs in each stack and once keeping only the first SNP from each stack. We found that increasing both *M* and *n* had little effect on the number of loci retained but resulted in a plateau at *M*=6 and *n*=6, and thus used these settings for downstream analyses. Decreasing *r* or *p* increased the number of loci retained. While retaining a greater number of loci is desirable, missing data from all individuals in a population can bias tests that rely on allele frequencies at the population level. To reduce

this bias, we only retained loci with data in 100% of populations (p=1) and for at least 50% of individuals in each population (r=0.5) for downstream analyses. To reduce the effects of physical linkage, we used datasets retaining only a single SNP per stack for all analyses presented here, and used datasets retaining all SNPs per stack for validation. After filtering for loci with minor allele frequencies >0.01, we retained 4870 SNPs for *C. volutator* and 3820 SNPs for *H. diversicolor* for all further analyses.

Genomic diversity and structure

Introduced populations may suffer a loss of diversity due to subsampling from their sources and the increased effects of genetic drift on small founding populations, which if severe may be a source of differentiation between native and introduced ranges. To assess the severity of founding effects associated with introduction we calculated observed and expected heterozygosity using Adegenet (Jombart and Ahmed 2011).

To investigate the geographic structure of independently evolving lineages, we explored genomic subdivision in *C. volutator* and *H. diversicolor* using ancestry proportion estimates based on sparse non-negative matrix factorization implemented in the R package SNMF 1.2 (Frichot et al. 2014). SNMF estimates ancestry coefficients for individuals from *K* ancestral panmictic gene pools, making no assumptions about the sampled or ancestral populations that individuals belong to. We calculated ancestry coefficients for *K* clusters from 1 to the number of populations in 5 replicate runs with a maximum of 5000 iterations and 5% of data masked using all populations, native populations only, and introduced populations only. We assessed the robustness of our

results by repeating these analyses for five values of the regularization parameter *alpha* on a logarithmic scale from 1-10,000. We determined the best number of clusters across replicate runs according to the minimum cross-entropy across all runs.

To test for differences in inter-population genomic differentiation between the native and introduced ranges of *C. volutator* and *H. diversicolor*, we calculated pairwise F_{ST} amongst populations in Europe and North America using the R package hierfstat (Goudet 2005). We then tested for a difference in mean inter-population F_{ST} between ranges using Student's t test. To test for correlation between pairwise inter-population genetic distance and geographic distance, we performed Mantel tests (Legendre et al. 2010) on Slatkin's Similarity Index (Slatkin 1995) calculated from genetic distances within each range.

Complex evolutionary scenarios can cause phylogenetic conflict between different genomic regions, making network methods useful in examining intraspecific relationships in the presence of recent gene flow. To examine phylogenetic relationships in *C. volutator* and *H. diversicolor* we created unrooted networks based on uncorrelated *p*-distances between all pairs of individuals using a neighbor-net method implemented in Splitstree (Huson and Bryant 2005). To determine phylogenetic support for branches we performed 1000 bootstrap iterations.

Simulated demographic scenarios

To explore the impact of human-mediated dispersal on patterns of divergence, we simulated SNPs from 10,000 unlinked neutrally evolving diploid loci of 85 base-pairs

each under scenarios of differing magnitudes of human-mediated migration in continuous-time simulations implemented in fastcoalsim2 (Excoffier et al. 2013). While extensive human-mediated gene flow precludes both formal testing of demographic scenarios or estimation of demographic parameters from site frequency spectra, forward simulations can provide insight into how particular demographic parameters theoretically affect spatial genetic patterns under strictly neutral evolution.

We created a base model of demographic events representing the major factors expected to shape genetic structure in our system: population expansion and colonization of new habitat in a native range from a glacial refugia to two island populations exchanging migrants and 8 mainland populations with stepping-stone migration, with 5 introduced populations with stepping-stone dispersal founded independently from a single mainland population (Figure 4.3). Upon founding, each population undergoes population expansion. An effective population size (N_e) of 2000 was used for every population except 7 (N_e = 1000), 15 (N_e = 1000), and 16 (N_e = 200), with these variations intended to demonstrate the effects of decreased *Ne* on population structure. We used a mutation rate of $2.5x10e^{-8}$ per generation per site, which generated distributions of SNPs per 85bp locus that approximated the distribution of SNPs per locus observed in RADseq data from *C. volutator* and *H. diversicolor*. Migration rates are scaled by the effective population size (N_em). We set the magnitude of within-range human-mediated migration (H_m) active in this system over 1800 generations (corresponding to the ~900 year window over which semi-dry ballast was in widespread use and ~2 generations per year) to be an order of magnitude less than, equal to, and an order of magnitude more than the magnitude of natural migration (N_m). After the window of human-mediated gene flow,

populations returned to natural migration matrices for 200 generations. We believe that human-mediated migration exceeding the magnitude of natural migration by 10x is a biologically reasonable estimate, as *C. volutator* and *H. diversicolor* limited natural long-distance dispersal capacities but were likely moved en masse by semi-dry ballast dredged and relocated in Europe (in excess of millions of tonnes per year in some ports: Steel 1832). To visualize the results of our simulations, we created unrooted phylogenetic networks bootstrapped for phylogenetic support between branches (as previously).

Scans for selection

We performed scans for positive selection using two methods that are reported perform well in demographic scenarios involving hierarchical spatial structure and population expansion: an F_{ST} based approach implemented in OutFLANK (Whitlock and Lotterhos 2015) and a principle components based approach implemented in PCAdapt (Luu et al. 2017). To correct for bias arising from multiple tests, we calculated *q*-values from *p*-values produced by each test using the *R* package *qvalue* (Dabney et al. 2010) with a false discovery rate threshold of 5%. For PCAdapt, we chose the optimal number of genetic compenents *K* based on Cattell's rule applied to scree plots (Cattell 1966). To assess false positive rates, we performed both scans for selection on SNPs from data simulated using a strictly neutral model of evolution.

Results

Introduced populations show reductions in genetic diversity consistent with founder effects

Founding effects can facilitate evolutionary change between introduced populations and their sources via the random or selective establishment of some genotypes over others during the introduction process (Wares et al. 2005; Wood et al. 2016). To compare diversity in native and introduced populations, we genotyped 121 *C. volutator* and 78 *H. diversicolor* individuals from sites covering the geographic extent of their known distributions (Figure 4.1). Sequencing 2 x 100 base-pair paired-end reads of double-digest restriction-site associated DNA at >10X average depth of coverage and keeping 1 single-nucleotide variation (SNP) per locus produced a total of 4870 SNPs for *C. volutator* and 3820 SNPs for *H. diversicolor*. Average heterozygosity in North American populations was 14.9% lower for *C. volutator* and 16.6% lower for *H. diversicolor* than in European populations (Table 4.1), suggesting that both species underwent mild genetic bottlenecks during introduction to North America. This is similar to the 18.7% average difference between ranges observed in other introduced species (reviewed across 70 species; Dlugosch and Parker 2008), suggesting that the bottlenecks experienced by *C. volutator* and *H. diversicolor* contribute to genetic divergence between introduced populations and their sources at a similar rate to those observed in other introduced species.

Isolation and divergence between introduced and native ranges

Introduced populations are expected to be isolated from their ancestral sources (Lee 2002; Montesinos et al. 2012), especially in systems where introduction vectors have become defunct. To determine whether populations of *C. volutator* and *H. diversicolor* in North America represent genetic lineages that are evolving independently from European populations, we calculated the probability of assignment of each individual to *K* ancestral gene pools in *C. volutator* and *H. diversicolor* using Sparse Non-negative Matrix Factorization (SNMF, Frichot et al. 2014). In *C. volutator* the best number of clusters (*K*=2) differentiates introduced and native individuals, and in *H. diversicolor* the best number of clusters (*K*=4) differentiates introduced individuals from three genetic clusters in Europe (Figure 4.2). Increasing *K* beyond its optimal value reveals increasingly fine-scale clustering that corresponds to geography, particularly in the native range, but does not group North American individuals with European clusters (Figure 4.4). This divergence far exceeds that detected with mitochondrial DNA, as previous studies found North American haplotypes to be a subset of those found in European populations (Einfeldt et al. 2014; Einfeldt and Addison 2015). The strong divergence we detect here is unlikely to be an artefact of under-sampling native genetic diversity, as the geographic extent of populations we sampled encompasses the native ranges of these species including geographically isolated regions (e.g. the British Isles). The inability of genetic clustering to identify putative source regions is, however, consistent with high levels of historic gene flow followed by recent isolation amongst native populations, which we discuss below.

Weak genetic differentiation and isolation-by-distance between native populations

Native ranges are expected to have high levels of genetic divergence among geographically isolated populations, particularly in species with fragmented habitat continuity and poor natural dispersal capabilities (Bohonak 1999; Bilton et al. 2002). This assumption underlies a central paradigm of invasion genetics: introduced populations exhibit more genetic similarity to populations nearer their sources (Baker and Stebbins 1965). Contrary to these expectations, pairwise genetic differentiation amongst European populations was no greater than amongst North American populations for *C. volutator* ($F_{ST\text{-}NA}$=0.117; $F_{ST\text{-}EU}$=0.110; *P*=0.493) or for *H. diversicolor* ($F_{ST\text{-}NA}$=0.131; $F_{ST\text{-}EU}$=0.171; *P*=0.052). Although significant, isolation-by-distance amongst European populations of *C. volutator* (r^2=0.278; *P*<0.001) and *H. diversicolor* (r^2=0.057; *P*=0.03009) explains a low proportion of variance in genetic structure in the native range, even compared to species with greater natural dispersal capacities studied over smaller geographic scales (reviewed by Jenkins et al. 2010; Kelly and Palumbi 2010). These findings indicate that genetic structure amongst European populations of *C. volutator* and *H. diversicolor* cannot be explained by their innate dispersal capabilities alone, suggesting that human-mediated movement has been influential in restructuring native genetic diversity.

