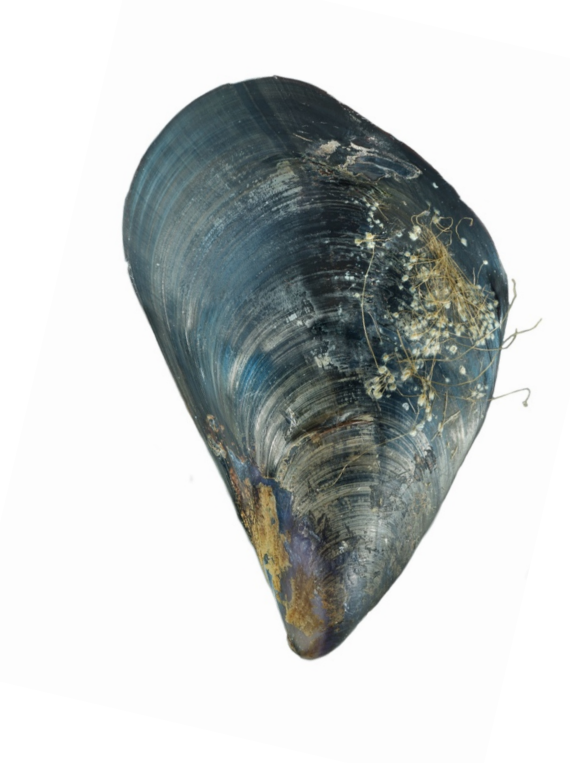
Evolutionary pathways to invasive success: Thermal adaptation, hybridisation and parallel evolution in a marine global invader

Iva Popovic
B.Sc. M.Sc.



A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2019
School of Biological Sciences

Abstract

The introduction and spread of non-native species pose numerous threats to ecosystems worldwide. While a large body of research has focused on the ecological factors contributing to the spread of non-native species, understanding the evolutionary causes and outcomes of biological invasions is a significant knowledge gap in evolutionary ecology. A key unresolved question involves whether successful non-native species (including invasive species) are already physiologically suited or 'pre-adapted' to new environments, or if colonisation implicates novel genetic changes following introduction. Hybridisation and introgression between native and introduced species can also facilitate genetic changes, potentially displacing parental species and enhancing invasive spread. Using a comparative transcriptome approach, I investigate hypotheses regarding the role of 'pre-adaptation' and recent genetic change (including hybridisation) in the spread of the Mediterranean marine mussel *Mytilus galloprovincialis*, one of the world's most widely established introduced species. In empirical Chapter 2, I explore whether the physiological traits that make warm-tolerant and invasive *M. galloprovincialis* distinct from cold-tolerant and non-invasive congeners are paralleled by evolutionary divergence at the molecular level. I test the hypothesis that genomic functions experimentally associated with adaptations to temperature stress in physiological studies show accelerated evolutionary divergence in invasive *M. galloprovincialis*. The results of this Chapter highlight several lines of evidence that positive selection has disproportionately affected genes encoding core elements of biochemical adaptation to temperature in the *M. galloprovincialis* lineage, suggesting an important role for thermal adaptation in explaining interspecific genetic differences that may also mediate different propensities for invasiveness in this genus. In empirical Chapters 3 and 4, I explore post-introduction evolutionary processes associated with *M. galloprovincialis* introductions within Australia and around the globe. In Chapter 3, I evaluate the extent of ongoing hybridisation and introgression between introduced northern *M. galloprovincialis* and a morphologically indistinguishable native Australian taxon, *Mytilus planulatus*. Contingent on resolving the demographic history between introduced and native congeners, my results demonstrate multiple recent introductions of two divergent source lineages of *M. galloprovincialis* into Australia and high levels of introgression between native and introduced populations. Estimated divergence times between congeners support a separate subspecies status for native *M. planulatus* and shed light on intrinsic challenges for invasive species research when native and introduced species boundaries are not well-defined. In Chapter 4, I sample multiple *M. galloprovincialis* populations to test the hypothesis that introduced populations have experienced parallel population genetic differentiation relative to their native genomic backgrounds. I also explore the relative contributions of major non-coding elements of the transcriptome (long non-

coding RNAs) to population differentiation and parallel genetic differentiation. I provide comparative evidence for parallel genetic changes between two introduced *M. galloprovincialis* populations in Australia and replicated introductions in California and South Africa. Analyses of variation among lncRNA transcripts suggest that non-coding elements evolve under selective constraints and are unlikely to underlie population-specific adaptations on the timescale relevant for species introductions (at least for *M. galloprovincialis*). These findings build on existing insight regarding the genomic scale at which selective processes may be predictable in introduced populations and shed light on the genomic architecture of post-introduction differentiation in *M. galloprovincialis*. Collectively, this thesis provides new perspectives on the putative genetic mechanisms underlying evolutionary changes in marine non-native species and contributes to our general understanding of the role of pre- and post-introduction evolutionary processes and their possible genetic outcomes in biological invasions.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

No publications included.

Submitted manuscripts included in this thesis

No manuscripts submitted for publication.

Other publications during candidature

Patiño S, Keever CC, Sunday JM, **Popovic I**, Byrne M, Hart MW. 2016. Sperm *bindin* divergence under sexual selection and concerted evolution in sea stars. *Molecular Biology and Evolution*. 33(8): 1988-2001

Contributions by others to the thesis

Cynthia Riginos contributed to the funding, conception and design of this thesis. She assisted with sample collection, interpretation of the data, and editing drafts of all Chapters of this thesis.

Nicolas Bierne contributed to the design of data analyses, interpretation of results and draft editing for empirical Chapter 3.

Ambrocio Matias contributed to design and execution of custom R scripts implemented for data analysis (approximate Bayesian computations) and assisted with data interpretation for empirical Chapter 3. He also contributed custom R scripts for nucleotide diversity implemented in empirical Chapter 4.

Federico Gaiti contributed to the execution of the bioinformatic pipeline for long non-coding RNA identification in empirical Chapter 4.

Statement of parts of the thesis submitted to qualify for the award of another degree

No works submitted towards another degree have been in this thesis.

Research involving human or animal subjects

No animal or human subjects were involved in this research.

Acknowledgements

Firstly, I would like to thank my supervisor, Cynthia Riginos, for being a wonderful mentor and advocate. I am grateful not only for the encouragement, patience and guidance she has provided during every step of this thesis, but also for the many thoughtful discussions that have helped shape how I think about evolution and science most generally. Thank you to Daniel Ortiz-Barrientos, Miloš Tanurdžić and John Dwyer for their involvement and contributions as my committee members. I would especially like to thank Miloš Tanurdžić for his support throughout the course of this thesis and for sparking my interests in the non-coding peculiarities of the genome. Thank you to Nicolas Bierne for his insightful comments on *Mytilus* evolution and advice as a collaborator on the third Chapter of this thesis. N. Bierne also provided Atlantic *Mytilus* samples that I utilised in all three empirical Chapters. I am grateful to Carson Keever and Michael Hart for their ongoing mentorship throughout this experience. I would also like to acknowledge the many scientists whose work has built the foundations of this thesis and without whom my research would not be possible.

This research was supported by generous funding from the Australian Biological Resources Study (ABRS), the National Sciences and Engineering Research Council of Canada and the University of Queensland (UQ). I am also grateful for additional financial support provided by the ABRS, the UQ Graduate School, the UQ School of Biological Sciences, the Australasian Evolution Society and the Society for the Study of Evolution. These awards have given me the opportunity to travel to conferences, meetings and marine stations, which have greatly enriched my academic experience.

I would like to acknowledge the many colleagues who have contributed useful discussions, assistance and comments on various parts of this thesis. I would like to thank George Somero for his invaluable insights on *Mytilus* physiology and for providing thoughtful comments on my writing. I would like to thank Lyn Cook for her useful feedback on grant writing. Thank you to Federico Gaiti for his help with lncRNA identification. I would like to acknowledge Stephan Palumbi and the Palumbi Lab members, Carmel McDougall, Stephanie Kerr, and Huanle Liu for providing bioinformatics assistance during the early stages of this thesis. I would also like to thank the members of Hacky Hour UQ for coding help, with a big thanks to Igor Makunin and Janette Edison. Thank you to all of the colleagues, friends and family who helped with international sample collection: Greg Rouse, Tom Chapman, Lucy Malard, Corinna Favaro, Brett Favaro, Ben Speers-Roesch and Bojana Popović. Thanks especially to Peter Teske for providing mussel samples from South Africa, and Victoria Cole for helping me collect mussels in the Sydney Harbour. Thank you

to Selena Popović for helping me with figures and Jessa Thurman for spending many hours photographing *Mytilus* shells under the microscope; one of these photos is shown on my title page.

I would like to warmly thank many other colleagues who have made my academic experience wonderful over the last four years. Thank you to my lab mates Josh Thia and Din Matias for their constant support, advice and friendship since my first day at UQ. I would especially like to thank Din Matias for being a collaborator on the third Chapter of this thesis, and for his help and insight, and many useful discussions over the last four years. I would also like to thank the past and current Riginos Lab members for their support: Andrew Mather, Jude Keyes, Dean Blower, Sinan Kassam, Kat Prata, Kim Dunbar and Vhon Garcia. Thank you to Maddie James, David Toews, Fabio Cortesi, Alicia Toon, James Hereward for many insightful conversations and advice about molecular lab work, bioinformatics, and life in science. I am also grateful to the School of Biological Sciences staff, who have made my experience at UQ wonderful. Thank you to my friends and colleagues for their support during the countless months of work on this thesis, especially Hugh Winwood-Smith, Owen Coffee, Stephanie Avery-Gomm, Julian Beaman, Cedric van den Berg, Hugo Perdomo, Aren Yazmaciyan, Nick and Lucy Smith, Emily Read, Marina Kamalova and the members of YUQ Kafica. I would like to thank Carmen da Silva, without whose friendship and support, it seemed at times, this thesis may not have been possible - I look forward to a lifetime of eco-evolutionary collaborations! Thank you to Serge Grosman for his encouragement and for making the last blurry months of this thesis a joy.

Finally, thank you to my family. My parents, Bojana Popović and Dragan Popović, and my sister, Selena Popović (and her other half, Filip Trajković) have supported me at every crossroad, highway and valley that have been a part of this experience. Thank you for your encouragement not only during the course of this thesis, but during the many years of studies that preceded it and for the many more that are yet to (hopefully) come.

Financial support

This research was supported by the following grants:

Australian Biological Resources Study (ABRS) National Taxonomy research Grant awarded to C. Riginos, I. Popovic and D. Colgan; ABRS Student Travel Award to I. Popovic

National Sciences and Engineering Research Council of Canada Postgraduate Scholarship Doctoral Program awarded to I. Popovic

University of Queensland (UQ) International Postgraduate Research Scholarship; UQ Centennial Scholarship; UQ Graduate School International Travel Award; and UQ Biological Sciences Conference Travel Award to I. Popovic.

Society for the Study of Evolution Rosemary Grant Research Award to I. Popovic

Australian Evolution Society Travel Conference Award to I. Popovic.

Keywords

invasive species, marine, *Mytilus galloprovincialis*, transcriptome, RNAseq, molecular evolution, introgression, demographic history, population genomics, genetic parallelism

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060411, Population, Ecological and Evolutionary Genetics, 70%
ANZSRC code: 060309, Phylogeny and Comparative Analyses, 20%
ANZSRC code: 060409, Molecular Evolution, 10%

Fields of Research (FoR) Classification

FoR code: 0603, Evolutionary Biology, 80%
FoR code: 0604, Genetics, 20%

I dedicate this thesis to my family.

Table of Contents

Chapter 1. Thesis overview **1**

Introduction 1

Evolutionary processes 6

'Pre-adaptation' 6

Post-introduction adaptation 7

Hybridisation and introgression 10

Mytilus as a model for invasive species research 12

Biogeographic history and thermal adaptation 13

Multiple global introductions 15

Interspecific hybridisation 17

Chapter 2. Comparative genomics reveals divergent thermal selection in warm- and cold-tolerant marine mussels **21**

Abstract 21

Introduction 22

Methods 25

Sample collection, assembly filtering and ORF identification 25

Transcriptome data processing and assembly 26

Transcriptome assembly quality and gene annotation 26

Ortholog identification and alignment 27

Branch-site models of molecular evolution 28

Analyses of positive selection using polymorphism variation 29

Identification of known thermotolerance candidate genes 31

Enrichment analyses of positively selected genes 31

Results 32

Sequencing and transcriptome assembly 32

Assembly quality, annotation and ortholog identification 32

Evidence of positive selection in the warm-tolerant *M. galloprovincialis* 33

Thermotolerance candidate genes and gene ontology enrichment analyses 34

Discussion 36

Temperature-related loci show the strongest signatures of positive selection 36

Functional concordance between positively selected thermotolerance candidate genes and temperature-relevant loci in *M. galloprovincialis* 38

Combining codon models with polymorphism and divergence analyses 40

Genetics of species divergence and implications for thermal adaptation 41

Chapter 3. Resolving the cryptic origins of Australia's native blue mussel despite high rates of admixture with twin introductions of independent non-native mussel lineages. 51

Abstract 51

Introduction 52

Methods 56

Sample collection, RNA extraction and sequencing 56

RNAseq data processing and reference transcriptome assembly 57

Genomic data filtering 57

Analyses of population structure and admixture 58

Genomic analysis of species relationships 59

Approximate Bayesian Computations (ABC) of demographic history 60

Results 64

Genomic analysis of species relationships 64

Evidence for genome-wide introgression with introduced *M. galloprovincialis* 65

Historical demographic inference 67

Discussion 69

Population relationships and introgression with multiple introduced source populations 69

Late-Pleistocene divergence between native and introduced Australian blue mussels 72

Modelling the effects of selection on genome-wide variation 73

Implications for species delimitation and endemic diversity 74

Monitoring *M. galloprovincialis* introductions in Australia 75

Chapter 4. Parallel genetic differentiation in multiple introduced populations of the marine mussel *Mytilus galloprovincialis* sheds light on long non-coding RNA evolution. 87

Abstract 87

Introduction 88

Methods 92

Study system 92

Sample collection, RNA extraction and sequencing 93

Reference transcriptome assembly: Additional filtering 94

Identification of lncRNA transcripts 94

Mapping and identification of SNPs 95

Analyses of overall population structure 95

Assessment of introduced population independence 96

Genomic outlier detection 97

Nucleotide diversity 98
Results 99
Contributions to genetic differentiation: Full and lncRNA datasets 99
Analyses of population independence 100
Repeatability of post-introduction evolution 102
Discussion 103
Repeatable patterns of post-introduction evolution 103
Contributions to genetic differentiation: Long Non-coding RNAs 109
Conclusions 111
Chapter 5. General Discussion 124
Overview 124
'Pre-adaptation' 125
Post-introduction evolution: Hybridisation and introgression 126
Post-introduction evolution: Repeatability of genetic differentiation 129
Conclusions 132
Literature Cited 134
Supporting Information 168
Chapter 2 168
Chapter 3 173
Chapter 4 180

List of Figures

Figure 1-1. 20
Figure 2-1. 43
Figure 2-2. 45
Figure 2-3. 46
Figure 3-1. 78
Figure 3-2. 79
Figure 3-3. 80
Figure 3-4. 83
Figure 3-5. 84
Figure 4-1. 113
Figure 4-2. 114
Figure 4-3. 116
Figure 4-4. 117
Figure 4-5. 118
Figure 4-6. 119

List of Tables

Table 2-1. 47
Table 2-2. 48
Table 2-3. 50
Table 3-1. 85
Table 3-2. 86
Table 4-1. 120
Table 4-2. 121
Table 4-3. 122
Table 4-4. 123

Chapter 1. Thesis overview

Introduction

Human-mediated introductions and the spread of non-native species pose numerous threats to ecosystems worldwide (Molnar et al. 2008). Invasive species, or introduced species that become widespread following human-mediated transport outside of their natural geographic ranges (Bock et al. 2015), are associated with profound impacts on receiving communities at multiple levels of biological organisation. Successful invaders may directly or indirectly displace native species though predation or competition (e.g., Branch and Steffani 2004; Arcella et al. 2014). Introduced species may also hybridise with native taxa, imposing less discernible, but equally severe, genetic impacts that can quickly erode or replace native genetic diversity (e.g., Blum et al. 2010; Fitzpatrick et al. 2010; Glotzbecker et al. 2016). In other cases, introduced species may inflict damage to local habitats (e.g., Robinson et al. 2007) and prompt cascading community-level impacts that can transform entire ecosystems (e.g., Griffiths et al. 1992; Shine 2010). The diverse and high ecological costs of invasive species provide substantial incentives for minimising their spread and for gaining a deeper understanding of the ability of non-native species to survive and reproduce in introduced environments (Molnar et al. 2008; Stewart et al. 2009). While a large body of research has focused on the anthropogenic (Carlton and Geller 1993; Rius et al. 2015a) and ecological factors (Carlton 1996; Branch and Steffani 2004; Lockwood et al. 2009; Richardson 2011; Ibanez et al. 2014) contributing to the spread of non-native species, understanding the evolutionary processes that create opportunities for some species to become invasive and the long-term outcomes of successful invasions remains a significant knowledge gap in evolutionary ecology (Stewart et al. 2009; Chown et al. 2015; Bock et al. 2015).

Since the earliest studies on biological invasions (Baker 1965; Baker 1974; reviewed in Bock et al. 2015), species introductions have received increasing attention for their utility as natural experiments for studying evolution over contemporary timescales. Indeed, a growing number of studies over the last two decades have focused on identifying the genetic changes associated with invasive taxa and the evolutionary processes involved in their establishment and spread (reviewed in Prentis et al. 2008; Bock et al. 2015; Dlugosch et al. 2015; Colautti and Lau 2015). Studies have broadly focused on two key challenges in invasion genetics: (1) understanding the demographic features of introduced and invading populations with an emphasis on the effects of genetic drift, population bottlenecks and multiple introductions on within-population genetic diversity; and (2) investigating the frequency and

nature of adaptive evolution as invasive species become established in novel environments. Evolution through stochastic processes involving the demographic history of introduced populations is comparatively better understood than adaptive evolution both empirically (Roman and Darling 2007; Uller and Leimu 2011) and theoretically (Wright 1931; Nei et al. 1975) and has been recently and extensively reviewed (e.g., Dlugosch and Parker 2008; Bock et al. 2015; Dlugosch et al. 2015; Estoup et al. 2016). Now that high-throughput approaches are enabling genome-scale investigations in non-model organisms, including a variety of invasive species, biologists can make stronger inferences about the role of natural selection in invasive success and begin to identify the genetic changes that may predispose, accelerate or impede biological invasions (Prentis et al. 2008; Rius et al. 2015b; Chown et al. 2015).

Although the history of invasive species research is rooted within terrestrial systems, which continue to be the predominant focus of studies investigating biological invasions (Prentis et al. 2008; Bock et al. 2015), genetic and genomic approaches are increasingly being employed to study invaders in the marine environment (Holland 2000; Geller et al. 2010; Rius et al. 2015a; 2015b; Tepolt 2015; Viard et al. 2016). Marine biological invasions offer two key advantages for studying the role of natural selection in successful species introductions. First, many marine species with sessile or slow-moving adults tend to exhibit high fecundities and high dispersal potential (through long-lived planktonic larvae) that support elevated rates of demographic connectivity across marine environments (Mayr 1954; Lessios et al. 2001; Palumbi and Lessios 2005). Because the transport of large numbers of planktonic larvae likely promotes successive and widespread introductions through human-mediated vectors (e.g., ballast water discharge and aquaculture transplantation; Carlton 1996; Carlton and Geller 1993), the prospect of multiple introductions provides naturally replicated experiments of introduced populations and taxa well-suited for comparative population genetic inferences (e.g., Tepolt and Palumbi 2015; Guzinski et al. 2018). Second, empirical studies to date have rarely documented strong signatures of genetic bottlenecks and or reduced genetic diversity in introduced marine populations compared to native populations (Riquet et al. 2013; Rius et al. 2015a; Viard et al. 2016). The combination of multiple introductions and high propagule pressure is a primary factor suspected to help founding populations maintain high genetic diversity and avoid the negative impacts of genetic drift (Palumbi 1992; Gagnaire et al. 2015; Dlugosch et al. 2015; Blackburn et al. 2015). Introduced populations originating from high numbers of propagules from diverse sources are, therefore, likely to retain large effective population sizes and sufficient standing genetic variation that may increase opportunities for adaptation to new introduced environments (Roman and Darling 2007; Riquet et al. 2013; Viard et al. 2016, but see Bierne 2016).

Despite these advantages, few studies have explored questions about the outcomes of local adaptation and the genetic changes associated with marine invasions (e.g., Riquet et al. 2013; Tepolt and Palumbi 2015; Bernardi et al. 2016). Instead, genetic studies have focused on collecting simple demographic data on marine introductions, with researchers seeking to resolve cryptic species diversity, or provide basic information on the native sources of introduced populations, patterns of genetic diversity across native and introduced ranges, and the frequency of repeated introductions (reviewed in Le Roux and Wieczorek 2009; Geller et al. 2010). Indeed, marine invasive species have also presented several challenges for classical genetic approaches (Tepolt 2015). Large effective population sizes that slow down genetic drift also maintain high levels of shared ancestral polymorphisms, high levels of genetic exchange at neutral loci, and weak population genetic differentiation on the scale of thousands of kilometres (Slatkin 1985; Palumbi 1992; Gagnaire et al. 2015). Such population genetic features inherently complicate differentiation-based inferences of introduction histories and make distinguishing among the sources of introduced populations difficult using few numbers of markers (Waples 1998; Tepolt and Palumbi 2015). Additionally, an abundance of cryptic and reproductively semi-isolated species complexes in the marine environment are a persistent issue for detecting marine invasions, and in some cases, distinguishing between closely related native and introduced species (Geller et al. 2010; Appeltans et al. 2012; Brunetti et al. 2015; Bouchemousse et al. 2016; Pante et al. 2015b; Viard et al. 2016). Given the limited utility of traditional genetic markers for delineating high gene flow marine populations and the very recent timescales of most contemporary invasions (e.g., Branch and Stefani 2004), more loci are required to resolve population genetic structure and selectively favoured genetic variation in introduced marine populations (Sherman et al. 2016; Guzinski et al. 2018).

High-throughput genomic and transcriptomic approaches are becoming increasingly accessible methods to characterise and analyse genetic variation in introduced species adapting to new physiological and ecological conditions (Rius et al. 2015a; Sherman et al. 2016). Genomic outlier scans are widely used methods to identify loci with atypical patterns of genetic differentiation relative to the neutral genomic expectations, either by detecting reductions in genetic diversity (selective sweep; e.g., Puzey and Vallejo-Marín 2014; Messer and Petrov 2013), strong differences in allele frequencies between populations (e.g., Foll and Gaggiotti 2008) or allelic associations with the environment that may be indicative of natural selection acting on these genomic regions (e.g., reviewed in De Wit et al. 2015; Chown et al. 2015; Jeffery et al. 2017). Other statistical methods for detecting selection involve neutrality tests based on allele frequency spectra (e.g., Tajima's D) or the variability of sequence data within and between species (e.g., HKA test; McDonald-Kreitman test) (reviewed in Anisimova and Liberles 2012). Alternatively, codon model tests of molecular evolution

are a powerful approach to detect changes in protein evolution and signatures of selection acting on genes and amino acids underlying potentially adaptive phenotypes (e.g., Hart 2014). Explicit modelling of demographic parameters from genomic data has also been highly informative for detangling the effects of demographic history from other post-introduction evolutionary processes. Notably, Approximate Bayesian Computation (ABC) approaches have been valuable for reconstructing invasion routes (Barker et al. 2017; reviewed in Estoup et al. 2016) and have allowed comparative inferences of past and ongoing admixture in human-mediated contact zones (e.g., *Ciona spp*. Roux et al. 2013; *Drosophila sp*., Pascual et al. 2007). Indeed, recent applications of a variety of population genomic methods in high gene flow marine species have provided evidence supporting the long-standing hypothesis that selection has a central role in promoting population differentiation in marine populations (Bierne et al. 2003a; Lessios 2007; Sanford and Kelly 2011), despite the homogenising effects of gene flow (e.g., Atlantic cod, Nielsen et al. 2009a; 2009b; *Littorina sp*, Galindo et al. 2010; Atlantic herring, Limborg et al. 2012, Lamichhaney et al. 2012, Corander et al. 2013; purple sea urchins, Pespeni and Palumbi 2013; Pacific lamprey, Hess et al. 2013; Pacific bat star, Hart et al. 2014; rough periwinkle, Westram et al. 2014; American lobster, Benestan et al. 2015; reviewed in Sherman et al. 2016). As genomic approaches gain broader use and accessibility for marine taxa, new resources for genomic studies in marine invasive species offer opportunities to test among evolutionary processes that may facilitate biological invasions (Tepolt 2015).

The overarching goal of this thesis is to advance our understanding of the evolutionary processes associated with successful species introductions in the marine environment. There are three broad evolutionary scenarios by which non-native species are able to persist in environments outside of their native range (Rius et al. 2015a; Figure 1-1): (1) successful introduced species may be already physiologically suited or 'pre-adapted' to new environmental conditions. At the population-level, 'pre-adaptation' may also encompass changes in the frequency of standing genetic variation already present in the native range; (2) alternatively, colonisation may implicate novel genetic changes in genes underlying favorable phenotypes in the introduced range; (3) introgressive hybridisation between native and introduced taxa can also facilitate genetic changes, potentially contributing to a species' ability to persist in new environments. In this thesis, I use the globally invasive marine mussel *Mytilus galloprovincialis* as a model system to investigate these three non-mutually exclusive evolutionary hypotheses and their implications for understanding adaptation in non-native species in the marine environment.

Chapter 2 explores whether the past environments under which certain species evolved can make them prone to ecological and physiological tolerance and thus, an evolutionary propensity or 'pre-

adaptation' to becoming successful invaders. In this Chapter, I combine comparative molecular evolutionary methods with knowledge from whole-organism physiological studies to compare the genomes of closely related invasive and non-invasive *Mytilus* species that have evolved under divergent thermal environments. My findings provide support for the hypothesis that genomic functions experimentally linked to species-specific temperature stress responses show correlated evolutionary divergence in warm-tolerant and invasive *M. galloprovincialis*, compared to three cold-tolerant and non-invasive congeners. These findings suggest an important contribution of temperature-dependent selection leading to the genomic divergence of closely-related invasive and non-invasive species and provide new perspectives regarding the putative molecular mechanisms that may contribute to different propensities for invasiveness.

In Chapter 3, I investigate the genomic outcomes of human-mediated secondary contact between introduced *M. galloprovincialis* and a morphologically indistinguishable native Australian species, *Mytilus planulatus*. Using a series of population genomic analyses and approximate Bayesian computations, I clarify the demographic history between *M. planulatus* and northern *M. galloprovincialis* and validate that mussels from Tasmania are representative of the endemic Australian taxon. The results of this study provide strong evidence that two contemporary *M. galloprovincialis* introductions into southeastern Australia are derived from genetically divergent northern source populations and that both introductions are associated with high levels of recent admixture between native and introduced populations. Estimated divergence times between congeners support a separate subspecies status for native *M. planulatus* and shed light on intrinsic challenges involved in understanding invasive species spread and the consequences of secondary contact when native and introduced taxa fall within the 'grey zone' of the speciation continuum (Roux et al. 2016).

Chapter 4 draws upon major findings from Chapter 3 to explore the genetic outcomes of post-introduction evolution in replicated worldwide introductions of *M. galloprovincialis* and provides insight regarding the genetic architecture of differentiation associated with introduced populations. In this chapter, I test the hypothesis that introduced *M. galloprovincialis* populations in Australia (Chapter 3) share strong deviations in genetic differentiation with replicated introductions in California and South Africa that originate from the same source populations. I also explore the relative contributions of long non-coding RNA (lncRNA) transcripts to population structure and parallel genetic differentiation; lncRNAs are major expressed components of the genome recently implicated (but rarely explicitly investigated) as an important source of genetic variation during invasion. Through these investigations, I provide comparative evidence for repeatable genetic

changes in multiple introduced *M. galloprovincialis* populations across its present-day distribution in the northern and southern hemispheres and build on existing insights regarding the genomic scale at which the outcomes of selection may be repeatable. Focusing on the contribution of lncRNAs to population structure, I conclude that these non-coding elements evolve under selective constraints and are unlikely to underlie population-specific adaptations on the timescale relevant for species introductions.

These three broad evolutionary scenarios (pre- and post- introduction evolutionary genetic changes, including hybridisation) and their implications for understanding marine invasions are discussed in detail below.

Evolutionary processes

'Pre-adaptation'

In the first scenario, the initial establishment of non-native species depends on the fitness or broad-scale physiological and ecological tolerance of introduced genotypes (Baker 1965; Bock et al. 2015; Bernardi et al. 2016). Because selection favours genotypes well suited or 'pre-adapted' to the conditions in the introduced range, species originating in similar biotic and abiotic conditions may be more likely to successfully establish and spread (Hamilton et al. 2015). This scenario, therefore, in its simplest form does not implicate evolution after the initial introduction of a population into the invaded range (Figure 1-1). Corroborative evidence of the 'pre-adaptation' hypothesis involves examining whether specific genetic changes associated with the evolutionary history of an invasive species have contributed to the evolution of invasiveness, or the potential for a species to become widespread following the initial introduction outside of its natural range (Bock et al. 2015; Chown et al. 2015). Bottom-up approaches such as genomic comparisons among closely related invasive and non-invasive species can reveal genetic changes that may predispose certain species or lineages to becoming successful invaders (Tepolt 2015). Bottom-up approaches are particularly advantageous in that *a priori* identification of candidate genes or traits under selection is not required (Hodgins et al. 2014; Renaut et al. 2014; Chown et al. 2015). Rather, this approach allows the examination of large numbers of genes to investigate whether changes in allele frequency, differences in gene expression, or signatures of positive selection between native and invasive taxa that may underlie adaptive variation (Hodgins et al. 2014; Bock et al. 2015). While the number of such studies has been limited, key examples have illustrated the power of targeted (Fields et al. 2006; Lockwood and Somero 2012, discussed below) or genome-wide comparisons of orthologs among invasive and non-invasive

congeners for investigating adaptive genetic parallelism (Hodgins et al. 2014) and expression differences (Lockwood and Somero 2012) in the evolution of invasive lineages. In turn, such analyses have identified re-occurring candidate genes or functional groups of genes as targets of past natural selection, forming the basis of predictions about which molecular mechanisms contributed to the evolution of invasive potential in the groups studied (Hodgins et al. 2013; 2014).

At the population level, a form of 'pre-adaptation' may also involve the introduction and persistence of only a subset of pre-existing alleles and genotypes that are particularly well-matched to the new introduced range (Figure 1-1). This scenario is somewhat analogous to Baker's single 'general purpose-genotype' that is best-suited for successful invasion (Baker 1955; 1974), but does not imply colonisation of the same lineage into a wide range of environments. Local adaptation throughout different parts of a species' native range may increase the probability that some native populations are more 'pre-adapted' than others to their new introduced environment (Keller and Taylor 2010; Galindo et al. 2010; Rius et al. 2015a; Tepolt 2015). For example, evidence for local adaptation across the native range of the invasive European green crab, *Carcinus maenas,* suggests that variation in the thermal tolerance between different source populations has played an important role in the success of multiple introduced populations throughout North America (Darling et al. 2008; Tepolt and Somero 2014; Tepolt and Palumbi 2015). Similarly, a genetic survey of microsatellite and Amplified Fragment Length Polymorphism (AFLP) markers in the slipper limpet, *Crepidula fornicata,* identified a small number of outliers confined to population comparisons in the native range, but found no evidence for differentiation between invasive and native populations (Riquet et al. 2013). While these findings suggest that pre-existing genotypes may be well adapted to the introduced environment, the ambiguity around the selected or neutral status of length polymorphism markers makes such patterns difficult to interpret regarding the effects of natural selection during invasion (Riquet et al. 2013).

Post-introduction adaptation

While some invasive species may be well-suited to their introduced environment, there is growing evidence that invasive species are frequently exposed to strong selection for local adaptation following introduction (Colautti and Lau 2015), and that adaptation can facilitate invasive spread (García-Ramos and Rodríguez 2002; Vandepitte et al. 2014). The second broad scenario in which invasive species are able to survive in their introduced environment involves adaptive evolution following introduction through selection for beneficial *de novo* mutations arising in the introduced range (Figure 1-1; reviewed in Tepolt 2015; Bock et al. 2015; Chown et al. 2015). Post-introduction

evolution may also involve changes in the frequencies of alleles and genotypes already present in the native range such that selection acts on pre-existing standing genetic variation following introduction (Figure 1-1; Vandepitte et al. 2014). Although substantial reductions in allelic richness and (to a lesser extent) heterozygosity are common features of introduced populations experiencing founder events and demographic bottlenecks during the course of introduction (Nei et al. 1975; Dlugosch and Parker 2008), examples of rapid evolutionary change in invasive species suggest that most invading populations have sufficient genetic variation to adapt to new ecological conditions (e.g., Hodgins et al. 2013; Vandepitte et al. 2014), particularly in marine systems (Rohfritsch et al. 2013; Rius et al. 2015a; Viard et al. 2016). It is now recognised that the paradox of genetic bottlenecks in biological invasions (i.e. the ability of invasive species to become widespread despite low genetic diversity and low evolutionary potential; Frankham 2005), depends on a variety of demographic and biological parameters that are of idiosyncratic nature (Allendorf and Lundquist 2003; Dlugosch et al. 2015; Estoup et al. 2016); such parameters include the frequency of multiple introductions and the number of individuals released during introduction (Carlton and Geller 1993; Lockwood et al. 2005; Simberloff 2009; Blackburn et al. 2015), the prevalence of intraspecific admixture from divergent source populations (Roman and Darling 2007; Keller and Taylor 2010; Uller and Leimu 2011; Rius and Darling 2014), as well as the mating systems of invaders (Ferrero et al. 2015).

Indirect evidence for post-introduction adaptation comes from parallel evolution of latitudinal clines for morphological traits or abiotic tolerance in both native and invaded ranges (e.g., *Drosophila subobscura*, Huey et al. 2000; Gilchrist et al. 2008, Sotka et al. 2018), as well as differentiation in ecologically important traits related to life history strategies, reproduction, growth and competition among introduced and native populations (e.g., Phillips et al. 2006; Hodgins and Rieseberg 2011; Turner et al. 2014; Ferrero et al. 2015; reviewed in Bock et al. 2015). In cases where conditions in the introduced range are predicted to be less stressful than native ones, evolutionary trade-offs between reduced abiotic stress tolerance and increased growth and reproduction have been proposed to account for observed trait differences between invasive and native plant populations (Grime 1977; Hodgins and Rieseberg 2011; Turner et al. 2014); however, support for this hypothesis in animal invasions, to my knowledge, remains to be explored. In contrast, evidence for genetic adaptation in marine non-native species has been largely limited to quantitative genetic studies involving trait differences within and between invasive and native populations, without investigating causative genetic variants (reviewed in Tepolt 2015). Other important examples in marine organisms have involved phylogeographic approaches used to link invasive lineages and haplotypes to environmental differences (e.g., Lee 1999; Geller et al. 2008; Asif and Krug 2012). For example, Lee (1999) demonstrated that independent invasive populations of the copepod *Eurytemora affinis* are more

closely related to their respective marine source populations than to each other, implying repeated evolution of low salinity tolerance in at least five independent invasions into fresh water. Later studies in the same biological system have provided evidence for replicated physiological shifts in salinity tolerance in the introduced range (Lee et al. 2003; Lee et al. 2011). Although genetic mechanisms were not examined explicitly, these findings implicate the role of strong selection for genetically based adaptations accompanying recent (<100 years) independent invasions into fresh water (Lee et al. 2003; Lee 2016).