Phylogenetic conflict within the native range

To explore patterns of divergence, we constructed neighbor-networks bootstrapped for phylogenetic support from genome-wide SNPs in *C. volutator* and *H.*

diversicolor (Fig. 4.2a-b). Trace fossils suggest that both these species had broad distributions in the North Sea throughout the Holocene (Buller et al. 1972; Allen and Haslett 2002; Streif 1972), and are thus expected to exhibit divergence arising from: 1) isolation between multiple refugia persisting during Pleistocene glaciations, as indicated by their distinct mitochondrial DNA lineages (Virgilio et al. 2009; Einfeldt et al. 2014; Einfeldt and Addison 2015); 2) isolation-by-distance due to their limited natural dispersal capabilities, and; 3) isolation between the British Isles and mainland Europe. Populations evolving in complete isolation are expected to be monophyletic and accumulate genetic differences over time, while gene flow between populations would cause conflicting phylogenetic signals (Grosberg and Cunningham 2001; Maggs et al. 2008).

Our networks support reciprocal monophyly between the introduced and native ranges of both species, with branch lengths between the two allopatric lineages representing divergence that has accumulated since their introduction. Compared to genetic differentiation between ranges, there is surprisingly little divergence in the native European range of either species. While individuals sampled from the same site cluster together with high fidelity, our networks show extensive reticulation indicating nearly equal phylogenetic conflict between all pairs of native populations. This pattern is not consistent with long-standing isolation-by-distance or regional barriers to gene-flow, which are expected to produce branching networks that reflect colonization history and geographic features affecting dispersal (Figure 4.3). Instead, the phylogenetic conflict between genetically distinct European populations suggests a high degree of past connectivity and contemporary isolation between all populations, consistent with a temporary period of widespread human-mediated dispersal.

Human-mediated gene flow erodes ancestral genetic structure

To determine whether gene flow enabled by human-mediated dispersal can erase ancestral genetic structure, we simulated genetic data for sixteen neutrally evolving populations with a demographic history reflecting post-glacial expansion and colonization, isolation between three regions in a native range, and introduction to a new range (Figure 4.3). We varied the strength of natural (stepping-stone) and human-mediated (active across all populations within each range) gene flow occurring at times that correspond to the onset of widespread semi-dry ballast use in Europe (1000 years ago), European colonization in North America (400 years ago), and changes to shipping practices that made semi-dry ballast defunct (100 years ago). We then constructed bootstrapped neighbor-networks from simulated SNPs. While the effects of human-mediated gene flow depend on the magnitude of natural gene flow between populations, these networks show two clear emergent patterns: 1) Increasing the strength of human-mediated gene flow causes the relationships between native populations to become less branch-like and more radiation-like by increasing phylogenetic conflict between populations; 2) Increasing the strength of human-mediated gene flow changes introduced populations from clustering with their ancestral source to being their own divergent cluster. These trends demonstrate that human-mediated gene flow generally erodes the genetic signatures of demographic history and long-standing barriers to natural migration, precluding inference of demographic history from genetic data.

Genetic patterns are consistent with neutral evolution and preclude inference of selection by genomic scans

Genetic structure among populations can be built or eroded by both random genetic drift and selection. To determine whether selection contributes to distribution-wide patterns of divergence, we performed scans for positive selection using two methods that perform well in demographic models with hierarchical spatial structure and population expansion: the F_{ST} based OutFLANK (Whitlock and Lotterhos 2015) and the principle components based PCAdapt (Luu et al. 2017). To approximate false positive ratios for both methods, we compared analyses of SNPs from *C. volutator* and *H. diversicolor* to data simulated under a strictly neutral model of evolution (as above). The F_{ST} approach detected no outlier SNPs in simulated data sets or in either species. The principle components approach identified a large proportion of SNPs as outliers in *C. volutator* (16% of SNPs) and *H. diversicolor* (23% of SNPs), but these results were well within the 13-46% ($\alpha = 0.27$, $SD = 0.11$) range of false positive ratios identified in our analyses of simulated data (Table 4.2). That there is no spatially structured positive selection between any sites across such a broad range seems unlikely. Rather, our results suggest that these species' demographic histories are responsible for large shifts in allele frequency differences between populations that mask differences resulting from directional selection, precluding accurate inference of selection using frequency-based scans. Although the role of selection remains uncertain, the patterns of genetic divergence observed from genome-wide SNPs in *C. volutator* and *H. diversicolor* are thus consistent with expectations under a strictly neutral model of evolution.

Discussion

Here, we use genome-wide SNPs from two sympatric invertebrates introduced during the Age of Exploration to examine how human-mediated dispersal shapes evolution. We find genetic patterns consistent with introduction via semi-dry ballast, which was used extensively from ~1000 years ago and throughout the Age of Exploration but is now defunct. Our results support the hypothesis that the isolation of introduced populations from their sources allows them to follow independent evolutionary trajectories. While this implies that human-mediated dispersal sets the stage for diversification, we detect a surprising degree of genetic similarity across the native range of both species, suggesting that human-mediated dispersal is also homogenizing. Together, these results provide evidence that human-mediated dispersal reshapes the geographic scale of diversification, which has important ramifications for our ability to investigate species' natural histories and predict the future impact of human movement on patterns of global biodiversity.

The evolutionary changes we found in this system have likely occurred, or are well underway, in many other species. We detected remarkably similar genetic patterns in *C. volutator* and *H. diversicolor*, suggesting that human-mediated dispersal has repeatable effects across taxa associating with a shared vector of transport. The number of other species introduced to the North Atlantic during the Age of Exploration is grossly underestimated due to colonization and trade preceding scientific observation by over a century (Carlton 2003b; Haydar et al. 2012). Globally, undocumented introductions are rampant due to similar temporal and spatial biases in observation (Carlton 1996; Ruiz et al. 2000). There is a growing body of evidence demonstrating that the broad ranges of

many cosmopolitan species are the product of human-mediated dispersal (Estoup and Guillemaud 2010; Cristescu 2015), and biogeographic re-evaluations suggest that many more undocumented introductions will be revealed in the light of DNA sequencing (e.g. Bouchemousse and Viard 2016; Marchini et al. 2017). It is probable that the genetic patterns we present here will become evident in other species introduced via semi-dry ballast and other historical vectors.

In more recent introductions, there is growing evidence that the same processes shaping evolution in *C. volutator* and *H. diversicolor* are already underway. Rapid evolution in introduced species is common (Colautti and Barrett 2013; Moran and Alexander 2014), which accelerates divergence between newly founded populations and their sources. Human transport has been shown to promote artificial connectivity within introduced ranges (Voisin et al. 2005; Lacoursière-Roussel et al. 2012; Reem et al. 2013), and is expected to affect native species associating with the same vectors in a similar way. Only a single previous study has addressed human-mediated gene flow between native populations of a contemporary invader, and found that anthropogenic activities have altered genetic structure (Hudson et al. 2016). While modern introductions offer clues that processes expected to shift the geographic scale of diversification from regional to global are underway, further studies of historical introductions across different time scales promise to fill in gaps in our understanding of how our movement impacts patterns of global biodiversity.

An unfortunate corollary to the extensive mixture between native populations we report here is that human-mediated movement could threaten to obscure ancestral genetic structure arising from natural processes, erasing the record of demographic history

written in DNA. The threat that extant vectors of human-mediated dispersal pose to native patterns of genetic diversity likely depends on the interplay of population sizes, propagule pressure, and both the geographic and temporal patterns of human activities that enable dispersal. Based on previously reported genetic patterns, we expected three factors to cause higher genetic differentiation amongst European populations compared to North American populations: 1) isolation-by-distance (e.g. Launey et al. 2002; Bockelmann et al. 2003) and genetic discontinuities (e.g. Roman and Palumbi 2004) between fragmented habitats along a larger European coast; 2) geographic barriers to dispersal between islands and continental Europe (e.g. Exadactylos et al. 2003; Panova et al. 2011); and 3) isolation between refugia during Pleistocene glaciations in Europe (e.g. Jolly et al. 2006). However, we did not detect a significant genetic between Europe and North America, with little evidence for factors shaping population genetic patterns in either species. While we detected weak isolation-by-distance in both species, it only became apparent over distances greater than 600 km. This casts doubt over whether the apparent isolation-by-distance reflects regional barriers to dispersal, ancestral regional structure, or has begun to emerge since semi-dry ballast became defunct. In *H. diversicolor* we found two Irish populations and a group of two populations from the Bay of Biscay were diverged from other European populations by a similar degree to that observed between North America and Europe. Once again, it is ambiguous whether this pattern reflects historic processes or has emerged due to regional barriers to natural gene flow since human-mediated gene flow ceased. Our analyses of simulated data further highlight the problem that human-mediated gene flow amongst native populations causes in inferring natural evolutionary processes. These findings lend credence to ethical

concerns that moving species within their native ranges may erode naturally arising genetic structure, potentially precluding the inference of past demography from genetic data and disrupting local adaptation (Crozier and Schulte-Hostedde 2015). An interesting consequence of these patterns is the inability to correctly reconstruct colonization and introduction pathways (Figure 4.3), which may pose a major challenge for invasion genetics research in the future if global human activity continues the current trend of increasing in breadth and magnitude.

Between native and introduced ranges, cessation of gene flow is predicted to promote divergence through neutral (i.e. genetic drift) or deterministic (i.e. selection) processes that ultimately lead to reproductive isolation and the formation of new species (Coyne and Orr 2004). Evidence of rapid adaptation implies that the diversification process is underway in many introduced species, but whether this generally leads to long-term divergence across the genome cannot be determined from recent introductions (Vellend et al. 2007). We find that isolation between introduced populations of *C. volutator* and *H. diversicolor* and their native European range has led to genome-wide divergence in both species. The low number of variants detected only in the introduced range suggests that this does not result primarily from new mutations, but rather divergent sorting of ancestral polymorphisms via genetic drift or selection. Divergence between ranges likely results from both bottlenecks that occurred during the introduction process and from introduced populations preserving ancestral lineages in the native range that later became homogenized with other native lineages, although how much each factor contributes is uncertain due to the unknown timing and source or sources of introduction. We refrain from speculating on the role of selection because our simulations

show the demographic changes associated with introduction can cause large allele frequency changes between populations that preclude inference of selection using frequency-based scans. Our simulations based on strictly neutral evolution and evidence from other historical introductions (Baker and Moeed 1987; Wang et al. 2003; Montesinos et al. 2012; Shultz et al. 2016) demonstrate neutral processes are sufficient for divergence to be more pronounced between introduced populations and their sources than amongst native populations. Regardless of the role of selection, our results clearly support the hypothesis that the formation of new populations via human-mediated dispersal is a major factor in shaping species' evolutionary trajectories.