Fewer studies, however, have examined the sources of genome-wide variation underlying differentiation and adaptation in introduced populations. A notably example includes repeated evolution of clines in chromosomal inversion frequencies in the introduced range of *Drosophila subobscura* resembling that of the native range (Pascual et al. 2007). In weedy plants, empirical research has provided evidence for differentiation in gene expression among native and introduced genotypes of common ragweed (*Ambrosia artemisiifolia*; Hodgins and Rieseberg 2011; Hodgins et al. 2013), and parallel shifts in gene expression among weedy populations with independent non-weedy origins (Lai et al. 2008). At the genomic level, Vandepitte et al. (2014) used genetic outlier analyses to detect temporal shifts in allele frequencies in six genes underlying flowering time adaptation in historical and contemporary colonising populations of Pyrenean rocket (*Sisymbrium austrium chrysanthum*), and whole-genome sequencing in introduced monkeyflower (*Mimulus guttatus*) populations revealed signatures of selective sweeps in genomic regions associated with flowering time and stress responses (Puzey and Vallejo-Marín 2014). Rapid responses to selection have also been documented in both introduced plants and animals undergoing range expansions (e.g., Preisser et al. 2008; Phillips et al. 2010).

In the marine realm, as far as I am aware, only four population genetic studies have applied high-throughput sequencing approaches to study post-introduction evolution in a marine species in both its native and introduced ranges (excluding anadromous fishes): (1) Tepolt and Palumbi (2015) conducted a transcriptome survey (RNAseq) across multiple populations of *C. maenas* in its invasive and native ranges (discussed above); (2) Saarman and Pogson (2015) used a reduced genome approach (i.e. restriction-site-associated DNA markers; RADseq) investigated patterns of introgression between invasive and native congeners in the marine mussel genus *Mytilus*; (3) Similarly, Bernardi et al. (2016) used RADseq markers to identify genomic outliers differentiating native and colonising populations of the bluespotted cornetfish, *Fisularia commersonii*, across a salinity gradient in the Mediterranean Sea; (4) Gagnaire et al. (2018) used whole-genome sequencing to investigate parallel genetic divergence between introduced and native populations of two oyster

species in the genus *Crassostrea*; however, the focus of this study was on heterogeneity in the rate of introgression between lineages, rather than the contribution of pre- or post-introduction adaptation to intraspecific variation. To date, all other genome-wide scans have been carried out using traditional genetic approaches (i.e. microsatellites, AFLP and single nucleotide polymorphisms; SNPs); however, small numbers of markers or inadequate resolution of population divergence have precluded strong inferences regarding adaptation (e.g., Riquet et al. 2013). In another example, Rohfritsch et al. (2013) identified shared genetic outliers in two invasive populations of the Pacific oyster *Crassostrea gigas*, (using microsatellites, AFLP and SNP markers) but it could not be determined whether the observed patterns reflected parallel responses to post-introduction selection or shared demographic histories of introduction.

Hybridisation and introgression

As non-native species are transported globally with increasing frequency, species introductions are likely to results in secondary contact between closely related native and introduced taxa. The third evolutionary scenario by which genetic changes may contribute to post-introduction adaptation involves interspecific hybridisation and introgression between native and introduced populations (Figure 1-1; Stebbins 1985; Abbot 1992). Intraspecific admixture between genetically distinct native sources populations may also benefit invaders though similar evolutionary mechanisms discussed below (Ellstrand and Schierenbeck 2000; Keller and Taylor 2010; Rius and Darling 2014); however, for the purpose of this overview, I will focus on interspecific hybrid interactions as they relate more explicitly to my thesis.

Introgression with the receiving community is a common phenomenon associated with biological invasions in both terrestrial (e.g., Zayed and Whitfield 2008; Lai et al. 2012; Kirkpatrick and Barret 2015) and marine systems (e,g., Saarman and Pogson 2015; Le Roux and Wieczorek 2009). In non-native plants, there is also growing support that hybridisation can facilitate adaptation to introduced environments by promoting the evolution of colonising traits such as increased fecundity and growth (Schierenbeck and Ellstrand 2009; Hovick and Whitney 2014); however, evidence for increased hybrid performance relative to parental taxa has been inconsistent among non-native animals (Hovick and Whitney 2014). There are several proposed molecular mechanisms by which hybridisation and introgression may benefit introduced populations (Rieseberg et al. 2007; Ellstrand and Schierenbeck 2000; Hovick and Whitney 2014). Hybridisation can effectively increase genetic variation in founding populations, potentially reducing the negative effects of genetic bottlenecks associated with introduction or masking deleterious alleles through increased heterozygosity (Abbot 1992; Ellstrand

and Schierenbeck 2000). Interspecific recombination can also create novel allele and gene combinations that may be selectively favoured, thereby increasing the likelihood of survival of hybrid individuals (relative to parental taxa) in the introduced range (Hovick et al. 2011). New combinations of parental genotypes have also been associated with novel phenotypic traits allowing hybrids to occupy habitats unsuitable for parental genotypes (e.g., Rieseberg et al. 2007; Sloop et al. 2009) or outcompete parental taxa under different environmental conditions (Perry et al. 2002; Fitzpatrick and Shaffer 2007; Hovick et al. 2011; Arcella et al. 2014). Hybridisation with native taxa may also aid founding populations through purely demographic processes by reducing the severity of Allee effects faced by introduced populations (Mesgaran et al. 2016). Strong Allee effects may be particularly relevant for marine invasive species with bentho-pelagic life cycles, in cases where larvae are likely to move away from the parents, thereby limiting mating opportunities for initial founders (Johannesson 1998; Gagnaire et al. 2018).

Studies of marine non-native species have not explicitly investigated whether introgression directly increases the propensity for successful introductions; it is evident, however, that introgression occurs frequency in marine taxa and that hybridisation has played a role in the evolution of some marine invasive species (Le Roux and Wieczorek 2009; Sloop et al. 2009; Roux et al. 2013). For example, there is evidence for hybrid origins for invasive populations of a marine alga, *Caulerpa racemosa* in the Mediterranean Sea (Durand et al. 2002), and adaptive interspecific introgression was identified at a nuclear locus in an invasive marine mussel in its native range (Fraïsse et al. 2014; discussed below). However, interspecific mating will also mediate gene flow between selectively neutral alleles, even in the absence of increased hybrid fitness (Currat et al. 2008; Zayed and Whitfield 2009). Because selection acts on genes and phenotypes, signals of introgression in the introduced range (e.g., Saarman and Pogson 2015) are not necessarily related to adaptive processes (i.e. increased hybrid fitness) and invasiveness (Estoup and Guillemaud 2010; Hovick and Whitney 2015). Given these challenges, population geneticists interested in the links between natural selection, introgression and successful introductions have applied transcriptome-based approaches (e.g., in plants; Kane et al. 2009; Lai et al. 2012; Hodgins et al. 2013) to assess genomic signatures of recent introgression between native and invasive species. Although genetic studies of marine invasive species have contributed to a better understanding of how introduced and native populations become differentiated following introduction (Geller et al. 2010; Rius et al. 2015a), the relative contributions of pre- and post-introduction adaptation, and interspecific introgression in the evolution of marine invasive taxa and their ability to exploit novel environments remain poorly understood (Viard et al. 2016).

Mytilus as a model for invasive species research

Blue mussels of the genus *Mytilus* are dominant sessile species in cold and temperate rocky marine communities throughout coastal habitats in the northern and southern hemispheres. *Mytilus* mussels are also highly abundant on docks and pilings, and can be found attached to other mussels, creating dense aggregations or mussel beds. Northern hemisphere *Mytilus* include the ribbed mussel, *Mytilus californianus* (Conrad 1837), and morphologically similar smooth-shelled sibling species *Mytilus trossulus* (Gould 1850), *Mytilus edulis* (Linnaeus 1758), and one of the world's most widespread invasive species, *Mytilus galloprovincialis* (Lamarck 1819) (McDonald et al. 1991; Seed 1992; Lowe et al. 2000). The Mediterranean native *M. galloprovincialis* is the only *Mytilus* species that has successfully established invasive populations globally. Because *Mytilus* species enjoy highly-dispersive larvae that swim in the plankton for several weeks (McQuaid and Phillips 2000), this genus provides an exceptional opportunity to investigate how pre- and post- introduction genetic changes, including hybridisation, contribute to the evolution of high gene flow marine non-native species.

A significant consideration, however, is that the genomes of *Mytilus* mussels are large and complex (28 chromosomes; estimated genome size of $1.56X10^{11}$ bp ~ 15,000 genes; Craft et al. 2010) and present some challenges for sequencing and analysing genome-wide variation in this genus. For example, an estimated ~30% of the *M. galloprovincialis* draft genome consists of highly repetitive elements and heterozygosity is predicted to be high and widespread (Murgarella et al. 2016). Indeed, previous attempts of genome-wide genotyping with microsatellite markers (C. Riginos, personal communication) and RADseq sequencing (A. Stambuk, personal communication) has been confounded by high within-species diversity and the prevalence of null alleles at some loci (i.e. failed amplification in PCR assays). Transcriptome sequencing is a useful alternative approach that is increasingly used to study a range of non-model species (Romiguier et al. 2014; reviewed in Gayral et al. 2013; De Wit et al. 2015) and has been recently reviewed in the context of marine invasive species (Sherman et al. 2016). In this approach, the genome is reduced non-randomly such that reads are derived only from expressed transcripts present in the sampled tissue at the time of preservation or extraction (Cahais et al. 2012; Gayral et al. 2013). Expressed transcripts include both protein-coding messenger RNA (mRNA) with flanking untranslated regions and non-protein-coding sequences. Non-coding transcripts can encompass a variety of other polyadenylated RNA transcripts (i.e. long non-coding RNAs (lncRNA), small regulatory RNAs, transfer RNAs (tRNA) and ribosomal RNAs (rRNA)) that may hold functionally relevant variation important for adaptation (Gaiti et al. 2015; Stapley et al. 2015).

Transcriptome sequencing is also advantageous because reference assemblies can be assembled *de novo* without alignment to a genome (De Wit et al. 2015). Additionally, transcriptome derived SNPs can be used for a variety of population genetic and phylogenomic analyses (e.g., Tepolt and Palumbi 2015), which can facilitate targeted examinations of amino acids and genes (e.g., Chapter 2) or groups of expressed transcripts of interest (e.g., Chapter 4). On the other hand, transcriptome sequencing requires additional precautions to exclude paralogous genes and alternatively spliced isoforms in reference assemblies (Cahais et al. 2012). Higher costs of per individual sequencing (De Wit et al. 2012; Tepolt 2015) also impose restrictions on the number of samples that can be economically sequenced per species and population. For the purpose of this thesis, however, the disadvantages of the approach were outweighed by its benefits: sequencing a large representative of the gene space that may be relevant to ecologically-important genetic variation was well-suited for asking questions regarding the role the sources of expressed variation that may be functionally-relevant and targeted by selective processes during invasion (Gayral et al. 2013; reviewed in Bourne et al. 2018).

In addition to the growing list of transcriptomic (e.g., Romiguier et al. 2014; Gerdol et al. 2014; Fraïsse et al. 2015; Saarman et al. 2017) and genomic resources (e.g., draft genome: Murgarella et al. 2016) available for *Mytilus* mussels, this study system offers at least three advantages for studying the genetic basis of differentiation and adaptation in marine invasions: First, *Mytilus* mussels have become an important comparative system for studying the links between species biogeography and thermal physiology and there is a growing body of evidence for understanding the molecular phenotypes (i.e. protein function) involved in environmental adaptation in this group (Lockwood et al. 2015). Second, replicated *M. galloprovincialis* introductions throughout the northern and southern hemispheres offer an opportunity to compare the genetic responses to introduction across multiple introduced populations compared to the native range. Third, because invasive *M. galloprovincialis* has a history of hybridising with native *Mytilus* congeners following secondary contact, it is possible to study the different genetic outcomes of introgression in replicated human-mediated contact zones between introduced and native populations. These advantages are outlined in more detail below.

Biogeographic history and thermal adaptation

The biogeographic history of northern hemisphere *Mytilus* mussels is relatively well understood and is supported by genetic and genomic studies (Śmietanka et al. 2015; Fraïsse et al. 2014). Speciation between North Atlantic *M. edulis* and *M. galloprovincialis* occurred 2.5 million years following the isolation of *M. galloprovincialis* in the Mediterranean Sea (Seed 1992; Roux et al. 2014a), and their divergence from North Pacific-derived *M. trossulus* is estimated at 3.5 million years ago (Rawson

and Hilbish 1995). Despite their close relatedness and morphological and ecological similarities, the biogeographic history of *Mytilus* species suggests that congeners likely evolved under disparate abiotic conditions (Lockwood and Somero 2011a). Ecophysiological studies investigating the role of environmental adaptation in setting species distribution limits have largely focused on *M. trossulus* and *M. galloprovincialis* in the the northeastern Pacific, where *M. galloprovincialis* has displaced native *M. trossulus* from the southern portion of its former distribution through southern and central California (Rawson et al. 1999; Geller 1999; Braby and Somero 2006a). The present day distributions of *M. trossulus* and exclusive establishment of *M. galloprovincialis* from Mexico to Monterey Bay, California suggests that physiological differences in abiotic tolerance between these two congeners underlie their current biogeographic limits: *M. trossulus* is more cold-adapted to the cooler, tidally influenced waters of the North Pacific where it is native, whereas *M. galloprovincialis* is more adapted to survive and compete in warmer and more saline waters like those characteristic of the Mediterranean Sea where it evolved. Furthermore, the northern distribution of *M. galloprovincialis* in California has shown to fluctuate with the influx of colder water associated with the Pacific Decadal Oscillation (Hilbish et al. 2010), lending additional support to the hypothesis that key abiotic factors, such as temperature and salinity underlie its invasive distribution and competitive advantage in this region (e.g., Schneider and Helmuth 2007).

Physiological and biochemical studies in one introduced *M. galloprovincialis* population in California have consistently demonstrated marked differences in temperature and salinity tolerance between native and invasive *Mytilus* congeners, with higher thermal tolerance but lower salinity tolerance in *M. galloprovincialis* compared to native *M. trossulus* (reviewed in Lockwood and Somero 2011a; Somero 2012; Lockwood et al. 2015). While *M. edulis* has not been the focus of many physiological investigations (e.g., Hilbish and Koehn 1985), recent studies have demonstrated greater warm temperature performance in *M. galloprovincialis* relative to *M. edulis* that is consistent with their biogeographic distributions (Fly and Hilbish 2012; Fly et al. 2015). Indirect evidence for interspecific differences in abiotic sensitivities also comes from variation in the distribution and abundance of introduced *M. galloprovincialis* and local water temperature, sun-exposure (Sarver and Foltz 1993; Schneider and Helmuth 2007; Schneider 2008), and ambient salinity (Braby and Somero 2006b). For example, Braby and Somero (2006b) showed that *M. trossulus* dominates higher latitudes characterised by high levels of fresh water input, despite relatively high temperatures at which warm-adapted *M. galloprovincialis* is predicted to thrive. Consistently, recent experimental studies have demonstrated differences in osmotic-stress responses among *Mytilus* species evidenced by differentiation in cardiac function (Braby and Somero 2006a), transcriptomic response (Lockwood and Somero 2011b) and protein abundance during osmotic-stress (Tomanek et al. 2012); however,

minor disparities in transcriptional responses between species (Lockwood and Somero 2011b) and inconsistent patterns at the proteomic level (Tomanek et al. 2012) have precluded a clear understanding of the genetic mechanisms underlying interspecific variation in salinity tolerance.

In contrast to salinity tolerance, the thermal physiology of *Mytilus* mussels is relatively better resolved. Experimental evidence based on targeted (Hofmann and Somero 1996) and broad-scale changes in gene expression (Lockwood et al. 2010), proteomic response (Tomanek and Zuzow 2010; Fields et al. 2012; Tomenak 2012), enzyme activity (Fields et al. 2006; Lockwood and Somero 2012), and cardiac physiology (Braby and Somero 2006a) have consistently demonstrated higher warm-temperature tolerance in *M. galloprovincialis* relative to *M. trossulus*. Most notably, two functional genetic studies have provided insight into the genetic basis of these physiological differences: Fields et al. (2006) identified a single amino acid change in the metabolic enzyme cytosolic malate dehydrogenase (cMDH) that accounted for the observed functional differences between orthologs mediating temperature tolerance in *Mytilus* species (Fields et al. 2006). While the presence of synonymous polymorphisms between the two *M. galloprovincialis* cMDH alleles sequenced may suggest substantial genetic variation, the authors posit that the observed polymorphism likely represents standing genetic variation from the native range segregating in the introduced population (Fields et al. 2006). In another study, Lockwood and Somero (2012) provided evidence that two fixed (and functionally non-conserved) amino acid differences in isocitrate dehydrogenase (IDH) were associated with differences in enzyme function and binding affinity between orthologs at high temperatures, potentially accounting for differences in heat sensitivity among congeners. While there are likely numerous molecular components underlying thermal tolerance difference between native and invasive *Mytilus* species, these findings demonstrate that few amino acid mutations may have large phenotypic effects governing temperature adaptation. Importantly, such data provide the basis of predictions about which genes may underlie (pre)adaptive genomic variation relating to apparent differences in invasive potential observed between *Mytilus* species. I explore these predictions in empirical Chapter 2.

Multiple global introductions

As a successful non-native species, *M. galloprovincialis* has established introduced populations in California (McDonald and Koehn 1988), South Africa (Grant and Cherry 1985; Branch and Steffani 2004), Australia and New Zealand (Westfall and Gardner 2010; Garder et al. 2016), Japan (Wilkins et al. 1983), Chile (Daguin and Borsa 2000; Westfall and Gardner 2010, 2013; Oyarzún et al. 2015; Larraín et al. 2018) and other parts of temperate East Asia (Hong Kong, Lee and Morton 1985; Korea,

McDonald et al. 1991; Daguin and Borsa 2000) as a result of accidental and deliberate human transport for aquaculture. Genetic surveys have also identified northern *M. galloprovincialis* haplotypes in Hawaii (Apte et al. 2000); however, there is no strong evidence to suggest that introduced populations are established in this region. Additionally, interspecific admixture with outgroup taxon, *M. edulis*, has led to pronounced genetic differentiation between *M. galloprovincialis* lineages from the Mediterranean Sea and the Atlantic coast of Europe (Quesada et al. 1995; Fraïsse et al. 2015), and both divergent lineages have been implicated in independent successful invasions into California (McDonald and Koehn 1988; Daguin and Borsa 2000) and South Africa (Branch and Steffani 2004) respectively. Given its global distribution, contemporary spread and impacts on native communities, Mediterranean *M. galloprovincialis* is recognised as one of the world's worst invasive species (Lowe et al. 2000). This species is currently classified under the highest threat categories for both 'Ecological Impact' and 'Management Difficulty' in a marine invasions database overseen by The Nature Conservancy (Molnar et al. 2008).

Invasive *M. galloprovincialis* has been particularly problematic in intertidal communities throughout California and South Africa, where its invasive status is well documented. Historical records suggest that Mediterranean *M. galloprovincialis* was first introduced to southern California in the 1900's (likely facilitated by ballast water from Japan; Geller et al. 1994), where its initial introduction and subsequent spread were overlooked due to inaccurate identification as native *M. trossulus* (McDonald and Koehn 1988; Geller et al. 1994). To date, *M. galloprovincialis* has spread over 1100 km across California and populations have displaced *M. trossulus* along the southern portion of its range, through Monterey Bay, California, where the two species hybridise (McDonald and Koehn 1988; Rawson et al. 1999; Geller 1999). Populations, however, have not become established in sites north of Monterey Bay despite the presence of *M. galloprovincialis* larvae (Geller et al. 1994) and haplotypes in some northern regions (reviewed in Wonham 2004; Anderson et al. 2002; Shields et al. 2010). In Africa, Atlantic lineage *M. galloprovincialis* currently occupies the southern and western coast of South Africa and south Namibia (Branch and Steffani 2004), where it was introduced in the 1970s (Grant and Cherry 1985; Branch and Steffani 2004). Its present-day distribution continues to expand at an average rate of approximately 115 km and 25 km year^{-1} on the western and southern coasts respectively (Branch and Steffani 2004), where several negative impacts have been documented, including local displacement of indigenous species (*e.g., Aulacomya ater* and *Perna perna*), and damage to coastal habitats (Robinson et al. 2007). In Australia, the prevalence of *M. galloprovincialis* haplotypes along temperate coastlines suggests that *M. galloprovincialis* may be established in some coastal regions and harbours (Dias et al. 2014). Documenting its introduced distribution and the sources of introduced populations, however, has been limited by the power of

traditional genetic approaches to differentiate introduced *M. galloprovincialis* from a morphologically indistinguishable and taxonomically contentious native *Mytilus* species, *Mytilus planulatus* (discussed below). Multiple *M. galloprovincialis* introductions worldwide provide an opportunity to comprehensively explore whether genetic responses to introduction occur in parallel across a wide range of geographically independent coastal habitats. Such data can shed light on the relative contributions of post-introduction genetic changes in the spread of introduced populations, and the traits that may underlie their resilience (e.g., Prentis et al. 2008; Bernardi et al. 2016). I address this hypothesis in Chapter 4.

Interspecific hybridisation

Mytilus galloprovincialis has a history of hybridisation and introgression with closely related *Mytilus* congeners where their ranges overlap throughout its native and invasive ranges (Bierne et al. 2003b; Japan, Brannock et al. 2009; California, Rawson et al. 1999). The *M. galloprovincialis* x *M. edulis* mosaic hybrid zone in southwestern Europe (where *M. galloprovincialis* is native) extends from Spain to Ireland and has been studied extensively (e.g., Gosling 1992; Bierne et al. 2003b; Hilbish et al. 2012; Roux et al. 2014a; Fraïsse et al. 2014; 2015). Genetic surveys of several well-defined contact zones in this region have demonstrated semi-permeable barriers to gene flow and asymmetric introgression between hybridising sister taxa (Rawson and Hilbish 1998; Bierne et al. 2003b; Gosset and Bierne 2013; also see Riginos and Cunningham 2005). For example, a multilocus scan of genetic variation in the Atlantic *M. galloprovincialis* x *M. edulis* hybrid zone identified locus-specific introgression from *M. edulis* into one *M. galloprovincialis* population, but not into another parapatric population (Fraïsse et al. 2014). While discriminating among selective and neutral hypotheses underlying signatures of introgression is a challenge (Bierne et al. 2011), the authors used a combination of molecular analyses (i.e. chromosomal walking around the outlier locus) and Bayesian demographic modelling to demonstrate that the transfer and fixation of alleles between species is likely a result of adaptive introgression at one locus (Fraïsse et al. 2014; also see Gosset and Bierne 2013). Genomic analysis of replicated zones of contact among *Mytilus* congeners (*M. galloprovincialis, M. edulis* and *M. trossulus*) in the same biogeographic region revealed that interspecific introgression from *M. edulis* into *M. galloprovincialis* has contributed significantly to strong intraspecific differentiation within *M. galloprovincialis*, and a large number of new genetic candidates for adaptive introgression were identified (Fraïsse et al. 2015). While several reproductive isolating mechanisms are known among *Mytilus* species, including asynchronous spawning (Secor et al. 2001; Gilg et al. 2007), differences in habitat preference (Gosling and McGrath 1990), assortative fertilisation (Bierne et al. 2002) and the presence of Dobzhansky-Muller incompatibility substitutions

at several loci (Bierne et al. 2006), these findings suggest that a history of past interspecific introgression may be an important source of adaptive evolution in *Mytilus* species (Fraïsse et al. 2015). Such evidence has important implications for interpreting genomic divergence among introduced and native *M. galloprovincialis* populations and patterns of introgression with native *Mytilus* taxa.

In its introduced range in California, invasive *M. galloprovincialis* and native *M. trossulus* form a hybrid zone that has extended more than 540 km along the coastline from Monterey Bay to Humbolt Bay in central California (Rawson et al. 1999; Braby and Somero 2006b; Hilbish et al. 2010). A recent study found weak asymmetric patterns of introgression from native *M. trossulus* into the invasive *M. galloprovincialis* neutral genomic background (Saarman and Pogson 2015) that are consistent with the prediction that populations at the edge of a contact zone should experience stronger introgression into the invading species (Currat et al. 2008). Because only 2.2 % of the filtered RADseq markers fell within coding regions (Saarman and Pogson 2015), gene-specific introgression patterns relating to adaptive processes could not be determined. In Australia, evaluating the nature and extent of introgression between introduced *M. galloprovincialis* and the endemic *M. planulatus* lineage has been limited to traditional genetic approaches implementing one or two loci (e.g., Borsa et al. 2007; Daguin and Borsa 2000; Westfall and Gardner 2010; 2013). The existence of *M. planulatus* in Australia is evidenced by fossil shells identified in Aboriginal middens in Australia (i.e. New South Wales, Donner and Jungner 1981; South Australia, Hope et al. 1977; Tasmania, Colhoun et al. 1982; reviewed in McDonald et al. 1991; Hilbish et al. 2000), suggesting that the presence of endemic *Mytilus* mussels predating European contact. Despite a number of genetic investigations, however, distinguishing among introduced and native *Mytilus* taxa has been hampered by low genetic differentiation between populations and high levels of discordance between nuclear and mitochondrial markers (Hilbish et al. 2000; Gérard et al. 2008; Westfall and Gardner 2010). Specifically, taxonomic delineation of the endemic taxon has been confounded by the presence of interspecific northern *Mytilus* genotypes at few putatively diagnostic nuclear markers, implicating either possible contemporary introgression or close genetic affinities to both northern *M. galloprovincialis* and *Mytilus edulis* (Borsa and Daguin 2000; Borsa et al. 2007; Westfall and Gardner 2010, Ab Rahim et al. 2016). Because both incomplete lineage sorting and interspecific gene flow can lead to shared polymorphism among species (Marko and Hart 2011), a key unresolved question involves whether signatures of mixed ancestry reflect unsorted ancestral polymorphisms or recent introgression, following secondary contact with introduced congeners. In Chapter 3, I investigate the genomic outcomes of secondary contact in two contact zones between introduced northern *M.*

galloprovincialis and native Australian *M. planulatus* to provide insight regarding the role of hybridisation and introgression in successful marine introductions.

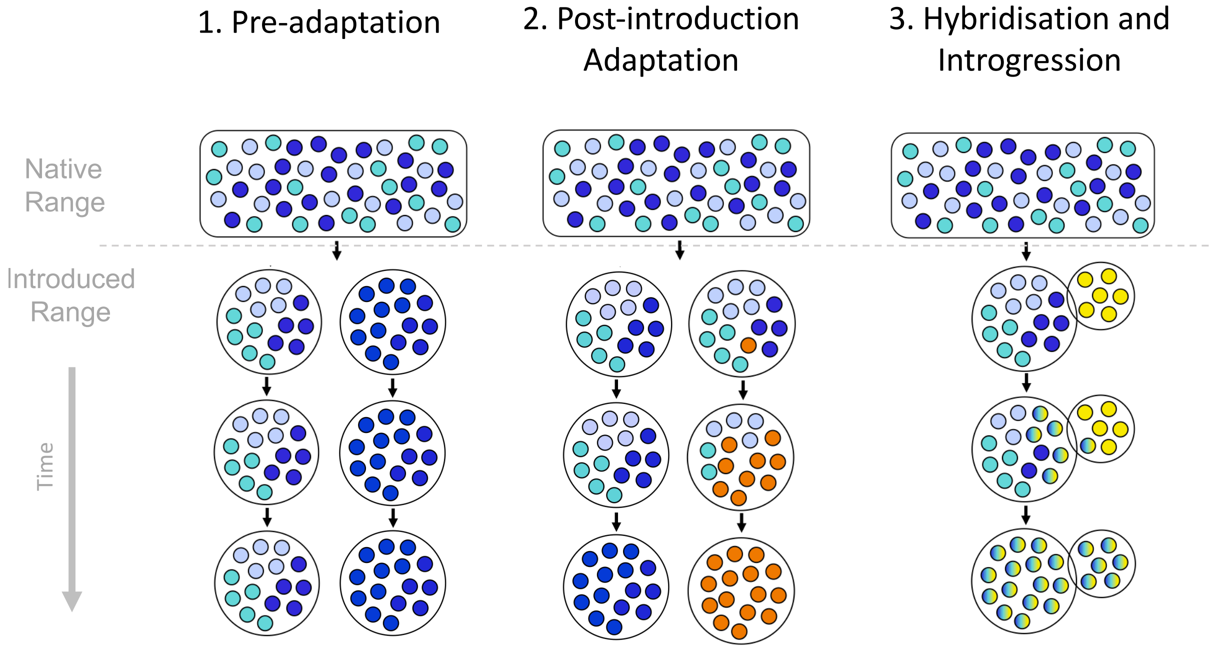


Figure 1-1. Conceptual illustration of three broad evolutionary scenarios (pre- and post- introduction genetic changes, including hybridisation) by which introduced species are able to persist in environments outside of their natural geographic range. Coloured circles indicate colonising genotypes.

Chapter 2. Comparative genomics reveals divergent thermal selection in warm- and cold-tolerant marine mussels

Abstract

Investigating the history of natural selection among closely related species can help delineate how genomes diverge in response to disparate environmental pressures. Integrating molecular evolutionary approaches with knowledge of genomic functions can elucidate how these evolutionary outcomes may affect ecologically-relevant traits such as species distribution limits and invasive success. Here, I integrate transcriptome-wide sequence data with *a priori* knowledge of genomic functions associated with temperature adaptation in physiological studies to assess the contributions of thermal selection to the divergence of the warm-tolerant and invasive marine mussel *Mytilus galloprovincialis,* from three cold-tolerant, non-invasive congeners. Concordant evidence of positive selection in codon model tests and intraspecific polymorphism and divergence analyses revealed the strongest signatures for positive selection in a small set of loci involved in oxidative stress. These selective targets are consistent with the hypothesis that genomic functions differentiating the strongest species-specific responses to thermal stress in physiological studies have also experienced parallel sequence divergence at the molecular level. Furthermore, positively selected loci share functional similarities with known thermotolerance candidate genes also under positive selection in independent tests. These findings implicate a contribution of temperature-dependent selection in the divergence of warm- and cold-adapted *Mytilus* species and provide molecular evidence supporting the hypothesis that differential strategies in coping with intracellular reactive oxygen species may explain interspecific variation mediating the relative successes of native and introduced species. Broadly, these findings demonstrate that independent lines of experimental evidence from whole organism physiological studies can be leveraged to inform hypotheses regarding functional genomic variation involved in the evolutionary divergence of closely related non-model species.

Introduction

Comparisons among closely related species adapted to contrasting niches affords the opportunity to investigate how genomes diverge in response to disparate environmental conditions (Oliver et al. 2010). Integrating molecular evolutionary approaches with knowledge of genomic functions can reveal how natural selection has shaped genes underlying ecologically-relevant traits such as species distribution limits and invasive species success. In cases where specific amino acids are known to affect protein function, analyses of intraspecific polymorphism and divergence in candidate sequences can be used to directly study functional variation in natural populations or taxa adapted to divergent environmental conditions (Dean and Thornton 2007; Storz and Wheat 2010; Storz et al. 2015). Such methods can yield insight into the selective mechanisms by which functional variation evolves or is maintained, and provide corroborative evidence about the adaptive significance of candidate loci (e.g., Storz et al. 2009; Linnen et al. 2009; Natarajan et al. 2015). The polygenic basis of many traits governing ecological and physiological tolerance, however, presents significant challenges for linking causative genetic loci to phenotypic variation mediating ecologically important traits (Storz and Wheat 2010; Rockman 2012; Le Corre and Kremer 2012). Furthermore, the lack of annotated reference genomes for many non-model species coupled with difficulties in establishing inbred lines and multigenerational pedigree relationships, precludes the use of quantitative genetic tools for informing hypotheses about ecologically relevant genetic variation.

Functional genomic studies investigating genome-wide expression are a useful approach to interrogate the complex relationship between genomic variation and species adaptive responses (Somero et al. 2017). Whole-genome patterns of gene expression across thousands of genes can be used to classified loci into putatively causative functional groups underlying to physiological responses ecological stressors (Lockwood et al. 2010; Somero 2012). Such functional classifications can be highly informative for identifying molecular traits under selection even in cases where genomic resources are scarce or traits governing ecological and physiological tolerance are controlled by many co-dependent genes (Rockman 2008). Alternatively, reverse-ecology approaches such as comparative genome scans of substitution rate variation are a powerful method for detecting protein-coding loci and genomic functions as targets of natural selection (Anisimova and Liberles 2007; Li et al. 2008; Ellegren 2008). Indeed, studies involving genomic datasets from multiple lineages have demonstrated high relative rates of molecular evolution and positive selection among loci related to reproduction, immune defense and sensory perception between diverging taxa (e.g., Clark et al. 2003; Nielsen et al. 2005; Holloway et al. 2007; Roux et al. 2014b). Yet, without the context of the organism's environment or specific hypotheses regarding the role of key environmental pressures on

species divergence (e.g., Oliver et al. 2010; Barreto et al. 2010; Ladner et al. 2012; Koester et al. 2013), genome scans may encourage false speculations about the adaptive significance of selected loci (Pavlidis et al. 2012; Storz et al. 2015). As a result, analyses relating molecular signatures of positive selection to lineage-specific adaptations could be strengthened by independent experimental evidence under which *a priori* hypotheses about the ecologically-relevant genomic traits involved can be formulated (Pavlidis et al. 2012). Such analyses can complement direct experimental measurements of gene expression to provide independent molecular evidence regarding the adaptive significance of the genomic functions involved in species-specific physiological responses.

Here, I apply an integrative approach combining comparative molecular evolutionary methods with knowledge from whole-organism physiological studies to compare the genomes of closely related warm-adapted and cold-adapted species in the marine mussel genus *Mytilus* that have evolved under disparate temperature environments. I leverage a large body of physiological experimental information to develop hypotheses about the genomic functions that have experienced positive selection for species divergence in warm-tolerant *Mytilus galloprovincialis*, one of the world's most widespread invasive species (McDonald et al. 1991; Lowe et al. 2000), relative to three cold-tolerant and non-invasive congeners. Over the last two decades, the *Mytilus* genus has become an important comparative system for studying the links between species distribution limits and thermal physiology, and there is a growing body of experimental evidence for understanding key molecular phenotypes involved in environmental adaptation in this group (Lockwood et al. 2015). The *Mytilus* genus therefore holds many advantages of a nearly model organism, with a breadth of ecophysiological studies and functional insights into the genetic basis of thermal adaptation, despite life history features that do not allow for accessible inbred lines or multiple generations required to link physiological differences to their genetic basis.

Northern hemisphere *Mytilus* include the ribbed mussel, *Mytilus californianus,* and smooth-shelled blue mussel species *Mytilus trossulus*, *Mytilus edulis*, and *M. galloprovincialis*. The Mediterranean native *M. galloprovincialis* is the only *Mytilus* species that is known to have established invasive populations outside of its native range, including introductions in California, South Africa, Australia and New Zealand (McDonald and Koehn 1988; Branch and Steffani 2004; Westfall and Gardner 2010). Ecophysiological studies investigating the role of environmental adaptation in setting species distributions have largely focused on *M. galloprovincialis* in the Northeastern Pacific, where introduced populations have displaced native *M. trossulus* along the southern parts of its range (Figure 2-1A; Rawson et al. 1999; Geller 1999; Braby and Somero 2006b). The displacement of *M. trossulus* and exclusive establishment of *M. galloprovincialis* from Baja California, Mexico to Monterey Bay,

California, suggests that interspecific differences in environmental tolerance between *Mytilus* species underlie their present distributions and relative successes as native and introduced species (Lockwood et al. 2015; Figure 2-1B). Indeed, comparative physiological and biochemical investigations have demonstrated higher warm-temperature tolerance in introduced *M. galloprovincialis* relative to cold-adapted congeners along the coast of California. Specifically, marked species-specific differences have been evidenced by targeted (Hofmann and Somero 1996) and broad-scale changes in temperature-dependent gene expression (Lockwood et al. 2010), proteomic response (Tomanek and Zuzow 2010; Fields et al. 2012; Tomanek 2012), metabolic enzyme activity (Fields et al. 2006; Lockwood and Somero 2012), and cardiac function (Braby and Somero 2006a; reviewed in Lockwood and Somero 2011a; Lockwood et al. 2015).