Our results demonstrate that human-mediated dispersal disrupts genetic structure within native ranges while establishing new isolated populations that follow independent evolutionary trajectories from their sources. The consistency between genome-wide patterns in *C. volutator* and *H. diversicolor* demonstrates that vectors of human-mediated dispersal can have repeatable impacts on members of the same community, suggesting changes to the geographic scale of evolution are tied to transport networks and propagule pressure particular to the vectors with which species associate. We consider human-mediated dispersal to be a double-edged sword, with its net effect on diversification rates dependent on the strength and regularity with which each edge strikes: human activity can facilitate gene flow that erodes ancestral genetic structure across broad spatial scales, but can also preserve ancestral diversity and promote diversification by establishing new populations beyond the limits of natural dispersal. This shift of the spatial scale of diversification from regional to global is unprecedented, and suggests that the erosion of

ancestral patterns of genetic structure will become an emerging concern for species associated with human activity as it increases in magnitude and reach.

Acknowledgements

We acknowledge NSERC for funding provided to JAA, ACEnet and Compute Canada for computational resources and funding to ALE, the Eaton Foundation for funding to ALE, and the New Brunswick Wildlife Trust Fund for funding to JAA and ALE. We thank Jin-Hong Kim and Jana Grütner for help developing ddRADseq methods, Jim Provan for hosting ALE during sampling in Europe, and Spencer Barrett for comments and suggestions on the manuscript.

References

Allen, J. R. L., & Haslett, S. K. (2002). Buried salt-marsh edges and tide-level cycles in the mid-Holocene of the Caldicot Level (Gwent), South Wales, UK. *The Holocene*, *12*(3), 303-324.

Alvarez, N., Hossaert-McKey, M., Restoux, G., Delgado-Salinas, A., & Benrey, B. (2007). Anthropogenic effects on population genetics of phytophagous insects associated with domesticated plants. *Evolution*, *61*(12), 2986-2996.

Ansorge, J., Frenzel, P., & Thomas, M. (2011). Cogs, sand and beer–a palaeontological analysis of medieval ballast sand in the harbour of Wismar (southwestern Baltic sea coast, germany). *Umweltarchäologie–Naturkatastrophen und Umweltwandel im archäologischen Befund*, 161-73.

Aronesty, E. (2013). Comparison of sequencing utility programs. *The Open Bioinformatics Journal*, *7*(1), 1-8.

Baker, H. G., & Stebbins, G. L. (1965). Genetics of colonizing species, proceedings. In *International Union of Biological Sciences Symposia on General Biology 1964: Asilomar, Calif.)*. Academic Press.

Baker, A. J., & Moeed, A. (1987). Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). *Evolution*, *41*(3), 525-538.

Bilton, D. T., Paula, J., & Bishop, J. D. D. (2002). Dispersal, genetic differentiation and speciation in estuarine organisms. *Estuarine, Coastal and Shelf Science*, *55*(6), 937-952.

Bockelmann, A. C., Reusch, T. B., Bijlsma, R., & Bakker, J. P. (2003). Habitat differentiation vs. isolation-by-distance: the genetic population structure of *Elymus athericus* in European salt marshes. *Molecular Ecology*, *12*(2), 505-515.

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *The Quarterly Review of Biology*, *74*(1), 21-45.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Bouchemousse, S., Bishop, J. D., & Viard, F. (2016). Contrasting global genetic patterns in two biologically similar, widespread and invasive *Ciona* species (Tunicata, Ascidiacea). *Scientific Reports*, *6*, srep24875.

Buller, A. T., & McManus, J. (1972). *Corophium* burrows as environmental indicators of Quaternary estuarine sediments of Tayside. *Scottish Journal of Geology*, *8*(2), 145-150.

Carlton, J. T. (1996). Biological invasions and cryptogenic species. *Ecology*, *77*(6), 1653-1655.

Carlton, J. T. (2003). Community assembly and historical biogeography in the North Atlantic Ocean: the potential role of human-mediated dispersal vectors. In *Migrations and Dispersal of Marine Organisms* (pp. 1-8). Springer, Dordrecht.

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

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276.

Colautti, R. I., & Barrett, S. C. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, *342*(6156), 364-366.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sunderland, MA.

Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular Ecology*, *24*(9), 2212-2225.

Crozier, G. K. D., & Schulte-Hostedde, A. I. (2015). Towards improving the ethics of ecological research. *Science and Engineering Ethics*, *21*(3), 577-594.

Dabney, A., Storey, J. D., & Warnes, G. R. (2010). qvalue: Q-value estimation for false discovery rate control. *R Package Version*, *1*(0).

Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PloS one*, *8*(12), e85024.

Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17*(1), 431-449.

Einfeldt, A. L., Doucet, J. R., & Addison, J. A. (2014). Phylogeography and cryptic introduction of the ragworm *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. *Invertebrate Biology*, *133*(3), 232-241.

Einfeldt, A. L., & Addison, J. A. (2015). Anthropocene invasion of an ecosystem engineer: resolving the history of *Corophium volutator* (Amphipoda: Corophiidae) in the North Atlantic. *Biological Journal of the Linnean Society*, *115*(2), 288-304.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what?. *Molecular Ecology*, *19*(19), 4113-4130.

Exadactylos, A., Geffen, A. J., Panagiotaki, P., & Thorpe, J. P. (2003). Population structure of Dover sole Solea solea: RAPD and allozyme data indicate divergence in European stocks. *Marine Ecology Progress Series*, *246*, 253-264.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973-983.

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

Grosberg, R., & Cunningham, C. W. (2001). Genetic structure in the sea. *Marine community ecology. Sinauer, Sunderland, MA*, 61-84.

Haydar, D. (2012). What is natural? The scale of cryptogenesis in the North Atlantic Ocean. *Diversity and Distributions*, *18*(2), 101-110.

Hansen, B., & Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, *482*(7385), 357-362.

Hewitt, C. L., Gollasch, S., & Minchin, D. (2009). The vessel as a vector–biofouling, ballast water and sediments. *Biological Invasions in Marine Ecosystems*, 117-131.

Hudson, J., Viard, F., Roby, C., & Rius, M. (2016). Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. *Biology Letters*, *12*(10), 20160620.

Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254-267.

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., & McGuire, M. (2010). A meta-analysis of isolation by distance: relic or reference standard for landscape genetics?. *Ecography*, *33*(2), 315-320.

Jolly, M. T., Viard, F., Gentil, F., Thiébaut, E., & Jollivet, D. (2006). Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology*, *15*(7), 1841-1855.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.

Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, *5*(1), e8594.

Lacoursière-Roussel, A. N. A. Ï. S., Bock, D. G., Cristescu, M. E., Guichard, F., Girard, P., Legendre, P., & McKindsey, C. W. (2012). Disentangling invasion processes in a dynamic shipping–boating network. *Molecular Ecology*, *21*(17), 4227-4241.

Launey, S., Ledu, C., Boudry, P., Bonhomme, F., & Naciri-Graven, Y. (2002). Geographic structure in the European flat oyster (*Ostrea edulis L.*) as revealed by microsatellite polymorphism. *Journal of Heredity*, *93*(5), 331-351.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, *17*(8), 386-391.

Legendre, P., & Fortin, M. J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, *10*(5), 831-844.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67-77.

Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., & Viard, F. (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, *89*(sp11).

Marchini, A., & Cardeccia, A. (2017). Alien amphipods in a sea of troubles: cryptogenic species, unresolved taxonomy and overlooked introductions. *Marine Biology*, *164*(4), 69.

McGrail, S. E. A. N. (1989). The shipment of traded goods and of ballast in antiquity. *Oxford Journal of Archaeology*, *8*(3), 353-358.

Montesinos, D., Santiago, G., & Callaway, R. M. (2012). Neo-allopatry and rapid reproductive isolation. *The American Naturalist*, *180*(4), 529-533.

Moran, E. V., & Alexander, J. M. (2014). Evolutionary responses to global change: lessons from invasive species. *Ecology Letters*, *17*(5), 637-649.

Panova, M., Blakeslee, A. M., Miller, A. W., Mäkinen, T., Ruiz, G. M., Johannesson, K., & André, C. (2011). Glacial history of the North Atlantic marine snail, Littorina saxatilis, inferred from distribution of mitochondrial DNA lineages. *PLoS One*, *6*(3), e17511.

Peer, D. L., Linkletter, L. E., & Hicklin, P. W. (1986). Life history and reproductive biology of *Corophium volutator* (Crustacea: Amphipoda) and the influence of shorebird predation on population structure in Chignecto Bay, Bay of Fundy, Canada. *Netherlands Journal of Sea Research*, *20*(4), 359-373.

Reem, E., Mohanty, I., Katzir, G., & Rinkevich, B. (2013). Population genetic structure and modes of dispersal for the colonial ascidian *Botryllus schlosseri* along the Scandinavian Atlantic coasts. *Marine Ecology Progress Series*, *485*, 143-154.

Roman, J. O. E., & Palumbi, S. R. (2004). A global invader at home: population structure of the green crab, *Carcinus maenas*, in Europe. *Molecular Ecology*, *13*(10), 2891-2898.

Ruiz, G. M., Fofonoff, P. W., Carlton, J. T., Wonham, M. J., & Hines, A. H. (2000). Invasion of coastal marine communities in North America: apparent patterns, processes, and biases. *Annual Review of Ecology and Systematics*, *31*(1), 481-531.

Scaps, P. (2002). A review of the biology, ecology and potential use of the common ragworm *Hediste diversicolor* (OF Müller) (Annelida: Polychaeta). *Hydrobiologia*, *470*(1-3), 203-218.

Seebens, H., Gastner, M. T., & Blasius, B. (2013). The risk of marine bioinvasion caused by global shipping. *Ecology Letters*, *16*(6), 782-790.

Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M. & Bacher, S. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, *8*, 14435.

Shultz, A. J., Baker, A. J., Hill, G. E., Nolan, P. M., & Edwards, S. V. (2016). SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorhous mexicanus*). *Ecology and Evolution*, *6*(20), 7475-7489.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457-462.

Steel, D. (1832). *The ship-master's assistant, and owner's manual.[With]*.

Streif, H. (1972). The results of stratigraphical and facial investigations in the coastal Holocene of Woltzeten/Ostfriesland, Germany. *GFF*, *94*(2), 281-299.

Vellend, M., Harmon, L. J., Lockwood, J. L., Mayfield, M. M., Hughes, A. R., Wares, J. P., & Sax, D. F. (2007). Effects of exotic species on evolutionary diversification. *Trends in Ecology & Evolution*, *22*(9), 481-488.

Virgilio, M., Fauvelot, C., Costantini, F., Abbiati, M., & Backeljau, T. (2009). Phylogeography of the common ragworm *Hediste diversicolor* (Polychaeta: Nereididae) reveals cryptic diversity and multiple colonization events across its distribution. *Molecular Ecology*, *18*(9), 1980-1994.

Voisin, M., Engel, C. R., & Viard, F. (2005). Differential shuffling of native genetic diversity across introduced regions in a brown alga: aquaculture vs. maritime traffic effects. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5432-5437.

Vonlanthen, P., Bittner, D., Hudson, A.G., Young, K.A., Müller, R., Lundsgaard-Hansen, B., Roy, D., Di Piazza, S., Largiadèr, C.R., & Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, *482*(7385), 357-362.

Wang, Zhenshan, et al. (2003). Reconciling actual and inferred population histories in the house finch (*Carpodacus mexicanus*) by AFLP analysis." *Evolution* 57.12: 2852-2864.

Wares, J. P. (2005). Mechanisms that drive evolutionary change. Insights from species introduction and invasions. *Species Invasions: Insights into Ecology, Evolution, and Biogeography*, 229-257.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F ST. *The American Naturalist*, *186*(S1), S24-S36.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46.

Wood, J. L., Yates, M. C., & Fraser, D. J. (2016). Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, *9*(5), 640-657.

Wright, S (1943) Isolation by distance. *Genetics, 28*, 114–138

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2013). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614-620.

Zhang, B., Edwards, O., Kang, L., & Fuller, S. (2014). A multi-genome analysis approach enables tracking of the invasion of a single Russian wheat aphid (*Diuraphis noxia*) clone throughout the New World. *Molecular Ecology*, *23*(8), 1940-1951.

Table 4.1 Population genetic diversity in *Corophium volutator* and *Hediste diversicolor*. Site codes for locations sampled, country, number of individuals sequenced (n), observed heterozygosity (H_o), and expected heterozygosity (H_e), and private SNPs with minor allele frequency > 0.01.

Corophium volutator (4870 SNPs, maf>0.01)

Population	Country	n	H_o	H_e	Private SNPs
Europe		**79**	**0.13150**	**0.16283**	**2690**
GIR	France	6	0.12911	0.12950	45
LOI	France	6	0.12405	0.12468	33
MSM	France	6	0.13507	0.13455	30
LAV	France	6	0.12783	0.13026	13
BAT	Netherlands	6	0.12777	0.13057	10
HEL	Germany	5	0.12229	0.12498	11
CUX	Germany	6	0.13003	0.13465	11
MAR	Denmark	6	0.13375	0.13322	12
BAH	Denmark	5	0.11102	0.11397	36
PIL	UK	6	0.13854	0.13791	30
THA	UK	6	0.14735	0.14419	28
ALK	UK	6	0.14380	0.14242	26
ROS	Ireland	6	0.13646	0.13259	58
BAL	Ireland	3	0.13789	0.12383	34
North America		**42**	**0.11060**	**0.12770**	**267**
FMS	USA	6	0.12118	0.10658	2
LCN	USA	6	0.11488	0.10380	7
WAL	USA	6	0.11458	0.10742	7
LBS	USA	6	0.10831	0.10617	11
POC	Canada	6	0.08322	0.08634	33
AVE	Canada	6	0.11896	0.11378	10
PCN	Canada	6	0.11376	0.11202	5
All		**12**	**0.12359**	**0.16234**	

Hediste diversicolor (3820 SNPs, maf>0.01)

Population	Country	n	H_o	H_e	Private SNPs
Europe		**67**	**0.14042**	**0.17884**	**2235**
GIR	France	6	0.16790	0.15915	201
PAI	France	6	0.14343	0.14283	105
LAV	France	4	0.15262	0.14685	21
NIE	Belgium	5	0.13732	0.12887	33
BAT	Netherlands	6	0.14861	0.13874	36
BUS	Germany	6	0.13258	0.12906	26
MAR	Denmark	5	0.13665	0.12491	26
PIL	UK	6	0.14007	0.13434	57
KNL	UK	6	0.13487	0.13093	31
BAN	Ireland	5	0.14714	0.15210	20
BLK	Ireland	6	0.14688	0.13573	67
BLS	Ireland	6	0.12223	0.10984	76
North America		**11**	**0.11609**	**0.12351**	**117**
SJS	Canada	5	0.11725	0.10372	41
AVE	Canada	3	0.12308	0.10221	26
GAN	Canada	3	0.11069	0.09247	29
All		**78**	**0.13666**	**0.17722**	

Table 4.2 False positive rates of selection scans on simulated data. Natural (N_m) and human-mediated (H_m) migration rates, number of SNPs detected in 10,000 simulated loci of 85bp each, number of outliers detected, and false positive rates of outlier detection using PCAdapt.

Simulated data set	Natural migration N_M	Human-mediated migration H_M	Number of SNPs	Number of outliers	False positive rate
i	5E-06	5E-07	9452	4326	0.4577
ii	5E-06	5E-06	9493	3747	0.3947
iii	5E-06	5E-05	9101	1954	0.2147
iv	5E-05	5E-06	8001	2561	0.3201
v	5E-05	5E-05	7770	1502	0.1933
vi	5E-05	5E-04	7479	2355	0.3149
vii	5E-04	5E-05	6556	858	0.1309
viii	5E-04	5E-04	6445	1553	0.2410
ix	5E-04	5E-03	6400	1248	0.1950

Table 4.3 Parasite and algal genomes filtered from genomic sequence data. Species grouped by databases constructed with Kraken, genome size, and GenBank accessions.

Organism	Size (MB)	Genome accession
Trematodes	**7066**	
Clonorchis sinensis	547	GCA_000236345
Fasciola hepatica	1275	GCA_000947175
Opisthorchis viverrini	620	GCF_000715545
Schistosoma haematobium	376	GCF_000699445
Schistosoma japonicum	403	GCA_000151775
Schistosoma mansoni	365	GCA_000237925
Dicrocoelium dendriticum	548	GCA_000950715
Echinostoma caproni	835	GCA_000950555
Schistosoma curassoni	344	GCA_000951415
Schistosoma margrebowiei	367	GCA_000951435
Schistosoma mattheei	341	GCA_000951455
Schistosoma rodhaini	343	GCA_000951475
Trichobilharzia regenti	702	GCA_000950905
Nematodes	**6855**	
Ancylostoma ceylanicum	313	GCA_000402015
Ancylostoma duodenale	333	GCA_000816745
Angiostrongylus cantonensis	6	GCA_000950995
Ascaris suum	263	GCA_000187025
Brugia malayi	94	GCF_000002995
Brugia pahangi	84	GCA_001280985
Bursaphelenchus xylophilus	73	GCA_000231135
Caenorhabditis angaria	80	GCA_000165025
Caenorhabditis brenneri	190	GCA_000143925
Caenorhabditis briggsae	108	GCF_000004555
Caenorhabditis elegans	100	GCF_000002985
Caenorhabditis japonica	166	GCA_000147155
Caenorhabditis remanei	145	GCF_000149515
Caenorhabditis tropicalis	79	GCA_000186765
Dictyocaulus viviparus	161	GCA_000816705
Dirofilaria immitis	85	GCA_001077395
Ditylenchus destructor	111	GCA_001579705
Elaeophora elaphi	1	GCA_000951195
Globodera pallida	124	GCA_000724045
Haemonchus contortus	320	GCA_000469685
Heterodera glycines	82	GCA_000150805
Heterorhabditis bacteriophora	77	GCA_000223415
Loa loa	96	GCF_000183805
Meloidogyne floridensis	97	GCA_000751915
Meloidogyne hapla	53	GCA_000172435
Meloidogyne incognita	82	GCA_000180415
Necator americanus	244	GCF_000507365
Oesophagostomum dentatum	443	GCA_000797555
Onchocerca ochengi	95	GCA_001077375
Onchocerca volvulus	96	GCA_000499405
Oscheius	117	GCA_001513535
Panagrellus redivivus	65	GCA_000341325
Pristionchus pacificus	133	GCA_000180635
Romanomermis culicivorax	323	GCA_001039655
Rotylenchulus reniformis	314	GCA_001026735
Steinernema carpocapsae	86	GCA_000757645
Steinernema feltiae	83	GCA_000757705
Steinernema glaseri	93	GCA_000757755
Steinernema monticolum	89	GCA_000505645
Steinernema scapterisci	80	GCA_000757745
Strongyloides ratti	43	GCA_001040885
Subanguina moxae	90	GCA_000981365
Toxocara canis	317	GCA_000951555