Physiological investigations have also provided insight into the genetic basis of thermal tolerance differences between *Mytilus* species (Lockwood et al. 2015). Functional genetic studies species have demonstrated sequence divergence in two metabolic enzymes, including fixed amino acid substitutions that have been sufficient to explain observed functional shifts in protein heat sensitivity and physiological differences between warm-adapted and cold-adapted congeners (cytosolic malate dehydrogenase, cMDH, Fields et al. 2006; isocitrate dehydrogenase, IDH, Lockwood and Somero 2012). Investigations across broader arrays of expressed genes and proteins have also identified key genomic functions mediating interspecific variation in thermotolerance (Lockwood et al. 2010; Tomanek 2012). Specifically, genes encoding small molecular chaperones (i.e. heat shock proteins) and oxidative stress proteins were associated with the strongest species-specific responses to thermal stress between *M. galloprovincialis* and *M. trossulus* in both expression and proteomic studies (Lockwood et al. 2010; Tomanek and Zuzow 2010; Tomanek 2012; Fields et al. 2012). Other genomic functions associated with the most evident interspecific transcriptional and proteomic changes to thermal stress included proteolysis, energy metabolism, cell signalling, ion transport, and cytoskeletal organisation (Lockwood et al. 2010; Tomanek and Zuzow 2010). Collectively, this body of research has provided extensive experimental evidence supporting a role of temperature-dependent adaptation in the divergence of *M. galloprovincialis* from cold-adapted congeners. The *Mytilus* system therefore provides a unique opportunity to compare the genomes of closely related invasive and non-invasive species in the context of known physiological adaptations that may also mediate apparent variation in invasive potential.

In this study, I test the hypothesis that genomic functions experimentally linked to species-specific adaptations to temperature stress show correlated evolutionary divergence in warm-tolerant and invasive *M. galloprovincialis*, compared to three cold-tolerant, non-invasive congeners. I use a

transcriptome approach to focus on protein-coding genes and extend interspecific comparisons to a novel dataset of 2719 *de novo* assembled orthologous genes from four northern hemisphere *Mytilus* species: *M. californianus*, *M. trossulus*, *M. edulis* and *M. galloprovincialis*. Based on existing physiological information, I expected that temperature-relevant genomic functions associated with physiological responses to heat stress have experienced positive selection for species divergence in *M. galloprovincialis.* I combined codon model tests with intraspecific polymorphism and divergence data from the *M. galloprovincialis* native range to separate the effects of demography, purifying and positive selection on the evolution of orthologous sequences (Figure S2-1). I then looked for concordance between independent selection analyses and used the results to determine whether predicted temperature-relevant genomic functions (identified in physiological studies) were prevalent among genes showing evidence of positive selection considering both divergence and polymorphism data. As a secondary objective, I investigated whether specific loci previously identified as thermotolerance candidates were molecular targets of positive selection in *M. galloprovincialis* in independent selection tests. I examined whether positively selected loci were enriched for specific thermotolerance candidate genes, or key functional Gene Ontology (GO) categories as diagnostic of the molecular functions contributing to divergence in *M. galloprovincialis*. The findings of this Chapter corroborate independent lines of experimental physiological evidence for temperature adaptation with concordant signatures of accelerated evolution in the same genomic functions differentiating warm-adapted and invasive species.

Methods

Sample collection, assembly filtering and ORF identification

Mussels were collected from natural rocky intertidal or subtidal environments. To minimise the possibility of sampling hybrid individuals, specimens were collected from geographic locations where species are endemic and believed to occur in allopatry (Table 2-1). Smooth-shelled species *M. galloprovincialis*, *M. edulis* and *M. trossulus* were genotyped for the species diagnostic size polymorphic marker Glu-5' (Rawson et al. 1996) to confirm species identity. I also performed principal component analysis (implemented in R; R Development Core Team 2017) of preliminary transcriptome-derived variants as an additional measure to exclude introgressed individuals. Total RNA was extracted from 10-20 mg of mantle tissue (preserved in RNAlater) from three individuals per taxon (n=12) using the RNeasy Plant Mini Kit (Qiagen, MD, USA) and following the Animal Tissues protocol with an additional DNAse treatment to remove genomic DNA. Additional native *M. galloprovincialis* samples were collected from two Mediterranean (n=7) and one Atlantic (n=5)

population for population-level analyses (Table 2-1). RNA quality and concentration was quantified using an Agilent 2100 Bioanalyzer prior to sequencing. A total of 24 individual cDNA libraries were constructed and barcoded using the TruSeq stranded mRNA kit (Illumina), with average insert sizes of 250-300 bp. Paired-end fragments (125 bp) were sequenced across three lanes of an Illumina Hiseq2000 (v4) and took up 50% of each Illumina lane.

Transcriptome data processing and assembly

Demultiplexing of individual sequence data was performed by the sequencing provider at the Queensland Brain Institute (QBI, Brisbane, Australia). All subsequent processing and analysis was completed on the QRIScloud computer cluster managed at the University of Queensland. Read quality and adapter contamination was assessed using FastQC. I used Trimmomatic (v0.36) (Bolger et al. 2014) to remove residual adapter sequences from paired reads; quality trimming was performed using a 4 bp sliding window and a phred-scale average quality score of 20. Processed reads were also subjected to a minimum size filter of 50 bp. To reduce redundancy among high-coverage reads (including rRNA contaminants) and to discard associated sequence errors, each dataset was digitally normalised using Trinity's insilico_read_normalization.pl script with a default kmer size of 25 and maximum read coverage of 50 (Grabherr et al. 2011). Overlapping paired reads were merged using FLASH v1.2.11 (Magoč and Salzberg 2011) with a minimum overlap length of 10 bp.

The resulting read datasets were pooled for each taxon and used to create *de novo* assemblies using Trinity v2.0.6 (Grabherr et al. 2011) with default parameters. Because downstream analyses required obtaining putative full-length orthologous genes, each assembly was filtered to reduce the inclusion of redundant contigs representing alternatively spliced isoforms, partial transcripts or divergent alleles, as well as close paralogs and non-coding RNA transcripts. The output of the Trinity pipeline was filtered to retain only the longest isoform per gene group. Transcripts with high sequence similarity were clustered using Cd-Hit-Est (Li and Godzik 2006; Fu et al. 2012) with a minimum sequence identity threshold of 95% of the shortest sequence. Transdecoder (Haas et al. 2013) was used to reduce assemblies to protein-coding sequences with significant matches to the Pfam protein database. To minimise the possibility of misalignment of orthologous genes between species or the inclusion of non-homologous sequences, I retained only genes with complete predicted open reading frames (ORF) greater than 100 amino acids in which both a start and stop codon was identified.

Transcriptome assembly quality and gene annotation

Predicted coding sequences were queried against the Uniprot-Swissprot protein database using blastx with an e-value threshold of 10^{-3} for significant matches. Subsequently, contigs with significant blast matches to likely environmental contaminants, including bacteria, fungi, viruses, protists (Alveolata), green (Viridiplantae) and red algae (Haptophyceae), and other eukaryotic contaminants (i.e. Euglenozoa) were removed using Biopython v1.68 and R (R Development Core Team 2017). Additional annotation was carried out using the Trinotate 3.0.1 annotation pipeline using blastx and the Pfam protein database. To assess the completeness and expected gene content of each filtered assembly, the *M. galloprovincialis* assembly was also subjected to a blastn search (e-value=10^{-3}) against the *M. galloprovincialis* draft genome (including 10,891 protein coding genes) (Murgarella et al. 2016). I also looked for the proportion of Benchmarking Universal Single Copy Orthologs (BUSCO; Simão et al. 2015), a set of 429 single-copy genes selected from OrthoDB that are shared by higher eukaryotes. Additional assembly quality metrics, such as the proportion of aligned reads were assessed by mapping the FLASH-merged normalised reads onto each assembly using Bowtie2 and employing a very sensitive (end to end) alignment approach (Langmead and Salzberg 2012).

Ortholog identification and alignment

To identify putative orthologous genes, I used the best-hit blast method Proteinortho v5.13 (Lechner et al. 2011). Clusters of putative orthologous and co-orthologous genes (i.e. multiple orthologs per taxon) termed orthogroups, were identified under an e-value threshold of 10^{-10} and alignment parameters including 50% query sequence coverage, 25% minimum sequence identity, and 95% minimum similarity thresholds for additional blast hits. Only orthogroups containing at least one sequence from each taxon were retained for downstream analyses. In cases where multiple transcripts from the same taxon were identified as co-orthologous genes, the phylogeny-based PhyloTreePruner pipeline (Kocot et al. 2013) was used to choose a single species representative for each group. I collapsed poorly supported nodes with bootstrap support of less than 60% into polytomies and applied a pairwise distance method (-r flag; SCaFoS; Roure et al. 2007) to select the best orthologous sequence if multiple sequences formed a monophyletic clade from the same taxon. As input for PhyloTreePruner, nucleotide alignments of orthogroups were generated using a codon-based alignment algorithm in PRANK (-F -codon option; Löytynoja and Goldman 2008), which takes evolutionary relationships into consideration to increase the accuracy of aligning homologous sequences in the presence of indel variation (Jordan and Goldman 2012). Individual gene trees were generated for each nucleotide alignment using FastTree2 (Price et al. 2010) with default options (Jukes-Cantor with CAT approximation) and gamma likelihood approximation. Prior to selection analyses, alignments were screened for the effects of intragenic recombination (which can influence

the tree topology of individual sites) using the single break point method implemented in HyPhy (Kosakovsky Pond et al. 2006). The GARDprocessor.bf module was used to assess whether significant breakpoints are due to topological incongruence using the Shimodaira-Hasegawa test at a significance level of P=0.05. In the presence of significant recombination breakpoint, alignments were partitioned and a new topology was generated for each gene partition prior to selection analyses. Branch lengths and node labels were removed from the newick formatted tree files using the R package Ape (Paradis et al. 2004).

Branch-site models of molecular evolution

To identify genes under positive selection in the *M. galloprovincialis* lineage, I carried out branch-site tests of positive selection for each gene alignment using the codeml method in PAML v4.8 (Yang et al. 2005; Zhang et al. 2005; Yang 2007). Because the rate of synonymous substitution approximates neutral evolution, the ratio of nonsynonymous substitutions per nonsynonymous site (*dN*) relative to synonymous differences (*dS*), omega ($\omega = dN/dS$), can be inferred as positive selection ($\omega > 1$) for amino acid substitutions or purifying selection ($\omega < 1$) against nonsynonymous changes. Given an *a priori* hypothesis, the branch-site method implements a maximum likelihood codon model to compare ω variation among a set of foreground branches (i.e. the lineage to be tested for positive selection) and background branches in the focal gene tree to identify whether selection is significantly different on the specified foreground lineages. The selection model permits all sites in a foreground branch vary in ω, which can belong to one of three rate classes: $\omega_0 < 1$, $\omega_1 = 1$ and ω_2 estimated as a free parameter relative to the background rate of evolution, which can reflect purifying selection (class 2a) or neutral evolution (class 2b). Positive selection is inferred if $\omega_2 >> 1$ and the likelihood of the selection model is significantly greater than the likelihood of the null model where ω for the foreground lineages is set to neutral evolution ($\omega = 1$). The likelihood ratio test statistic (LRT), twice the log-likelihood difference ($-2\Delta l$), was compared to the X^2 distribution (critical value=3.84; P=0.05) with one degree of freedom to determine if the selection model significantly improved the fit to the genetic data. In cases where the LRT is significant, Bayesian inference (Bayes Empirical Bayes; BEB) is used to estimate the posterior probability that individual codons in the foreground branch are under positive selection (Yang et al. 2005).

To account for gene tree discordance among individual genes that may result from incomplete lineage sorting (Mendes and Hahn 2016), I generated individual gene trees for each alignment; I specified the *M. galloprovincialis* lineage as the foreground branch (Figure 2-1C) and used the inferred

topologies as input for the PAML analyses. Codeml analyses were conducted using pairwise deletion of sequence gaps to estimate ω for sites that included indels between orthologous sequences. Codons were inferred to be under positive selection if the BEB critical posterior probability value was P≥0.95. To ensure that codon misalignment does not affect our conclusions about site-specific positive selection (e.g., Mallick et al. 2009), alignments of genes under selection were visually inspected *post hoc* to identify unreliable alignments at the selected sites; I searched for potential signs of misalignment including clusters of codons under selection in relatively conserved regions and sites under selection occurring within five codons of an indel or stop/start codon in one or more taxa (Markova-Raina and Petrov 2011). Following visual inspection, one alignment was removed due to signs of obvious misalignment.

Analyses of positive selection using polymorphism variation

Additional power to detect lineage-specific selection can be obtained through combining divergence data and data on population-level genetic variation (Bierne and Eyre-Walker 2004; Christe et al. 2017). Given the relatively recent divergence between *M. galloprovincialis* and *M. edulis* (2.5 million years; Seed 1992; Roux et al. 2014a), as well as historical introgression (Fraïsse et al. 2015), I expected that incomplete lineage sorting may be pervasive in some genomic regions. Moreover, it is well known that positive selection may be overestimated if species have split only very recently (<10Ne generations; Keightley and Eyre-Walker 2012). To minimise the effects of shared polymorphism and to keep species comparisons consistent with ecophysiological studies, I compared intraspecific polymorphism to divergence relationships of replacement and synonymous substitutions in two genetically divergent native *M. galloprovincialis* lineages from the Mediterranean (n=10) and Atlantic (n=5) (Fraïsse et al. 2015), relative to *M. trossulus*. I aligned RNAseq reads to the reference dataset of 2719 genes to obtain multiple *M. galloprovincialis* sequences aligned against an orthologous *M. trossulus* sequence as the outgroup. RNAseq reads were mapped to reference contigs using Bowtie2 and the very sensitive local method (Langmead and Salzberg 2012) and PCR duplicates were removed using Picard MarkDuplicates (http://picard.sourceforage.net). Variants were identified using Freebayes (https://github.com/ekg/freebayes). To minimise the inclusion of erroneous or low quality variants, I removed singletons, indel or complex variants, and columns with more than 70% missing data from the resulting VCF file using vcftools (Danecek et al. 2011). The remaining Single Nucleotide Polymorphisms (SNPs) were filtered for a minimum quality of 30, and per individual genotype calls below a genotype quality score of 30 and depth coverage of 5 reads were removed. Using a total of 44,750 SNPs, I reconstructed consensus haplotype sequences for each individual using filtered variants and reference sequences in BCFtools v1.3.1, such that heterozygous

sites were consecutively assigned a strand. The resulting 2498 gene alignments were subset by population, merged with the corresponding outgroup sequence, and aligned using PRANK as outlined above (Löytynoja and Goldman 2008). From each dataset, I discarded sequences with premature stop codons (annotated in SNPeff; Cingolani et al. 2012) and genes for which no variant information was retained following filtering. I calculated divergence and polymorphism counts for synonymous and replacement substitutions for each gene, for each locality, using the Polymorphorama perl script (Bachtrog and Andolfatto 2006; Haddrill et al. 2008). I did not consider synonymous polymorphisms with frequencies below 10% (i.e. Fay et al. 2001; Bierne and Eyre-Walker 2004; Charlesworth and Eyre-Walker 2008).

I performed two tests of selection on each population dataset to determine whether adaptive fixation could be detected using a population genetic framework. Because strong population structure may lead to false classification of fixed substitutions (between divergent lineages) as polymorphisms and therefore obscure the true signal of positive selection, the subsequent tests were performed on Mediterranean and Atlantic *M. galloprovincialis* populations independently. Assuming that the contribution of polymorphism substitutions to adaptive evolution is relatively small, I calculated *alpha*, the proportion of nonsynonymous substitutions fixed in each lineage by positive selection for individual genes for which polymorphism information was obtained, based on equation 2 in Smith and Eyre-Walker (2002):

$$alpha = 1 - D_s P_n / D_n P_s \; ;$$

where D_n, D_s, P_n, P_s are the numbers of replacement and synonymous substitutions and polymorphisms per gene, respectively. A G-test of independence with the Williams correction was carried out in R to assess the significance of *alpha* at $P \leq 0.05$ (McDonald and Kreitman 1991). Both codon models and McDonald-Kreitman based methods, however, assume that sites fixed under selection do not contribute to polymorphisms within populations. In turn, such methods do not account for slightly deleterious nonsynonymous polymorphisms under weak purifying selection that may contribute to false signals of adaptive fixation (Keightley and Eyre-Walker 2012; Mugal et al. 2013; but see Eyre-Walker and Keightley 2009). To establish how the contribution of deleterious polymorphisms and the relative strength of natural selection may vary among genes and *M. galloprovincialis* populations, I calculated the *DoS* statistic at the level of each gene with significant values of *alpha* for each locality (Stoletzki and Eyre-Walker 2011)

$$DoS = D_n / (D_n + D_s) - P_n / (P_n + P_s) \; ;$$

where positive *DoS* values indicate positive selection and negative *DoS* values suggest slightly deleterious mutations segregating in the focal population. To validate that adaptive divergence has occurred within the *M. galloprovincialis* lineage (rather than divergence driven in the outgroup lineage), I repeated polymorphism and divergence statistics *post hoc* using *M. edulis* as an outgroup for seven candidate loci showing concordant evidence of positive selection in both codon models and significant and positive *alpha/DoS* values in the initial analyses conducted against a *M. trossulus* outgroup (Table 2-2).

Identification of known thermotolerance candidate genes

I obtained loci previously identified as thermotolerance candidate genes in physiological investigations. Gene sets included candidates for (A) divergent functional adaptation (i.e. *cMDH*, *IDH,* Fields et al. 2006; Lockwood and Somero 2012); (B) species-specific differential expression under heat-stress (Lockwood et al. 2010); and (C) shared transcriptional responses in three *Mytilus* congeners under acute heat stress (Connor and Gracey 2011; Lockwood et al. 2015). A total of 273 thermotolerance candidate genes were assigned to reference transcripts using reciprocal best hits in Proteinortho v5.13 (Lechner et al. 2011), but only a small proportion of candidate loci (36/273) were matched to complete orthologs in all four taxa. Because I used a relatively stringent filtering approach to assign orthologous sequences, it is likely that many thermotolerance candidate loci were excluded from the reference dataset and downstream analyses, including genes expressed at low levels. Indeed, the majority of matched transcripts were found in trace counts in the transcriptome assemblies, such that I could only recover partial open reading frames in one or more species. Given our stringent quality filtering to assign orthogroups, I did not consider these partial transcripts that may lead to unreliable interspecific alignments and false inferences of positive selection.