Trichinella	49	GCA_001447505
Trichinella britovi	52	GCA_001447585
Trichinella murrelli	49	GCA_001447425
Trichinella nativa	48	GCA_001447565
Trichinella nelsoni	47	GCA_001447455
Trichinella papuae	47	GCA_001447755
Trichinella patagoniensis	50	GCA_001447655
Trichinella pseudospiralis	49	GCA_001447445
Trichinella spiralis	64	GCF_000181795
Trichinella zimbabwensis	51	GCA_001447665
Trichuris muris	84	GCA_000612645
Trichuris suis	64	GCA_000797535
Trichuris trichiura	75	GCA_000613005
Wuchereria bancrofti	90	GCA_001555675
Myxozoans	**423**	
Enteromyxum leei	68	GCA_001455295
Kudoa iwatai	31	GCA_001407335
Sphaeromyxa zaharoni	174	GCA_001455285
Thelohanellus kitauei	150	GCA_000827895
Microsporidians	**166**	
Anncaliia algerae	12	GCA_000385855
Edhazardia aedis	51	GCA_000230595
Encephalitozoon cuniculi	2	GCF_000091225
Encephalitozoon hellem	2	GCF_000277815
Encephalitozoon intestinalis	2	GCF_000146465
Encephalitozoon romaleae	2	GCF_000280035
Enterocytozoon bieneusi	4	GCF_000209485
Hamiltosporidium tvaerminnensis	13	GCA_000180835
Mitosporidium daphniae	6	GCF_000760515
Nematocida parisii	4	GCF_000250985
Nematocida sp. 1	4	GCA_000738915
Nosema apis	9	GCA_000447185
Nosema bombycis	16	GCA_000383075
Nosema ceranae	8	GCF_000182985
Ordospora colligata	2	GCF_000803265
Pseudoloma neurophilia	5	GCA_001432165
Spraguea lophii	5	GCA_000430065
Trachipleistophora hominis	8	GCA_000316135
Vavraia culicis	6	GCF_000192795
Vittaforma corneae	3	GCF_000231115
Cryptococcus Yeasts	**185**	
Cryptococcus albidus	21	GCA_001468955
Cryptococcus bestiolae	24	GCA_000512585
Cryptococcus curvatus	16	GCA_001028165
Cryptococcus dejecticola	24	GCA_000512565
Cryptococcus flavescens	23	GCA_000442785
Cryptococcus gattii	18	GCF_000185945
Cryptococcus laurentii	19	GCA_000738825
Cryptococcus neoformans	19	GCF_000149245
Cryptococcus pinus	21	GCA_000512605
Apicomplexans	**1255**	
Ascogregarina taiwanensis	6	GCA_000172235
Babesia bigemina	14	GCF_000981445
Babesia bovis	8	GCF_000165395
Babesia divergens	11	GCA_001077455
Cryptosporidium	10	GCA_000831705
Cryptosporidium baileyi	8	GCA_001593455
Cryptosporidium hominis	9	GCF_000006425
Cryptosporidium meleagridis	9	GCA_001593445
Cryptosporidium muris	9	GCF_000006515
Cryptosporidium parvum	9	GCF_000165345
Cyclospora cayetanensis	45	GCA_001305735
Eimeria acervulina	46	GCF_000499425

Eimeria brunetti	67	GCA_000499725
Eimeria maxima	46	GCF_000499605
Eimeria mitis	60	GCF_000499745
Eimeria necatrix	55	GCF_000499385
Eimeria nieschulzi	63	GCA_000826945
Eimeria praecox	60	GCA_000499445
Eimeria tenella	52	GCF_000499545
Gregarina niphandrodes	14	GCF_000223845
Hammondia hammondi	68	GCF_000258005
Neospora caninum	58	GCF_000208865
Plasmodium berghei	18	GCF_000005395
Plasmodium chabaudi	17	GCF_000003075
Plasmodium coatneyi	28	GCA_000725905
Plasmodium cynomolgi	26	GCF_000321355
Plasmodium falciparum	23	GCF_000002765
Plasmodium fragile	26	GCF_000956335
Plasmodium gaboni	16	GCA_000576715
Plasmodium inui	27	GCF_000524495
Plasmodium knowlesi	23	GCF_000006355
Plasmodium reichenowi	24	GCF_000723685
Plasmodium vinckei	18	GCF_000709005
Plasmodium vivax	27	GCF_000002415
Plasmodium yoelii	23	GCF_000003085
Sarcocystis neurona	124	GCA_000875885
Theileria annulata	8	GCF_000003225
Theileria equi	12	GCF_000342415
Theileria orientalis	9	GCF_000740895
Theileria parva	8	GCF_000165365
Toxoplasma gondii	69	GCF_000006565
Oomycetes Yeasts	**2655**	
Albugo candida	33	GCA_001306775
Aphanomyces astaci	76	GCF_000520075
Aphanomyces invadans	71	GCF_000520115
Hyaloperonospora arabidopsidis	78	GCA_001414525
Phytophthora agathidicida	37	GCA_001314435
Phytophthora alni	236	GCA_000439335
Phytophthora cambivora	231	GCA_000443045
Phytophthora capsici	56	GCA_000325885
Phytophthora cinnamomi	54	GCA_001314365
Phytophthora cryptogea	103	GCA_000468175
Phytophthora fragariae	74	GCA_000686205
Phytophthora infestans	229	GCF_000142945
Phytophthora kernoviae	37	GCA_000333075
Phytophthora lateralis	60	GCA_000318465
Phytophthora multivora	40	GCA_001314345
Phytophthora nicotianae	71	GCA_001483015
Phytophthora parasitica	82	GCF_000247585
Phytophthora pinifolia	132	GCA_000500225
Phytophthora pisi	59	GCA_000751395
Phytophthora pluvialis	54	GCA_001314425
Phytophthora ramorum	67	GCA_000336535
Phytophthora rubi	48	GCA_000687305
Phytophthora sojae	83	GCF_000149755
Phytophthora taxon totara	56	GCA_001314925
Phytopythium vexans	34	GCA_000387545
Plasmopara halstedii	75	GCA_900000015
Pseudoperonospora cubensis	64	GCA_000252605
Pythium aphanidermatum	36	GCA_000387445
Pythium arrhenomanes	45	GCA_000387505
Pythium insidiosum	53	GCA_001029375
Pythium irregulare	43	GCA_000387425
Pythium iwayamai	43	GCA_000387465
Pythium oligandrum	36	GCA_001573145

Pythium ultimum	45	GCA_000387525
Saprolegnia diclina	63	GCF_000281045
Saprolegnia parasitica	53	GCF_000151545
Algae	**2097**	
Ectocarpus siliculosus	196	GCA_000310025
Saccharina japonica	543	GCA_000978595
Auxenochlorella protothecoides	23	GCF_000733215
Chlamydomonas reinhardtii	120	GCF_000002595
Chlorella pyrenoidosa	57	GCA_001430745
Chlorella variabilis	46	GCF_000147415
Chlorella vulgaris	37	GCA_001021125
Coccomyxa	12	GCA_001244535
Coccomyxa subellipsoidea C-169	49	GCF_000258705
Cymbomonas tetramitiformis	281	GCA_001247695
Gonium pectorale	149	GCA_001584585
Helicosporidium	12	GCA_000690575
Micromonas	20	GCF_000090985
Micromonas pusilla	22	GCF_000151265
Monoraphidium neglectum	70	GCF_000611645
Ostreococcus 'lucimarinus'	13	GCF_000092065
Ostreococcus tauri	13	GCF_000214015
Picochlorum	13	GCA_000876415
Trebouxia gelatinosa	62	GCA_000818905
uncultured Bathycoccus	5	GCA_000259855
Volvox carteri	138	GCF_000143455
Chondrus crispus	105	GCF_000350225
Cyanidioschyzon merolae	17	GCF_000091205
Galdieria sulphuraria	14	GCF_000341285
Porphyridium purpureum	19	GCA_000397085
Heterococcus	61	GCA_000498555

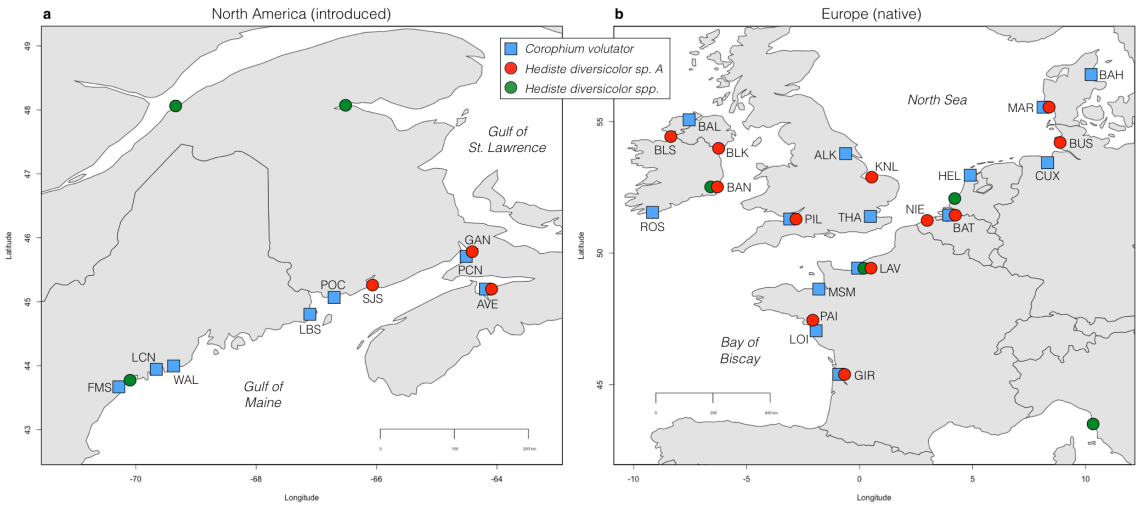


Figure 4.1 Distribution of introduced and native invertebrate populations. Specimens of *Corophium volutator* and morphologically identified *Hediste diversicolor* were collected from mudflats spanning their known ranges. a, Populations from the introduced range in North America. b, Populations from the native range in Europe.