Enrichment analyses of positively selected genes

To examine whether known thermotolerance candidate genes were overrepresented among positively selected gene sets relative to our complete dataset of 2719 analysed orthogroups, I applied Fisher's exact tests in R. I also determined whether any GO functional classes were significantly enriched among sets of genes with positive *alpha*/*DoS* values *or* high ω values in *M. galloprovincialis* using GO enrichment analyses in the R-bioconductor package topGO v2.26 (Alexa et al. 2006; Alexa and Rahnenfuhrer 2010). I assigned GO terms (based on blastx and the Pfam databases) to each reference orthogroup; I applied Fisher's exact tests using the topGO wieght01 algorithm and considered only GO-terms assigned to at least five genes to identify enriched *Biological Processes* categories at a

significance level of P≤0.001. Enrichment analyses were performed against two reference gene sets: 1) a dataset of 38,534 total protein-coding genes from the *M. galloprovincialis* reference transcriptome; and 2) a smaller annotated file of 2719 reference *M. galloprovincialis* transcripts analysed in this study (Supplementary Material).

Results

Sequencing and transcriptome assembly

Illumina sequencing yielded between 2.4 million to 28.5 million raw paired reads per RNA-seq library. Quality filtering and Trinity read normalisation resulted in a large reduction of reads (up to 84.6%) prior to assembly, which was largely due to the high presence of rRNA contaminants in the raw RNA libraries. Filtering and *de novo* assembly statistics are summarised in Table S2-1. The removal of alternatively spliced variants and transcripts with high sequence similarity resulted in 80,523 to 130,389 non-redundant transcripts per reference assembly. Based on these assemblies, I extracted transcripts containing likely ORFs, which resulted in 7,046 to 13,681 complete coding sequences per species for downstream analyses (Table S2-2).

Assembly quality, annotation and ortholog identification

Functional annotation against the Uniprot-Swissprot protein database gave significant blastx hits for ~70% of the predicted coding sequences for each species; only ~5% of transcripts were identified as environmental contaminants and removed from each dataset (Table S2-3). Greater than 97% of contigs in the *M. galloprovincialis* draft genome (although presumed to be highly fragmented; see Murgarella et el 2016; Table S2-4) recovered a significant hit to our *M. galloprovincialis* transcriptome assembly. Similarly, other quality assessments suggested high quality for *de novo* reconstruction of the *Mytilus* transcriptomes: The proportion of normalised RNAseq reads that successfully mapped back to each respective assembly ranged from 82.4% to 92.3%. Additionally, the percentage of full-length and partial BUSCO orthologs recovered from each assembly was greater than 87% (Table S2-5) suggesting high assembly completeness; this value is similar to proportions of single copy orthologs identified in other comparative genomic studies (e.g., Hodgins et al. 2014). Reciprocal best hits of complete coding sequences yielded 2719 high-quality orthologous groups including all four taxa, out of which 79 (3%) required pruning to exclude putatively paralogous sequences in at least one taxon. A significant recombination breakpoint was identified in nine

alignments; each alignment was partitioned and a new topology was generated for each gene partition priori to selection analyses.

Evidence of positive selection in the warm-tolerant M. galloprovincialis

I hypothesised that genomic functions significantly associated with temperature tolerance in physiological investigations have experienced positive selection for species divergence in warm-tolerant *M. galloprovincialis*. Branch-site analyses revealed 99 genes (3.6%) with significant evidence for positive selection ($\omega >> 1$) in *M. galloprovincialis* using the LRT (Figure 2-2). A smaller subset of 38 genes (1.4%) additionally returned at least one codon under positive selection at a BEB posterior probability cut-off of ≥0.95 (Figure 2-2). To minimise alignment error effects on selection tests, I filtered this subset of 38 genes to exclude loci with three characteristics that may be indicative of misalignments leading to high ω variation: 1) positively selected sites within five codons of an indel; 2) positively selected sites within five codons of a start/stop codon; and 3) positively selected sites within a cluster of selected codons. I termed this reduced gene set (n=19) as having the 'highest-confidence alignments' under positive selection.

In contrast to codon model tests, I identified a larger number of genes (n=175) with evidence of adaptive divergence in one or both populations ($P \leq 0.05$) (Figure 2-2). Estimates of positive *alpha* were greater than 0.5 for all 175 selected genes ($P \leq 0.05$), suggesting that more than half of sites fixed since the divergence of *M. galloprovincialis* and *M. trossulus* are adaptive. While the low number of polymorphic sites at some genes may account for relatively high *alpha* estimates, these findings are consistent with high proportions of fixed substitutions observed in other species (Smith and Eyre-Walker 2002; Galtier 2016).

I also calculated the direction of selection (*DoS*) statistic separately for each gene to assess the contribution of positive (*DoS*=positive) and relaxed purifying selection (*DoS*=negative) to inferences of evolutionary divergence and high ω variation in the *M. galloprovincialis* lineage. Significant and positive *DoS* values were obtained for 17 genes that were also positively selected in codon model LRTs, among which only seven genes also had one or more codons evolving under positive selection (BEB $P \geq 0.95$, Figure 2-2; Table 2-2); when only genes with the 'highest-confidence alignments' were considered, four genes showed evidence for positive selection in both analyses. Congruent findings among branch-site models and polymorphism and divergence analyses indicate strong evidence of positive selection for species divergence at these seven loci; concordant signatures of accelerated divergence in independent analyses are therefore unlikely to be false positive signals resulting from

alignment error or weakly selected mutations (Figure 2-2). I find that five out of seven genes showing the strongest signatures of positive selection have documented roles in temperature tolerance based on functional annotations in the GO database (Ashburner et al. 2000) and the UniProt Knowledge Database (The UniProt Consortium 2017), as well as relevant published literature (Table 2-2). When I considered only genes with the 'highest-confidence alignments', three out of the four remaining positively selected genes have known functions relevant to temperature tolerance. These selective targets fall within genomic functions associated with the strongest species-specific responses to temperature-related stressors in physiological studies, primarily molecular functions relating to oxidative defense (i.e. thiol regulation and cytoskeletal stabilisation; Table 2-2; Figure 2-2).

Because positive *alpha/DoS* estimates may be interpreted as divergence driven by adaptive evolution in the outgroup lineage rather than selection in the focal species for which polymorphism data is considered, I also conducted polymorphism and divergence analyses using *M. edulis* as an outgroup for the seven loci in which I identified adaptive divergence relative to *M. trossulus* (Table 2-2). Compared to *M. edulis*, I found significant and positive *alpha/DoS* values in one or both localities (Mediterranean or Atlantic) for all seven loci, implicating lineage-specific accelerated evolution in *M. galloprovincialis*. A small proportion of the genes analysed had significantly negative *DoS* values (17.6% and 21.8% in the Mediterranean and Atlantic populations respectively; Figure 2-2) that suggest weakly deleterious nonsynonymous substitutions segregating in the population; however, only one of these loci was also positively selected in branch-site analyses (using inference based on the LRT statistic), suggesting that accelerated rates of evolution along the *M. galloprovincialis* lineage are unlikely to be due to relaxed purifying selection at putative selected loci.

Thermotolerance candidate genes and gene ontology enrichment analyses

Out of 36 known thermotolerance candidate genes identified from physiological investigations (and present in our multispecies dataset), I found five genes under positive selection in either codon model tests or polymorphism and divergence-based analyses (Figure 2-2; Table 2-2); these specific thermotolerance candidates were not statistically overrepresented among any positively selected gene sets (Fisher's exact test $P>0.05$). Among these five loci, branch-site models identified two homologs under positive selection using inference based on the LRT statistic (Table 2-2). The first homolog had a blastx annotation to a small heat shock protein of the alpha-crystallin protein family and Heat Shock Protein 25 (HSP25; based on *M. californianus* EST annotations; Genbank accession ES737726), which is commonly up-regulated in heat-stressed *Mytilus* congeners (Connor and Gracey 2011; Lockwood et al. 2015). The branch site model also identified a single codon with a high

Bayesian posterior probability of positive selection at site 185 (albeit not significant; BEB P=0.905) encoding a functional amino acid substitution in *M. galloprovincialis* (Figure 2-3). The second homolog returned a best hit to protein *Shootin-1* (SHTN1) involved in actin filament binding (Connor and Gracey 2011; Lockwood et al. 2015). Among loci showing significantly positive *alpha/DoS* values, I also identified two commonly expressed thermotolerance candidate genes: oxidoreductase Glutathione S-transferase omega-1 (*GSTO1*) and Dipeptidyl peptidase 1 (i.e. cathepsin C; *CTSC*) (Table 2-2; Connor and Gracey 2011; Lockwood et al. 2015).

Notably, positive *alpha/DoS* values were identified for the thermotolerance candidate gene *cMDH*, corroborating that the observed amino acid differences between *M. trossulus* and *M. galloprovincialis* are likely adaptive (Fields et al. 2006; Table 2-2). *cMDH* has been implicated, through biochemical studies, as a genetic candidate for setting species-specific thermal limits in *Mytilus* mussels and other marine mollusks (*Mytilus* mussels, Fields et al. 2006; *Lottia* limpets, Dong and Somero 2008; Liao et al. 2017). *Post hoc* inspections of *cMDH* intraspecific read alignments revealed that *M. edulis* is polymorphic for valine and asparagine amino acids at the functional codon 114, while *M. galloprovincialis* appears to be fixed for asparagine based on available data. I therefore speculate that episodic selection for species divergence may have occurred at some point following the divergence of *M. trossulus* and a common ancestor predating the speciation of *M. galloprovincialis* and *M. edulis*. It is worth noting, however, that I did not find evidence of positive selection in the thermotolerance candidate gene IDH, which also has known point mutations involved in heat tolerance and antioxidant responses in *Mytilus* species (Tomanek and Zuzow 2010; Lockwood and Somero 2012).

I determined whether any GO functional categories were significantly overrepresented among positively selected genes showing either high ω or positive *alpha*/*DoS* values in *M. galloprovincialis*. I obtained GO terms for 21,185 out of 38,534 genes represented in the *M. galloprovincialis* protein-coding assembly. Positively selected groups of genes identified in PAML were consistently significantly enriched for GO terms associated with *sulfur compound metabolism* and *organonitrogen compound catabolic processes* (P<0.001) relative to the *M. galloprovincialis* assembly as the reference gene set (Table 2-3). A larger diversity of GO terms were enriched among genes with significant and positive *alpha*/*DoS* values; in addition to *sulfur compound metabolic processes*, significantly enriched GO categories included terms involved in RNA splicing, RNA-protein binding, and liver development (Table 2-3). Results of enrichment analyses using a smaller dataset of 2719 transcripts analysed in this study as the reference gene set are shown in (Table S6).

Discussion

The results of this study highlight how knowledge from physiological studies can be leveraged to inform genomic-enabled investigations of molecular evolution and environmental adaptation. I demonstrate several lines of evidence that divergent selection has disproportionately affected loci functionally linked to temperature stress expression responses in the warm-tolerant and invasive *M. galloprovincialis* lineage. Consistent with my predictions, genomic functions under positive selection are associated with the strongest species-specific responses to thermal stress in physiological studies, supporting the hypothesis that functions linked to temperature adaptation and invasive success at the whole organism level show correlated evolutionary divergence in warm- and cold-adapted congeners. Specifically, the strongest signatures of positive selection fell within a small set of loci involved in protecting the cell from oxidative stress. Furthermore, I find that specific thermotolerance candidate genes under positive selection in independent selection tests share close functional similarities with loci exhibiting the strongest evidence for positive selection in my transcriptome survey (Table 2-2). These findings implicate a contribution of temperature-dependent selection in the divergence between warm- and cold-adapted *Mytilus* species and provide insight regarding the genomic variation unique to *M galloprovincialis* that may also underlie interspecific differences in the evolutionary propensities for invasiveness.

Temperature-related loci show the strongest signatures of positive selection

Positively selected loci in the *M. galloprovincialis* lineage included those involved in immunity, reproduction and cellular processes relating to transport, signalling and protein modification. Such gene categories frequently evolve under positive selection in other taxa (Nielsen et al. 2005; Kosiol et al. 2008). Previous comparative genome scans have also suggested high relative rates of evolution in genes involved in co-evolutionary arms races driven by sexual conflict or host-pathogen defense (e.g., Roux et al. 2014b). I therefore reasoned that in the absence of a predominant contribution of thermal selection in the divergence of *M. galloprovincialis*, genes with the strongest signatures of positive selection would fall within such genomic functional categories. A striking result here, however, is the greater weight of evidence for selection acting on loci associated with thermal and oxidative stress responses. Of the six loci with both the strongest signatures of positive selection (i.e. genes with at least one codon with a high Bayesian posterior probability of falling within the positively selected site class ($\omega_2 >> 1$) and significantly positive *alpha/DoS* values indicating gene-wide adaptive divergence), and demonstrable homologies to annotated genes, five candidates have documented functions in temperature tolerance (Figure 2-2; Table 2-2). Importantly, these selective

targets fall within genomic functional categories differentiating species-specific responses to heat stress in both expression and proteomic comparative studies, supporting a contribution of thermal selection in the divergence of *M. galloprovincialis* from cold-adapted congeners.

These data highlight selection in genes encoding core elements of biochemical adaptation to temperature: proteins involved in controlling oxidative stress directly (e.g., glutathione S-transferases) and indirectly by regulating antioxidant osmolytes in the cellular milieu (e.g., sulfur amino acid proteases) and contributing to cytoskeletal stabilisation (e.g., tubulin binding proteins) (Table 2-2). Most notably, I identified consistent signatures of positive selection in a homolog of the cysteine protease CSAD, including a single positively selected codon conferring a functional group change from a polar and uncharged threonine to a non-polar and hydrophobic valine residue in *M. galloprovincialis* (Table 2-2; Figure 2-3). CSAD is a rate-limiting enzyme for taurine biosynthesis through targeted breakdown of sulfur amino acids (de la Rosa and Stipanuk 1985). Taurine is a major free organic acid in the tissues of osmoconforming marine invertebrates and there is evidence that its primary functions are in regulating osmotic desiccation, membrane stabilisation, and oxidative defense (Silva and Wright 1992; Schaffer et al. 2010; Yancey and Siebenaller 2015). Cellular levels of taurine have been shown to reduce oxidative damage by enhancing the electron transport system and preventing the generation of reactive oxygen species (ROS) by the mitochondria (Jong et al. 2012). Interestingly, recent physiological work in *M. californianus* has demonstrated significant and positive correlations of taurine levels in gill tissues with exposure to elevated temperatures, pointing to thermoprotective properties of taurine and its regulation in response to heat stress under natural conditions (Gleason et al. 2017). However, hypotheses regarding whether warm- and cold-tolerant *Mytilus* species vary in CSAD enzyme activity and taurine turnover in various tissues under thermal stress have not yet been investigated and represent valuable directions for future research.

Several other loci exhibiting the strongest signatures of positive selection are also known to respond to thermal and oxidative stress in *Mytilus* species (e.g., TNF, CPT1A; Table 2-2). Another significant finding is positive selection on the oxidative stress enzyme MGST3, at which I identified a single positively selected codon with three nucleotide substitutions encoding a nonpolar and hydrophobic tryptophan in *M. galloprovincialis* compared to polar glutamine acid residues in the other *Mytilus* species (Table 2-2; Figure 2-3). MGST3 is a glutathione-S transferase (GST) enzyme that catalyses the binding of cellular glutathione to cysteine residues of proteins for protection against oxidative stress (Abele and Puntarulo 2004; Dalle-Donne et al. 2009). GSTs also show elevated rates of sequence evolution in *Acropora* corals (Voolstra et al. 2011) and have been identified as candidates for population-level differentiation under divergent thermal selection (Bay and Palumbi 2014)

pointing to functions mediating thermal tolerance in other marine invertebrates. Taken together, these findings suggest that shifts in thermal environments experienced by warm- and cold-adapted *Mytilus* species are also paralleled by the evolutionary divergence of oxidative stress-related genomic functions. This result is also consistent with the hypothesis that differential strategies in coping with intracellular ROS are a primary factor differentiating thermal tolerance limits between *Mytilus* congeners in physiological studies (Tomanek 2012).

Under natural conditions, sessile intertidal organisms experience daily tidal cycles that expose individuals to surface seawater temperatures and aerial emergence leading to prolonged thermal stress and the accumulation of ROS generated through cellular metabolism (Abele and Puntarulo 2004; Tomanek 2015). Increases in body temperature and heat-induced oxidative damage are therefore common physiological stressors experienced by osmoconforming organisms in intertidal habitats (Tomanek 2015). Alternative evolutionary pressures, however, not involving thermal selection, could lead to observed signatures of accelerated divergence in *M. galloprovincialis*. Oxidative stress related genes can also function in general cellular stress responses, and thus, may evolve in response to various selective environments (Tomanek 2012). Exposure to hyposaline stress, marine pollutants, ocean acidification, pathogens, as well as increased rates of food consumption, have been shown to induce the production of ROS, and may leave signals of selection on genes involved in oxidative stress (e.g., Lockwood and Somero 2011b; Dowd et al. 2013, Tomanek et al. 2012; Tomanek 2015; Feis et al. 2018). Nevertheless, observed signatures of positive selection on oxidative stress-related genomic functions provide corroborative evidence that selective pressures acting on ROS-scavenging cellular pathways are primary feature differentiating *M. galloprovincialis* from cold-adapted congeners (Gracey et al. 2008; Tomanek 2012; 2015, also see Somero et al. 2017).