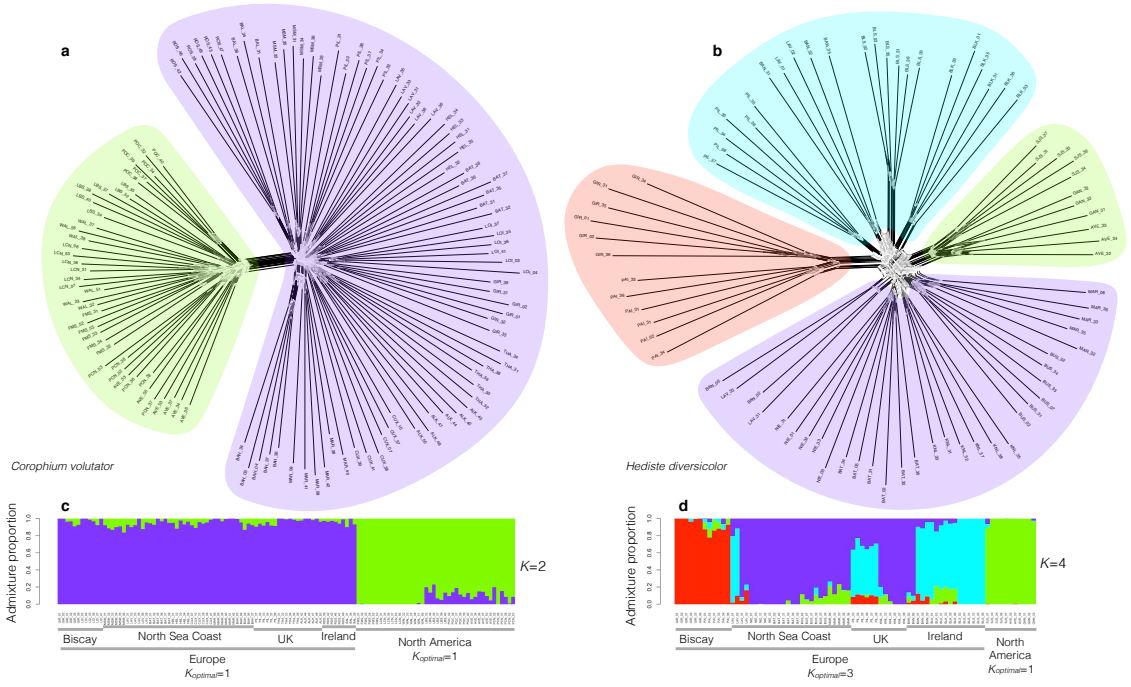


Figure 4.2 Genetic divergence between, and phylogenetic conflict within, introduced and native ranges. a-b, Bootstrapped neighbour-joining networks for *Corophium volutator* (4870 SNPs) and *Hediste diversicolor* (3820 SNPs) show clustering of individuals sampled from the same site, support for division between clusters of native vs. introduced ranges, and extensive reticulation between populations within each range. c-d, Assignment proportions for individuals to *K* optimal genotypic clusters according to minimum cross-entropy computed with SNMF support genetic subdivision between North America and Europe in both species, and regional genetic structure of *H. diversicolor* in Europe.

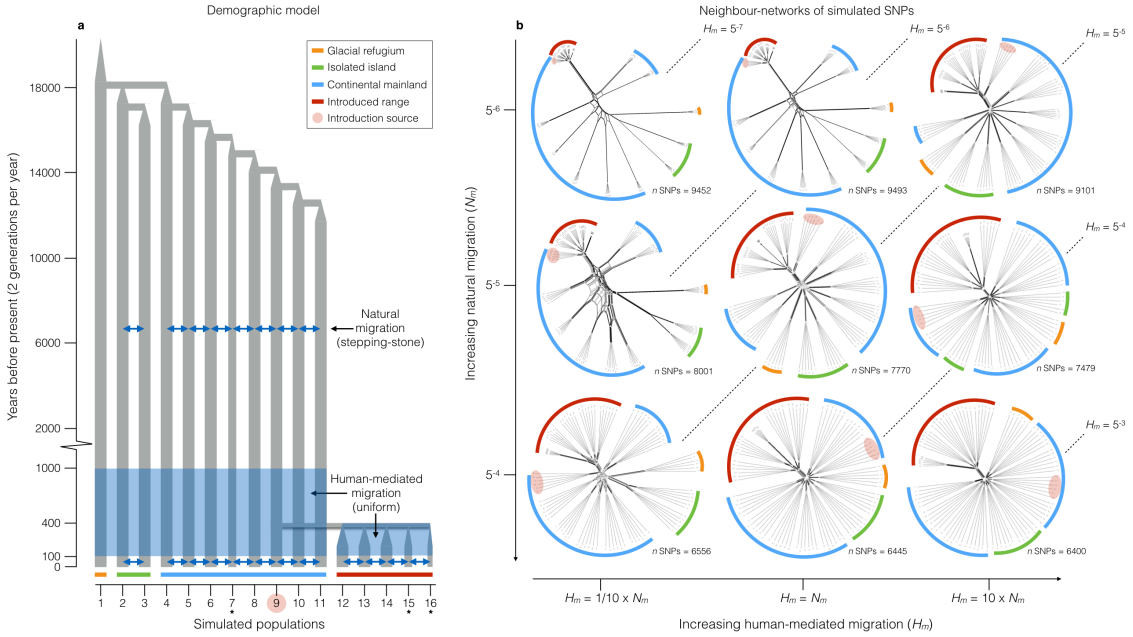


Figure 4.3 Human-mediated migration erases ancestral genetic structure. a, Demographic model of processes expected to shape diversity in North Atlantic intertidal species, used to simulate genetic data. The model incorporates population expansion and post-Pleistocene colonization from a glacial refugium, migration from natural dispersal (N_m) representing connectivity between adjacent populations in discontinuous habitat patches along a coastline, bottlenecks during introduction, and a temporary window of uniform human-mediated migration (H_m) among populations within each range. *Simulated populations 7 and 15 have 0.5 N_e, and population 16 has 0.1 N_e. b, Bootstrapped neighbor-networks of SNPs from simulated genetic data for different values of N_m and H_m show that increasing H_m increases reticulation between populations, precluding inference of ancestral demographic processes and recent introduction pathways.

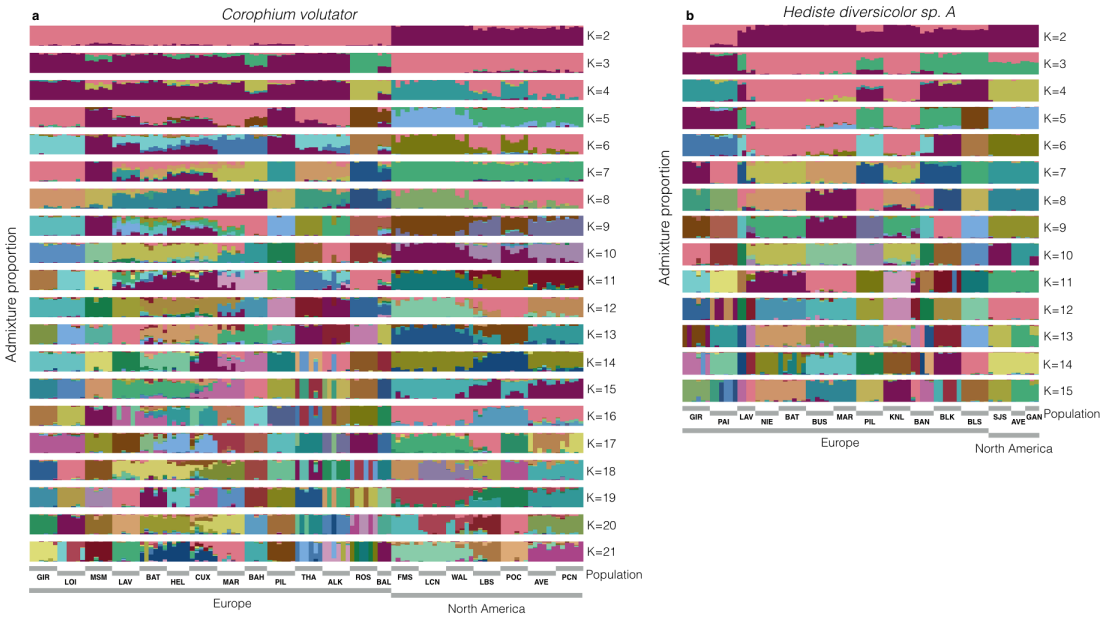


Figure 4.4 Hierarchical genetic structure. a-b, Assignment proportions for individuals to *K* optimal genotypic clusters according to minimum cross-entropy computed with SNMF show genetic subdivision at increasingly fine geographic scales as *K* is increased.

Chapter 5 – General discussion

The objectives of my thesis reflect a series of questions that were not formulated at the beginning of my dissertation, but rather grew from the unexpected results of each individual project along the way. Initially, my objective was to determine the geographic scale at which soft-sediment mudflats in the Bay of Fundy were ecologically connected by characterizing genetic structure in the most abundant invertebrate in these mudflats: *Corophium volutator*. I did this by sequencing mitochondrial DNA from individuals sampled from mudflats in two sub-basins of the Bay of Fundy. I detected low overall genetic diversity, genetic subdivision between the two basins, and no subdivision between sites within basins, providing evidence that the basins were both evolutionarily and ecologically isolated from each other. This motivated a number of practical questions about the system: What caused this separation? Was there subdivision at finer spatial scales that could not be detected due to the low overall diversity of the genetic markers I used? Why was overall genetic diversity so low? Where did *C. volutator* come from? And when? Was its history unique? While these questions are practical and system-specific, answering them provides insight into larger evolutionary questions about the processes underpinning current, and emerging, global patterns of biodiversity. In Chapters 2-4 of this thesis, I asked: 1) How does life-history interact with environmental features to shape patterns of connectivity? 2) What are the roles of natural and human-mediated dispersal in shaping species' distributions? 3) How are the evolutionary trajectories of species impacted by the changes to their inherent dispersal abilities brought on by human activity? The results of each chapter often necessitated re-interpretation of

both system-specific and general ecological and evolutionary assumptions. Together, these chapters provide insights into the natural and human-induced processes shaping evolution, improving the ability to predict emerging patterns of global biodiversity in a changing world.

My initial work characterizing subdivision between habitat patches in the Bay of Fundy using *C. volutator* demonstrated that apparently similar habitats and communities in the Bay of Fundy were in fact ecologically isolated from each other, potentially due to the activity of strong tidal currents within each of its two major branches. While *C. volutator* lacks a pelagic larval phase, passive dispersal within basins via currents could be made possible by vertical swimming behaviour during its juvenile and adult phases, which would allow it to move into the water column and drift with currents during tidal cycles (Drolet et al. 2012). Such passive adult dispersal is predicted to cause genetic admixture within basins and subdivision across other areas where currents split. In Chapter 2 I tested how hydrology influences population structure by assessing subdivision and gene flow with coalescent analyses of mitochondrial and nuclear DNA sequence markers across the North American range of *C. volutator*. I found that subdivision was generally concordant with divisions predicted by splits in major currents between basins, with the strongest divide occurring between the Bay of Fundy and Gulf of Maine. I also found that the magnitude of gene flow between basins was generally concordant with the direction of these currents. My results demonstrated that currents can shape evolution even in species that lack the capability of dispersing via a pelagic larval phase, suggesting that juvenile and adult behavior is an important but often overlooked

factor affecting demographic connectivity in estuarine and marine invertebrates. As most marine invertebrates disperse via passive means, my findings suggested that hydrological barriers may structure entire marine communities in the Northwest Atlantic over small spatial scales, and that the Bay of Fundy is a more isolated region than previously assumed[1].