Functional concordance between positively selected thermotolerance candidate genes and temperature-relevant loci in M. galloprovincialis

Some of the thermotolerance candidate genes known to mediate temperature adaptation from physiological studies also show evidence for accelerated evolution in *M. galloprovincialis* (Table 2-2). Although these loci were previously linked to temperature tolerance through heat-induced expression responses (with the exception of *cMDH*), I find three lines of evidence that positively selected thermotolerance candidates share close functional similarities with temperature-relevant loci exhibiting the strongest signatures of positive selection (i.e. those discussed in the previous subsection). First, proteins involved in cytoskeletal reorganisation (i.e. SHTN1, MDM1, small HSP25; Haslbeck et al. 2005; Tomanek 2012) were among the major functional categories

significantly affected by temperature and heat-induced oxidative stress in expression and proteomic studies in *Mytilus* species (Lockwood et al. 2010; Tomanek 2012), and are also under positive selection in this study (Table 2-2). Second, thermotolerance candidate *GSTO1* and *MGST3* both encode enzymes controlling oxidative stress through the metabolism of cellular glutathione, the most abundant free thiol responsible for redox regulation of intracellular ROS (Table 2-2; Dalle-Donne et al. 2009; Tomanek 2015). Third, functional similarities are evident between cysteine proteases *CSAD* and the thermotolerance candidate *CTSC;* both enzymes regulate organic osmolytes and ROS in the intracellular milieu through targeted breakdown of sulfur-containing amino acids (Table 2-2; McGuire et al. 1992; Connor and Gracey 2011; Lockwood et al. 2015).

Taken together, I propose that accelerated divergence among functionally analogous, but non-homologous proteins, implicates a common selective factor acting on oxidative stress-related genomic functions in *M. galloprovincialis* that is consistent with strong selective pressures on ROS-scavenging proteins as a major component of thermal adaptation among marine invertebrates (Tomanek 2015). Specifically, sulfur catabolism was the top biological process under selection for species divergence, and was significantly enriched among positively selected gene datasets in both codon model tests and polymorphism and divergence-based analyses (Table 2-3). I therefore posit that genes involved with oxidative defense though the targeted metabolism of sulfur-containing thiol groups (e.g., glutathione, cysteine) are important molecular candidates differentiating warm-adapted and cold-adapted congeners. Genes involved in sulfur metabolism have also been shown to diverge under positive selection between ecologically divergent sea urchins occupying shallow and deep water habitats (*Allocentrotus sp.,* Oliver et al. 2010), suggesting that accelerated evolution of genes regulating sulfur metabolism may be a common evolutionary outcome associated with temperature-related ecological shifts experienced by closely related taxa.

More broadly, accelerated divergence in specific thermotolerance candidate genes identified from expression studies suggests that differential transcriptional regulation of key genomic functions between warm- and cold-adapted species may be paralleled by sequence evolution across longer evolutionary timescales (Koester et al. 2013). Significant correlations between gene expression divergence and coding sequence divergence and evolutionary rate have been identified in a number of terrestrial taxa (e.g., *Drosophila*, Nuzhdin et al. 2004; Lemos et al. 2005; Holloway et al. 2007; *Solenopsis* fire ants, Hunt et al. 2012; conifers, Hodgins et al. 2016). However, inconsistent support for this relationship (e.g., yeast, Tirosh and Barkai 2008; sunflowers, Moyers and Rieseberg 2013) indicates that expression divergence may not always predict sequence evolution. For example, I did not find evidence for positive selection among thermotolerance candidate genes with differential

interspecific expression responses in the reference dataset (Figure 2-2). Nevertheless, striking functional concordance among positively selected sets of loci in the present study illustrates how independent lines of experimental evidence from expression studies of whole organism adaptation can inform evolutionary hypotheses regarding genomic divergence between closely related species adapted to contrasting environmental niches.

Combining codon models with polymorphism and divergence analyses

Why did I not always identify congruent signatures of gene-specific positive selection in both codon model tests and polymorphism and divergence analyses? Observed discrepancies among analytical approaches may reflect inherent differences in the methods used in this study. Global McDonald-Kreitman-based tests of adaptive divergence (i.e. *alpha*/*DoS*) rely on summed counts of polymorphism and divergence substitutions within individual genes (Bierne and Eyre-Walker 2004; Stoletzki and Eyre-Walker 2011). As a result, the power of such methods is limited to gene-wide signals of selection operating over long evolutionary timescales. Hence, such methods may underestimate adaptive fixation if current polymorphism is maintained by polygenic selection or soft selective sweeps (Storz and Wheat 2010; Messer and Petrov 2013). Furthermore, because signatures of selection cannot be specified to individual codons, methods based on pooled polymorphism statistics are likely to miss selective targets if adaptive evolution is constrained to small gene regions, such as protein binding domains in long conserved proteins (e.g., Popovic et al. 2014; Hart et al. 2014).

In contrast, likelihood based branch-site tests detect variation in positive selection among individual amino acids and lineages in a phylogeny (e.g., Anisimova and Yang 2007). The small number of taxa analyzed in this study, however, will have limited the power of the branch-site model to estimate gene-wide ω variation and positive selection in a single focal lineage (Anisimova et al. 2001; Stoletzki and Eyre-Walker 2011; Lu and Guindon 2014). The sensitivity of such gene-tree based methods may be further diluted among closely related *Mytilus* species in which shared ancestral polymorphism is predicted to be high, synonymous substitutions to be low (Stoletzki and Eyre-Walker 2011) and interspecific introgression common (Fraïsse et al. 2015). On the other hand, the branch-site model assumes that interspecific sequences harbour only fixed differences; lineage-specific estimates of ω are therefore difficult to interpret for loci in which the rate of shared polymorphism is high and reciprocally monophyletic gene trees are predicted to be rare. The inclusion of population-level polymorphism data provides one way to validate these assumptions; however, small sample sizes coupled with missing individual genotypes and the loss of variants that did not meet filtering

thresholds (e.g., rare variants or complex structural polymorphisms), provides conservative estimates of the extent of shared variation segregating within *M. galloprovincialis*. Nevertheless, congruence among selection tests in this study for focal loci illustrates the power of combining these complementary approaches for (i) distinguishing between positive selection and slightly deleterious amino acid substitutions leading to high ω variation; and (ii) identifying genes that are among the most likely candidates for adaptive species divergence, rather than artefacts of neutral evolutionary processes or methodological biases.

Genetics of species divergence and implications for thermal adaptation

The results of this Chapter highlight evidence that genomic variation differentiating *M. galloprovincialis* from cold-adapted congeners is largely limited to the divergence of loci mediating responses to temperature-related stressors, primarily oxidative stress. While we cannot determine the past selective pressures that a species has experienced, identifying genes and codons as targets of positive selection provides a step forward in delineating lineage-specific adaptations under divergent environmental selection (Storz and Wheat 2010). To date, cMDH and IDH are the only enzymes for which there is genetic and biochemical evidence for functional variation affecting thermal tolerance limits in *Mytilus* species (Fields et al. 2006; Lockwood and Somero 2012). The candidate genes identified in this study build on existing knowledge and predictions regarding which genomic functions that may have predisposed *M. galloprovincialis* to warm temperature adaptation, and thus perhaps, an enhanced competitive ability as an invasive species (Lockwood and Somero 2011a). Determining the adaptive significance of candidate proteins and potential contributions to biological invasions, however, will require functional validation and comparative analyses of thermal physiology between species and populations before strong conclusions can be made (Dean and Thornton 2007). Sequence data can be used in combination with computational approaches to test hypotheses about how specific amino acid substitutions affect three-dimensional protein structure and thermal stability, or whether these sites comprise binding regions undergoing large conformational changes during catalysis (e.g., Fields et al. 2012; Fields et al. 2015; Liao et al. 2017; Saarman et al. 2017). Such data will have important implications for understanding the evolution of invasive characteristics that may be favoured under warming ocean conditions (Tepolt 2015). Identifying genetic changes associated with invasive lineages and providing links to phenotypes involved in their establishment is of particular importance for marine invasive species. Several lines of evidence suggest that the spread of marine invaders are likely to be accelerated by warming ocean temperatures (Stachowicz et al. 2002; Sorte et al. 2010a; 2013; Tepolt and Palumbi 2015), and that colonisation rates are more rapid in the marine environment compared to terrestrial systems (Sorte et al. 2010b).

As genomic tools expand our understanding of a variety of non-model organisms, including marine invasive species, we can gain better perspectives on whether the past environments under which species evolved may make them prone to ecological and physiological tolerance and an evolutionary propensity to become successful invaders.

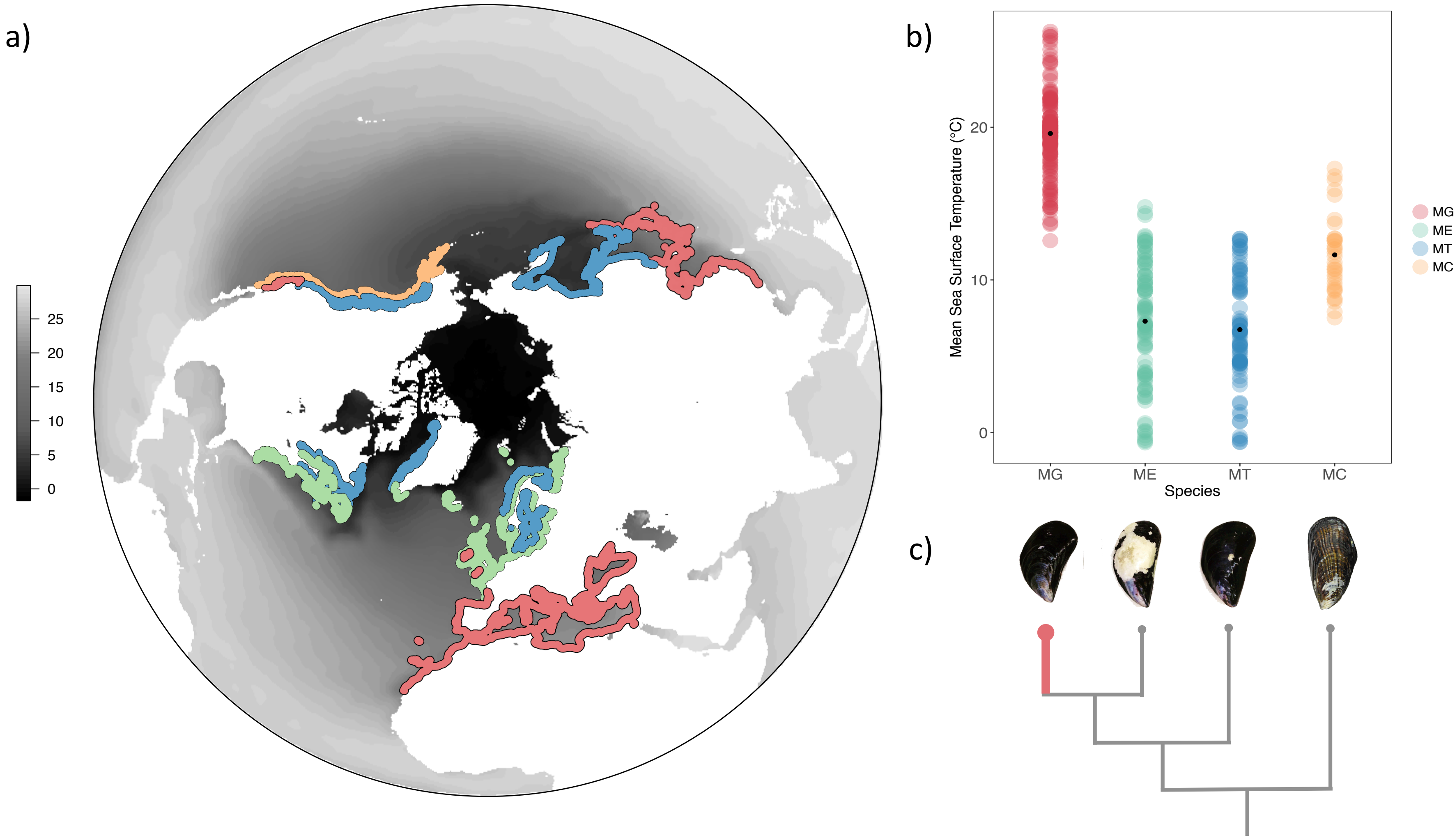
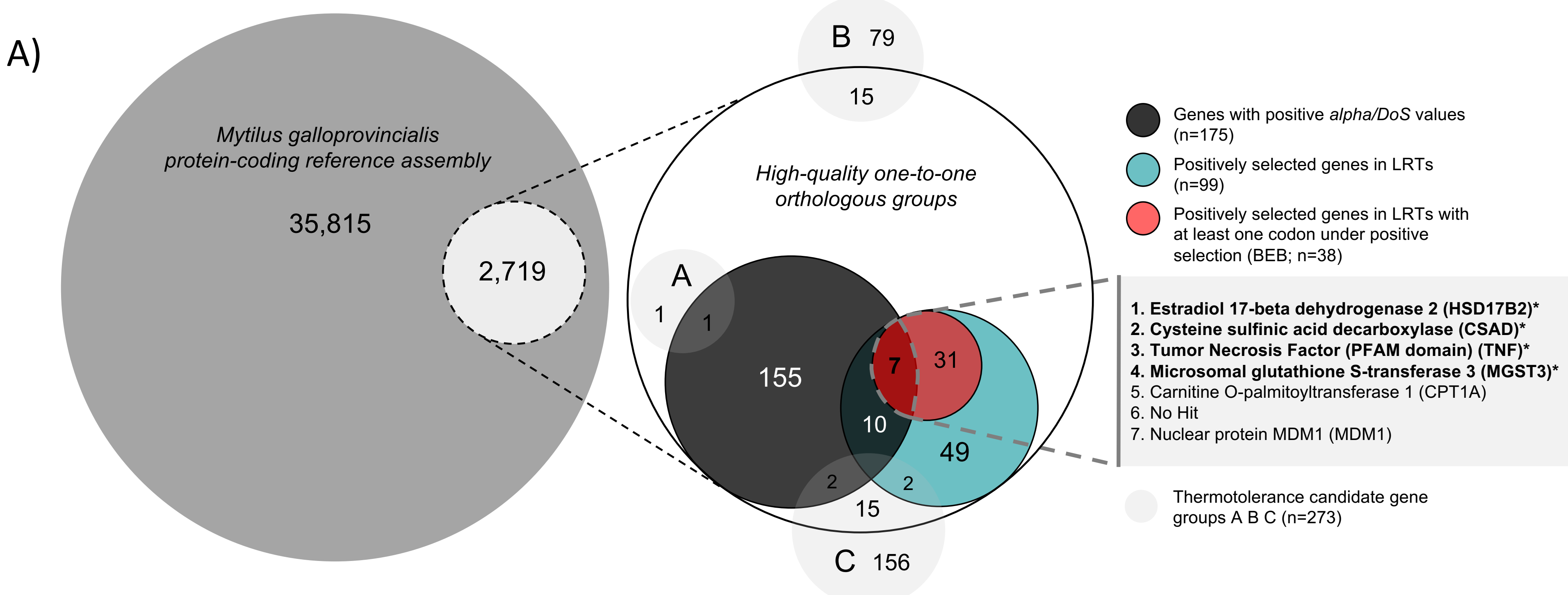


Figure 2-1. Geographical and thermal distributions of *M. galloprovincialis* (MG), *M. edulis* (ME), *M. trossulus* (MT) and *M. californianus* (MC). **a)** Present day distributions in the northern hemisphere based on published genetic literature; **b)** Mean Sea Surface Temperature (SST) data corresponding to species range; **c)** Cladogram illustrating taxonomic relationships. Branch-site model hypothesis test of high relative rates of amino acid substitution along the MG foreground lineage is indicated in red.


A)
Mytilus galloprovincialis
protein-coding reference assembly
35,815
2,719
High-quality one-to-one
orthologous groups
B 79
15
A
1
1
155
7
31
10
49
2
2
15
C 156
Genes with positive alpha/DoS values
(n=175)
Positively selected genes in LRTs
(n=99)
Positively selected genes in LRTs with
at least one codon under positive
selection (BEB; n=38)
1. Estradiol 17-beta dehydrogenase 2 (HSD17B2)*
2. Cysteine sulfinic acid decarboxylase (CSAD)*
3. Tumor Necrosis Factor (PFAM domain) (TNF)*
4. Microsomal glutathione S-transferase 3 (MGST3)*
5. Carnitine O-palmitoyltransferase 1 (CPT1A)
6. No Hit
7. Nuclear protein MDM1 (MDM1)
Thermotolerance candidate gene
groups A B C (n=273)

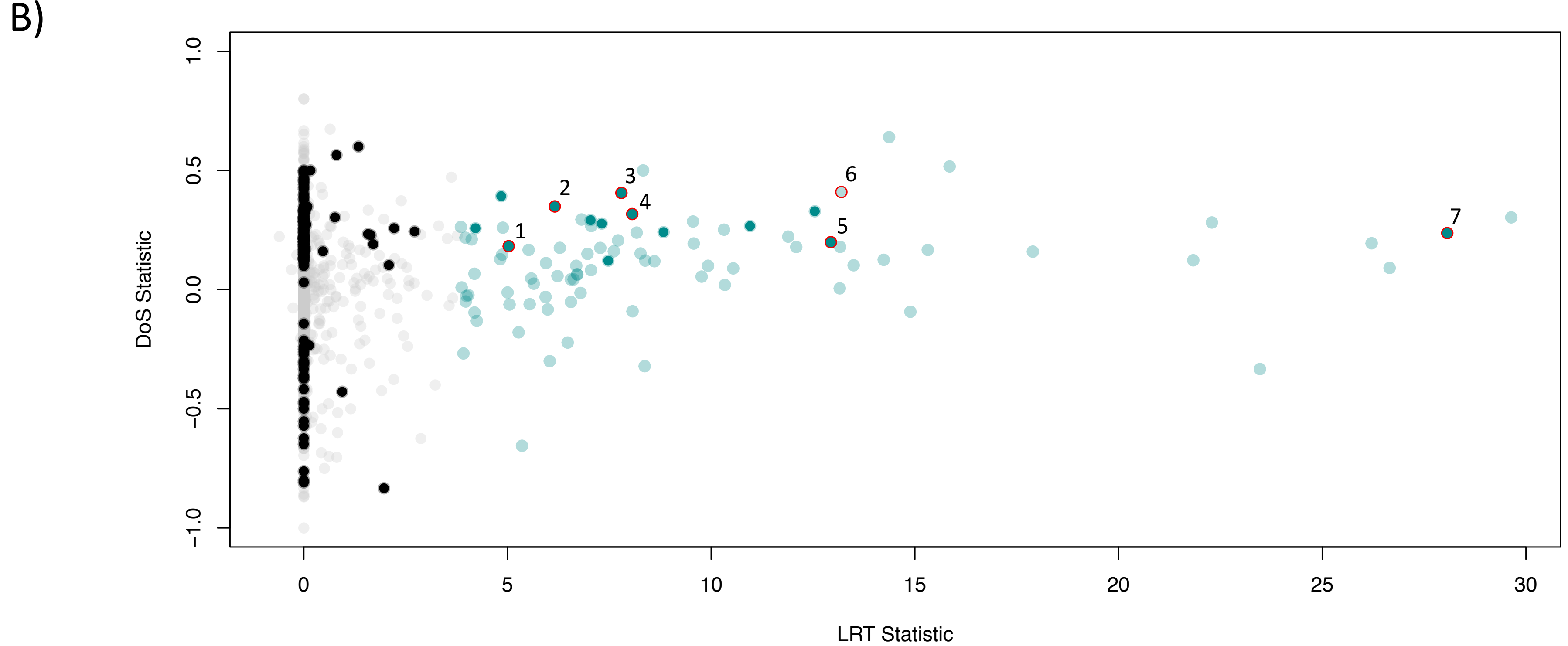


Figure 2-2. Summary of selection test results. **a)** Venn diagram indicates total counts of *de novo* assembled protein-coding transcripts; orthologous groups, known thermotolerance candidate genes, and positively selected gene sets are indicated with values in each corresponding circle. Thermotolerance candidate genes groups correspond to known candidates for (A) divergent genetic adaptation, (B) species-specific differential expression, and (C) shared transcriptional responses among *Mytilus* congeners under acute heat stress. Seven genes showing concordant evidence of positive selection in codon model tests (i.e. Likelihood ratio test (LRT), $P \leq 0.05$; Bayesian Empirical Bayes (BEB), $P \geq 0.95$) and polymorphism and divergence statistics are indicated in the grey box. Four positively selected genes with the 'highest-confidence alignments' are marked with an **asterisk;** **b)** Strength of selection profiles for orthologous gene groups, showing the relationship between per-gene LRT statistic and *DoS* statistic calculated for the Mediterranean *M. galloprovincialis* population (refer to Figure S2-2 for Atlantic population *DoS* values). Significant LRT values are indicated in cyan. Significant *DoS* values are solid circles. Seven genes showing the strongest signatures of positive selection are outlined in red.