Some aspects of the results presented in Chapter 2 were not satisfactorily explained by hydrological patterns alone. All populations showed low genetic diversity at all markers, calling to question whether more fine-scale demographic subdivision existed, but was not detectable using only a few genetic markers with low levels of diversity (which I addressed in Chapter 4). Having assumed that *C. volutator* is native to the Northwest Atlantic coast, and thus expecting to find higher genetic diversity in southern populations due to post-glacial recolonization from southern refugia, I found surprising low levels of genetic diversity in the southern populations. Furthermore, a lack of reciprocal monophyly across the Fundy/Maine divide suggested that a more complex evolutionary history than previously assumed may have shaped *C. volutator*'s distribution, but fell short of providing sufficient information to test different hypotheses about the species' origins. *C. volutator* has a large ecological impact in the Northwest Atlantic due to its abundance, re-engineering of intertidal sediment qualities, competition

[1] Since publishing these findings, further population genetic (Einfeldt et al. 2014), phylogeographic (Einfeldt et al. 2017), and biogeographic (Pappalardo et al. 2015) work has been found to support this hypothesis.

with other species, and role as a prey item for many invertebrate and vertebrate species. Resolving its natural history is thus important to interpret ecological dynamics in the North Atlantic, as this provides evolutionary context for the many direct and indirect interactions that *C. volutator* has with marine, benthic, and terrestrial communities.

In Chapter 3 I investigated the evolutionary origins of *C. volutator* across its entire range by comparing mitochondrial and nuclear sequence data. In the process of obtaining samples, I was faced with much uncertainty regarding the actual known range of *C. volutator*. I sampled extensively from the northernmost reported population in North America (Îles de la Madeleine in Québec) to the southernmost population (Strawberry Cove in Maine) and in seemingly suitable habitat as far south as Connecticut. However, the furthest south I was able to find *C. volutator* was at the Gilsland Farm Audubon Center in Falmouth Maine (only 27 km south of the previously reported southernmost population). I wrote to request samples from hundreds of biologists that had reported *C. volutator* in the North Sea, the Mediterranean, India, and Japan. From the specimens I received, and closer examinations by some of these biologists, all reports of the species outside of the Gulf of Maine, Bay of Fundy, and the Atlantic coast of Europe turned out to be misidentifications (and some samples from Europe turned out to be misidentifications, cryptic, or un-described species). This suggests a closer relationship between Europe and North America than had previously been assumed.

Using mitochondrial and nuclear sequence data, I found that North American diversity is a subsample from much more diverse European populations, consistent with colonization of North America from Europe. Coalescent-based model testing supported a

history of multiple introductions, which is unlikely under a model of natural colonization but plausible if the species was introduced by human activity. I combined my results with biogeographic evidence that *C. volutator* was introduced by historic human activity rather than natural dispersal or recent activity[2]. This research suggested that the marine coastal environment of the Northwest Atlantic, and likely other systems, may harbor many more undocumented introductions and have been more impacted by humans than previously thought. The ecological repercussions of *C. volutator*'s recent history are still emerging. My findings added evidence supporting the hypothesis that *C. volutator* was not as essential a prey item as once thought, particularly for semi-palmated sandpipers (*Calidris pusilla*) that were present in the Northwest Atlantic before *C. volutator*'s introduction. Recent work led by Diana Hamilton suggests that the diet of semi-palmated sandpipers broadly tracks prey availability (Quinn et al. 2017) and that their tongues may be adapted to harvesting biofilm (personal communication). While the arrival of *C. volutator* in the Northwest Atlantic likely disrupted long-standing ecological relationships between native species, it is not clear whether this has been to the benefit or detriment of individual species.

---

[2] Upon submitting this chapter for publication, I found that I was not the first to suspect this species was introduced, but merely the first to present convincing genetic evidence. Based on natural history, taxonomy, systematics, geography, and geological history, John Chapman had made a similar argument over a decade earlier, but, "The idea that *C. volutator* (Cv) was introduced was strongly (vehemently?) doubted when I was working on it." (Chapman 2014 personal communication). Except for a single conference proceeding, none of Chapman's work on the system reached publication.

Despite resolving this contentious case of history, genetic data from European populations did not reveal sufficiently fine-scale population structure to identify the source(s) from which North American populations descended. At the time of publication I (naively) attributed this to the technical limitations of using a small number of genetic markers to resolve spatial genetic structure. To determine how human activity shaped colonization pathways and evolution in the North Atlantic, in Chapter 4 I thoroughly sampled the North American and European coasts for *C. volutator* and *H. diversicolor*, a co-occurring annelid with similar life-history traits, and characterized SNPs across the genomes of both species. A common assumption in invasion genetics is that native ranges are structured predominately by slow-acting and long-standing natural processes, and that introduced populations are more related to their sources than native populations generally are to each other. I thus expected to reconstruct colonization routes using genetic measures of relatedness between introduced populations and their source regions. This paradigm is based on the assumption that human-mediated dispersal occurs at sufficient strength to establish new populations but not to cause gene flow between established populations, which is likely true only when vector activity is novel. Over time vectors may become more active within native and introduced ranges than between them, homogenizing native patterns of genetic diversity but allowing introduced populations to diverge from their sources. Historically introduced species can provide insight into how human-mediated dispersal shapes evolution after vectors have been active for longer periods of time, and how increasing rates of global human activity may be restructuring species' evolutionary trajectories.

In Chapter 4 I presented evidence that human-mediated dispersal changes the geographic scale at which evolution occurs, by simultaneously homogenizing native genetic diversity and establishing new isolated populations that follow independent evolutionary trajectories from their sources. I found that both *C. volutator* and *H. diversicolor* had undergone mild genetic bottlenecks during introduction, that populations were genetically distinct from each other, that introduced and native ranges were genetically divergent, and that there was extensive phylogenetic conflict between populations within each range. These results were incompatible with gene flow via natural dispersal alone, precluding the ability to reconstruct colonization pathways in this system. They also explained the inability of mitochondrial and nuclear sequence data to resolve phylogeographic patterns in Europe in Chapter 3; the lack of phylogeographic signal of glacial refugia was not an inadequacy of the genetic markers, but a result of human activity moving genetic diversity between native populations. However, the data presented in Chapter 3 were unable to resolve genetic structure at a fine enough geographic scale to adequately address this problem, demonstrating the improved power of inference granted by increasing the amount of genetic information analyzed.

The results of Chapter 4 add to a small but growing body of evidence suggesting that the rapidly increasing number of introduced species may be laying the foundation for an increase in the global rate of speciation, by effectively causing vicariance between introduced and native ranges (Montesinos et al. 2012; Thomas 2015). Studying the degree of reproductive isolation between historic invaders and their native congeners through crossing is likely to provide further insight into whether introductions contribute

to the generation of new taxa. However, whether the increase in divergence enabled by the establishment of new allopatric populations outweighs the divergence lost when native ranges are homogenized, and whether introduced populations stay isolated, remains to be seen. That human-mediated dispersal can facilitate gene flow amongst native populations is a new concern for researchers aiming to elucidate natural history from contemporary distributions of genetic variation, as human activity has the potential to erase natural patterns of genetic diversity and disrupt the stream of heredity that otherwise allows us peer into the history of the natural world. This may require a re-evaluation of inferred colonization pathways in systems where there is potential for human-mediated dispersal within native ranges. The ethical issues surrounding reshaping native genetic diversity have been expressed in the context of transplant experiments in ecological research (Crozier and Schulte-Hostedde 2015), but the wide-reaching nature of unintentional human-mediated dispersal suggests that this issue is an emerging concern that requires immediate attention in a broad range of systems.

Chapters 2-4 of this thesis progress from narrow to broad surveys of both geographic locations and genomic information sampled. However, increasing the number of markers analyzed did not preclude the major findings of the previous studies. The increased information instead offered additional insight into the system that the previous approaches had lacked the power to detect. Further increasing the scale of genomic data surveyed is likely to improve ability to differentiate individuals, and thus regional structure, but the types of processes that can be inferred may in some cases be limited

regardless of the amount of genetic information obtained. This is particularly true in the case of inferring selective processes.

Inference of selection using SNPs surveyed from anonymous genomic positions in species without an assembled genome resource is limited by reliance on changes in allele frequency that could arise due to neutral processes (e.g. Shultz et al. 2016). In Chapter 4, I use forward simulations to demonstrate that neutral processes can lead to allele frequencies that register as "selective", particularly when populations have vastly different effective population sizes. While SNPs provide fine-scale resolution of differentiation that is useful in answering many questions, the genealogical information conferred by DNA sequence data promises that genomic information will be superior in disentangling the roles of genetic drift and adaptive processes in evolution. While it is tempting to think that sequencing entire genomes may allow more robust expected patterns of selection to be tested, caution is still required when relying solely on genetic data. For instance, ratios of non-synonymous to synonymous substitutions in genes is a useful method of inferring selection between species, but at the population level these ratios are relatively insensitive to the strength of selection (Kryazhimskiy and Plotkin 2008). A general improvement to the ability of genomic surveys to detect selection is to observe genetic change of populations over time, using either repeated surveys (Weider et al. 1997; Wandeler et al. 2007) or formal experimentation (e.g. Tobler et al. 2013; Egan et al. 2015) to complement initial surveys. However, apparent selection based on inference from genomic data alone may be the byproduct of unobserved processes, and true tests of adaptation require linking genomic information to phenotype and fitness

(Barrett and Hoekstra 2011). As this is a conceptual limitation of using genetic data alone, more advanced sequencing technologies will likely not be able to replace the need for combining genetic information with other approaches in resolving the role of selection in evolution. In the future, population genomic data combined with ecological and environmental data in study designs with temporal or experimental components promises to provide improved ability to predict responses to changing ecological and environmental conditions.

The geographic extent of the surveys I present in this thesis increases with each chapter, but other factors are likely to impact evolution in intertidal environments over shorter distances than what I addressed. The distributions of macro-invertebrates in soft-sediments are structured, often over a scale of kilometers or meters, by environmental factors such as temperature, salinity, immersion time, current velocity, substrate type, food availability, turbidity, and interspecific interactions with prey, competitors, predators, and parasites (McLusky et al. 1993; Bocher et al. 2007). These same factors may shape patterns of genetic diversity in species whose distributions span either opposing extremes or continuums of such environmental stressors, especially in species capable of moving between and choosing environments suited to their innate tolerances. Characterizing a high degree of genomic content or targeting genes expected to be involved in tolerance to specific conditions will likely be a productive approach to investigating the role of environmental conditions in evolution at small geographic scales. For example, transcriptome sequencing of *H. diversicolor* shows that populations with increased and heritable tolerance to elevated copper concentrations have higher

expression levels of genes related to copper homeostasis than populations at uncontaminated sites (McQuillan et al. 2014). In Chapter 4 I showed that discrete patches of soft-sediment habitat are inhabited by genetically distinct populations of *C. volutator* and *H. diversicolor*, suggesting that environmental differences between mudflats, even at small spatial scales, may contribute to genetic divergence amongst populations. Ideally, future surveys of genetic variation in this system should consider both small-scale and distribution-wide processes as contributors to evolution.

In summary, the evolutionary patterns I present in this thesis provide an improved ability to predict emerging trends of biodiversity in a changing world. In Chapter 2 I demonstrated that currents can create population structure at finer spatial scales than previously assumed in marine environments, and are a major mechanism structuring connectivity amongst habitat patches. Climatic changes in the Northwest Atlantic are expected to change hydrological patterns in the near future (Hurrell et al. 2001; Drinkwater et al. 2009), which can thus be expected to change patterns of connectivity amongst marine and intertidal ecosystems. In Chapter 3 I showed that an ecologically important invertebrate was introduced to North America by early exploration and trade, demonstrating that the many interactions it has with native species are recent associations, as opposed to being the products of a long co-evolutionary history. Introduced species are increasing globally in number and in the breadth of the systems they impact, and my findings present an optimistic reminder that native species can adapt to, or even exploit, new arrivals. In Chapter 4 I presented human-mediated dispersal as an evolutionary force that reshapes the geographic scale at which diversification occurs, by

facilitating gene flow amongst otherwise isolated populations and creating new populations that then become isolated from their sources. As the reach and magnitude of human movement continues to increase, species associated with human vectors of transport can be expected to undergo similar changes to their evolutionary trajectories. While contemporary connectivity will continue to be affected by regional processes (such as currents), uncurbed human activity will likely disrupt diversification arising from barriers at regional scales while promoting the formation of new lineages at a global scale.

References

Barrett RD and Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature Reviews Genetics. 12: 767-780.

Bocher P, Piersma T, Dekinga A, Kraan C, Yates MG, Guyot T, Radenac G. 2007. Site- and species-specific distribution patterns of molluscs at five intertidal soft-sediment areas in northwest Europe during a single winter. Marine Biology. 151: 577-594.

Crozier GKD and Schulte-Hostedde AI. 2015. Towards improving the ethics of ecological research. Science and engineering ethics. 21: 577-594.

Drinkwater KF, Mueter F, Friedland KD, Taylor M, Hunt GL, Hare J, Melle W. 2009. Recent climate forcing and physical oceanographic changes in Northern Hemisphere regions: A review and comparison of four marine ecosystems. Progress in Oceanography. 81: 10-28.

Drolet D, Bringloe TT, Coffin MR, Barbeau MA, Hamilton DJ. 2012. Potential for between-mudflat movement and metapopulation dynamics in an intertidal burrowing amphipod. Marine Ecology Progress Series, 449, 197-209.

Egan SP, Ragland GJ, Assour L, Powell TH, Hood GR, Emrich S, Feder JL. 2015. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. Ecology letters. 18: 817-825.

Einfeldt AL, Doucet JR, and Addison JA 2014. Phylogeography and cryptic introduction of the ragworm *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. Invertebrate biology. 133: 232-241.

Einfeldt AL, Zhou F, and Addison JA. 2017. Genetic discontinuity in two high dispersal marine invertebrates in the northwest Atlantic. FACETS. 2: 160-177.

Hudson J, Viard F, Roby C, Rius M. 2016. Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. Biology letters. 12: 20160620.

Hurrell JW, Kushnir Y, and Visbeck M. 2001. The north Atlantic oscillation. Science. 291: 603-605.

Kryazhimskiy S, and Plotkin JB. 2008. The population genetics of dN/dS. PLoS genetics. 4: e1000304.

McLusky DS, Hull SC, and Elliott M. 1993. Variations in the intertidal and subtidal macrofauna and sediments along a salinity gradient in the upper Forth Estuary. Netherland Journal of Aquatic Ecology. 27: 101-109.

McQuillan, JS, Kille P, Powell K, Galloway TS. 2014. The regulation of copper stress response genes in the polychaete *Nereis diversicolor* during prolonged extreme copper contamination. Environmental science & technology. 48:, 13085-13092.

Montesinos D, Santiago G, and Callaway RM. 2012. Neo-allopatry and rapid reproductive isolation. The American Naturalist. 180: 529-533.

Pappalardo P, Pringle JM, Wares JP, and Byers JE. 2015. The location, strength, and mechanisms behind marine biogeographic boundaries of the east coast of North America. Ecography, 38(7), 722-731.

Quinn JT, Hamilton DJ, and Hebert CE. 2017. Fatty acid composition and concentration of alternative food of Semipalmated Sandpipers (*Calidris pusilla*) in the upper Bay of Fundy, Canada. Canadian Journal of Zoology, 95(8), 565-573.

Shultz AJ, Baker AJ, Hill GE, Nolan PM, Edwards SV. 2016. SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorhous mexicanus*). Ecology and Evolution. 6: 7475-7489.

Thomas CD. 2015. Rapid acceleration of plant speciation during the Anthropocene. Trends in ecology & evolution. 30: 448-455.

Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, Hermisson J, Schlötterer C. 2013. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. Molecular biology and evolution. 31: 364-375.

Wandeler P, Hoeck PE, and Keller LF. 2007. Back to the future: museum specimens in population genetics. Trends in ecology & evolution. 22: 634-642.

Weider LJ, Lampert W, Wessels M, Colbourne JK, Limburg P. 1997. Long–term genetic shifts in a microcrustacean egg bank associated with anthropogenic changes in the Lake Constance ecosystem. Proceedings of the Royal Society of London B: Biological Sciences. 264: 1613-1618.

Curriculum Vitae

Anthony Leon Einfeldt

University of British Columbia, BSc, 2008

Publications:

Peer-reviewed

Einfeldt AL, Zhou F, and Addison JA. 2017. Genetic discontinuity in two high dispersal marine invertebrates in the northwest Atlantic. Facets 2: 160-177. DOI 10.1139/facets-2016-0044.

Mlynarek JJ, Moffat CE, Edwards S, **Einfeldt AL**, Heustis A, Johns R, MacDonnell M, Pureswaran DS, Quiring DT, Shibel Z, and Heard SB. 2017. Enemy escape: a general phenomenon in a fragmented literature? Facets 2: 1-30. DOI 10.1139/facets-2017-0041.

Einfeldt AL and Addison JA. 2015. Anthropocene invasion of an ecosystem engineer: Resolving the history of *Corophium volutator* (Amphipoda: Corophiidae) in the North Atlantic. Biological Journal of the Linnean Society 115: 288-304.

Addison JA, **Einfeldt AL**, Kang NN and Walde SJ. 2015. Small-scale genetic structure in a stream dwelling caddisfly in Eastern Canada. Marine and Freshwater Research DOI: 10.1071/MF13268.

Einfeldt AL, Doucet J, and Addison JA. 2014. Phylogeography and cryptic introduction of *Hediste diversicolor* (Annelida, Nereididae) in the Northwest Atlantic. Invertebrate Biology 133: 232-241.

Einfeldt AL and Addison JA. 2013. Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the northwest Atlantic. Marine Biology 160: 1015-1027.

In Preparation

Einfeldt AL, Jesson LK, and Addison JA. The geographic scale of evolution was reshaped by shipping in the Age of Exploration. In review, Nature Communications Biology.

Conference Presentations:

Einfeldt AL and Addison JA. 2017. Divergence in introduced species: How changing technology shaped evolution in two common marine invertebrates. European Society of Evolutionary Biology 16th congress, Groningen, the Netherlands. (Poster presentation).

Einfeldt AL and Addison JA. 2017. Comparative genomics of historic marine invaders: Neutral and selective processes influencing the success of the amphipod *Corophium volutator* and polychaete *Hediste diversicolor*. Canadian Society for Evolution and Ecology 2017, Victoria, British Columbia, Canada. (Oral presentation).

Einfeldt AL and Addison JA. 2014. Historic introduction of *Corophium volutator* and *Hediste diversicolor* in the Northwest Atlantic. Genomes to Biomes 2014, Montreal, Quebec, Canada. (Poster presentation).

Einfeldt AL and Addison JA. 2012. Phylogeography of *Corophium volutator*: a biological legacy of colonialism. 1st Joint Congress on Evolutionary Biology, Ottawa, Ontario, Canada. (Oral presentation).

Einfeldt AL and Addison JA. 2012. Connectivity in the East Atlantic: Population genetics of the key benthic invertebrate *Corophium volutator*. Canadian Society for Zoology, Sackville, Nova Scotia, Canada. (Oral presentation).

Einfeldt AL and Addison JA. 2011. A tale of two bays: tidal barriers to gene flow drive genetic divergence among mudflats in the upper Bay of Fundy. Canadian Society for Evolution and Ecology, Banff, Alberta, Canada. (Oral presentation).

Einfeldt AL. 2011. Population genetics of *Corophium volutator* in the Bay of Fundy. Mudflat Ecology Meeting, Sackville, Nova Scotia, Canada. (Oral presentation).

Einfeldt AL. 2010. Population genetics in the Bay of Fundy: *Corophium volutator* and *Ilyanassa obsoleta*. Working Group of the Bay of Fundy Ecosystem Partnership, Mudflat Ecology Discussion Series. University of New Brunswick. Fredericton, New Brunswick, Canada. (Oral presentation).