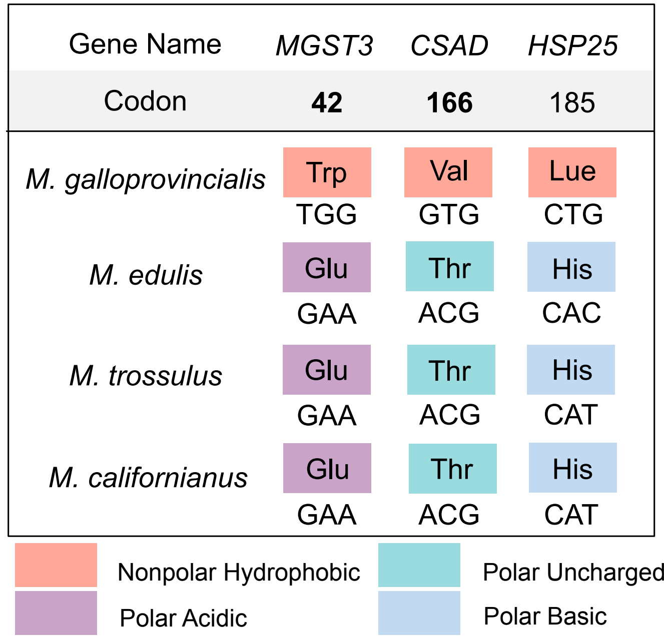


Figure 2-3. Amino acid substitutions and their functional group attributes in selected genes with evidence of positive selection in *M. galloprovincialis*. Significantly positively selected codons with a Bayesian posterior probability P≥0.95 are indicated in **bold.**

Table 2-1. Details of sample collection locations. Populations marked with an asterisk* were also used for population-level analyses.

Analysis	Taxon	Location	Population	Samples sequenced
Species-level analyses				
	Mytilus californianus	Scripps Institution of Oceanography, CA, USA	-	3
	Mytilus trossulus	Light House Park, Vancouver, BC, Canada	-	3
	Mytilus edulis	Darling Marine Station, Maine, USA	-	3
	Mytilus galloprovincialis	Herceg Novi, Montenegro (Eastern Mediterranean)	-	3*
Population-level analyses				
	Mytilus galloprovincialis	Herceg Novi, Montenegro (Eastern Mediterranean)	Mediterranean	5
		Crique des Issambles, France (Western Mediterranean)	Mediterranean	5
		Primel, France (Atlantic)	Atlantic	5

Table 2-2. Summary of loci with evidence of positive selection: **a)** Seven loci showing the strongest signatures of positive selection evidenced by significant Likelihood Ratio Test (LRT) statistics (P≤0.05) and at least one codon with a Bayesian Empirical Bayes (BEB) posterior probability ≥95%; and significantly positive *alpha* and *DoS* values indicating gene-wide adaptive divergence in the Atlantic (A) or Mediterranean (M) population relative to a *M. trossulus* outgroup (P≤0.05); **b)** Positively selected known thermal tolerance candidate genes from physiological studies. Genes with the 'highest-confidence alignments' are in **bold,** three of which have documented roles in temperature tolerance. Significant p-values are marked with an **asterisk.**

Gene Annotation	Abbreviation	Branch-site Tests		Polymorphism-Divergence Tests		Functional Description [ab]; Uniprot ID	Functional Category
		LRT (ω)	BEB (codon site)	*alpha*	*DoS*		
a) Temperature-Related Loci under Positive Selection							
No Hit	*NA*	**13.20 (15)***	0.917 (8) 0.946 (14) 0.922 (18) **0.951 (53)***	**1.000 A*** 0.853 M	**0.569 A*** 0.409 M	NA	NA
Estradiol 17-beta dehydrogenase 2	*HSD17B2*	**5.030 (4)***	0.911 (129) **0.956 (191)***	**0.577 A*** **0.522 M***	**0.211 A*** **0.182 M***	Estrogen biosynthetic process [a]; P37059	Reproduction
Microsomal glutathione S-transferase 3	*MGST3*	**8.062 (999)***	**0.979 (42)***	1.00 A **1.00 M***	0.317 A **0.317 M***	Glutathione peroxidase activity [a]; Q3T100	Oxidative Stress
Nuclear protein MDM1	*MDM1*	**28.07 (998)***	**0.973 (7)*** **0.972 (8)*** **0.990 (9)*** **0.976 (19)***	0.723 A **1.00 M***	0.155 A **0.237 M***	Microtubule binding [a]; Q9D067	Cytoskeletal Reorganisation
Cysteine sulfinic acid decarboxylase	*CSAD*	**6.160 (26)***	**0.951 (166)***	**0.912 A*** **1.00 M***	**0.305 A*** **0.349 M***	Taurine biosynthesis, sulfur amino acid catabolism [a]; Q64611	Oxidative Stress
Tumour necrosis factor (PFAM protein domain)	*TNF*	**7.800 (999)***	0.917 (7) **0.996 (39)***	**0.880 A*** **0.859 M***	**0.425 A*** **0.406 M***	Heat-stress response [b]; PF00229.17 [c]	Stress Response
Carnitine O-palmitoyltransferase 1	*CPT1A*	**12.93 (94)***	**0.999 (703)***	0.775 A **0.858 M***	0.166 A **0.199 M***	Carnitine metabolism, fatty acid beta-oxidation [a], heat-stress response [b]; P32198	Lipid Metabolism

b) Thermotolerance Candidate Genes							
Glutathione S-transferase omega-1	*GSTO1*	0 (1)	-	**0.844 A*** 0.786 M	**0.433 A*** 0.367 M	Glutathione oxidoreductase activity [a], heat-stress response [b] ; P78417	Oxidative Stress
Dipeptidyl peptidase (cathepsin C)	*CTSC*	0 (1)	-	**0.739 A*** **0.719 M***	**0.229 A*** **0.223 M***	Sulfur amino acid catabolism [a], heat-stress response [b] ; P53634	Oxidative Stress
Cytosolic malate dehydrogenase	*cMDH*	0 (1)	-	**1.00 A*** **1.00 M***	**0.176 A*** **0.200 M***	Oxidoreductase activity, gluconeogenesis [a]; heat-stress response [b] ; Q3T145	Energy Metabolism
Small heat shock protein 25	*HSP25*	**7.601 (999)***	0.905 (185)	1.00 A 0.643 M	0.250 A 0.161 M	Heat-stress response [a b]; Q17849	Molecular Chaperone
Shootin-1	*SHTN1*	**4.821 (72)***	None Identified	0.607 A 0.520 M	0.155 A 0.127 M	Actin filament binding [a], heat-stress response [b]; A0MZ66	Cytoskeletal Reorganisation

[a] Functional description based on Gene Ontology Consortium classification

[b] Functional description based on *Mytilus* ecophysiological studies

[c] Functional description based on PFAM protein domain classification

Table 2-3. Enrichment of Gene Ontology (GO) terms among positively selected gene sets ($P \leq 0.001$). Gene sets are delineated by groups with (1a) significant Likelihood Ratio Test (LRT) statistics; (1b) at least one codon with a high Bayesian Empirical Bayes (BEB) posterior probability of falling within the positively selected site class with $\omega_2 >> 1$ ($P \geq 0.95$); (1c) genes with the 'highest-confidence alignments' at putatively selected sites; and (2) significantly positive *alpha* and *DoS* values relative to a *M. trossulus* outgroup ($P \leq 0.05$). Analyses were conducted against the complete *M. galloprovincialis* protein-coding assembly as the reference gene set.

Analysis of Selected Gene List	Gene Count	GO ID	Biological Process Term	P-value	Number of Significant Annotated Genes
1. Branch-site tests (PAML)					
1a. LRT p≤0.05	99	GO:0006790	*sulfur compound metabolic process*	0.00076	1*
		GO:1901565	*organonitrogen compound catabolic process*	0.00076	1*
1b. LRT p≤0.05; BEB p≥0.95	38	GO:0006790	*sulfur compound metabolic process*	0.00051	1*
		GO:1901565	*organonitrogen compound catabolic process*	0.00051	1*
		GO:0016054	*organic acid catabolic process*	0.00081	1*
1c. LRT p≤0.05; BEB p≥0.95 *'highest-confidence alignment'*	19	GO:0006790	*sulfur compound metabolic process*	0.00025	1*
		GO:1901565	*organonitrogen compound catabolic process*	0.00025	1*
		GO:0016054	*organic acid catabolic process*	0.00041	1*
2. Positive *alpha* and *DoS*					
		GO:0000375	*RNA splicing, via transesterification reaction*	1.6e-06	4
	175	GO:0006790	*sulfur compound metabolic process*	1.8e-06	2*
		GO:0022618	*ribonucleoprotein complex assembly*	9.3e-06	3
		GO:0001889	*liver development*	3.4e-05	2

* Cysteine sulfinic acid decarboxylase (CSAD) is annotated with significantly enriched GO terms

Chapter 3. Resolving the cryptic origins of Australia’s native blue mussel despite high rates of admixture with twin introductions of independent non-native mussel lineages.

Abstract

Introduced species can impose profound impacts to the evolution of the receiving communities with which they interact. If native and introduced taxa remain reproductively semi-isolated, human-mediated secondary contact may promote genetic exchange across hybrid zones, potentially eroding native genetic diversity or enhancing invasive species spread. In this Chapter, I investigate the genetic outcomes of past and ongoing (post-introduction) gene flow between the invasive marine mussel, *Mytilus galloprovincialis* and a morphologically indistinguishable and taxonomically contentious native Australian taxon, *Mytilus planulatus*. Population genomic analyses of transcriptome-wide species relationships and approximate Bayesian computations of demographic history validated that Tasmanian mussels are representative of the endemic taxon but share a strong genetic affinity to northern *M. galloprovincialis*. Genomic analyses provided strong evidence that two recent introductions of *M. galloprovincialis* into southeastern Australia are derived from genetically divergent Mediterranean and Atlantic lineages, and that both introductions are associated with high rates of admixture between introduced and endemic populations. Demographic inferences indicated recent divergence times and low levels of historical gene flow between northern *M. galloprovincialis* and the Tasmanian endemic lineage, suggesting historical separation between endemic and introduced taxa of at least 100,000 years prior to present day contact. The results of this study build upon previous genetic studies investigating *M. galloprovincialis* introductions and its interactions with endemic southern hemisphere lineages. This Chapter also demonstrates the utility of genomic data for detangling contemporary invasive introgression from signatures left by historical gene flow and recent divergence histories in native and introduced marine taxa when species boundaries are not well-defined.

Introduction

The ability of introduced species to alter the ecology and evolution of the receiving communities with which they interact is a fundamental issue for understanding the long-term impacts of biological invasions (Rius et al. 2014a; Colautti and Lau 2015). When introduced species are distinct in morphology, life history or ecology from native residents, studies have documented profound effects on native communities, including direct displacement of native populations through competitive or predatory exclusion (e.g., Branch and Steffani 2004). In many cases, however, introduced populations are only detected following successful establishment. For instance, identification of introduced taxa may be obscured by the existence of closely related but morphologically similar native species (Geller et al. 2010). An additional consideration is that morphologically cryptic native and introduced taxa may display variable levels of reproductive isolation: At one extreme, pre- and post-zygotic reproductive barriers are established during long periods of isolation, prohibiting genetic exchange following human-mediated secondary contact (Wu 2001; Roux et al. 2014a; Harrison and Larson 2016). On the other end of the spectrum, species that have been isolated for shorter periods of time or have experienced historical contact throughout their evolutionary histories may retain genomes semi-permeable to gene flow (Roux et al. 2016). If native and introduced taxa remain reproductively semi-isolated, secondary contact may promote ongoing genetic exchange across hybrid zones, resulting in complex evolutionary outcomes for endemic populations (Ellstrand and Schierenbeck 2000).

In the absence of complete reproductive barriers, native and introduced parental species may form stable hybrid zones maintained by environmental and intrinsic barriers to gene flow (Bierne et al. 2011; Jeffery et al. 2018). Introduced and native species may also dissolve into hybrid swarms, potentially eroding the genetic integrity of endemic populations (Harrison 1993; Blum et al. 2010; Fitzpatrick et al. 2010) or eliminating parental genotypes entirely through introgression swamping (Riley et al. 2003; Arcella et al. 2014; Todesco et al. 2016; Glotzbecker et al. 2016). Despite significant consequences for endemic genetic diversity, however, hybrid invasions are more likely to go undetected if species boundaries between introduced and native taxa are not well-defined (Geller et al. 2010). Indeed, comparative genomic studies have revealed a high occurrence of weakly differentiated and semi-isolated species within the 'grey zone' of the speciation continuum (De Queiroz 2007; i.e. 0.075%-2% molecular divergence; Roux et al. 2016) in both terrestrial and marine systems, highlighting general taxonomic issues pertinent for delineating closely related lineages. For invasive species research, however, the 'grey zone' raises additional significant challenges for

understanding invasive species spread and the consequences of secondary contact for endemic populations.

In the marine environment, an abundance of cryptic and cryptogenic species complexes is a persistent issue for detecting marine invasions and documenting introgression between closely related native and introduced species (Carlton 1996; Geller et al. 2010; Appeltans et al. 2012; Brunetti et al. 2015; Bouchemousse et al. 2016; Pante et al. 2015; Viard et al. 2016). Many marine species exhibit high fecundity and dispersal potential (through planktonic larval stages) that support elevated rates of gene flow and weak patterns of genetic differentiation between populations (Palumbi 1992; Gagnaire et al. 2015). Weak differentiation is also sustained by large effective populations sizes that slow down the effects of genetic drift, such that high levels of ancestral polymorphisms are also common features of many diverging marine taxa (e.g., Gagnaire et al. 2012; Fraïsse et al. 2015). Therefore, a primary issue for marine invasions is that high rates of gene flow quickly erode past genetic differentiation that has accumulated during divergence. Once genome-wide homogeneity is restored, distinguishing recently diverged native and introduced species (and their hybrids) from randomly mating panmictic gene pools becomes a key challenge for delineating introduced marine populations (Waples 1998; Gagnaire et al. 2015; Riginos et al. 2016; Crandall et al. 2018) and the processes involved in marine invasive spread (Rius et al. 2015a; Viard et al. 2016).

Without clear taxonomic boundaries, genomic methods based on genetic differentiation (i.e. F-statistics) may fail to distinguish between recently diverged native and invasive marine taxa or to identify divergent sources of introduced populations (Tepolt 2015; Tepolt and Palumbi 2015; Viard et al. 2015). Additionally, because lineages with large effective populations sizes are not expected to reach genome-wide reciprocal monophyly for long periods of time (>$4N_e$ generations; Kuhner 2009; Marko and Hart 2012), both incomplete lineage sorting and recent (post-introduction) gene flow may lead to shared polymorphisms between semi-isolated species (Marko and Hart 2011; Fontaine et al. 2015). Ongoing introgression in contact zones is therefore difficult to recognise and quantify when native and invading taxa show either weak levels of divergence or genomes shaped by complex speciation histories of historical contact (e.g., Fraïsse et al. 2015). Additionally, because both F_{ST} and linkage disequilibrium-based population clustering approaches assume a mutation-drift equilibrium and a single demographic model (i.e. Wright's island model; Wright 1951), such methods cannot provide explicit tests of migration or various demographic histories underlining patterns of genetic ancestry (Patterson et al. 2012; Pickrell and Pritchard 2012). In turn, neglecting complex demographic scenarios that have shaped the genetic backgrounds of closely related taxa may mislead direct

interpretations of population relationships and the strength of introgression between endemic and introduced marine populations (Rougemont and Bernatchez 2018).

Methods that explicitly model demographic histories across many loci offer powerful approaches for resolving the contributions of ancestral polymorphism and recent introgression to shared variation between species at various stages of speciation (e.g., Fu and Li 1997; Pritchard 1999; Fagundes et al. 2007). Coalescent genealogy samplers, for example, allow explicit inferences of divergence and migration rate parameters between natural populations based on highly likely genealogies given the genetic data (e.g., isolation-with-migration models; Hey and Nielsen 2004; 2007; Kuhner 2009; Marko and Hart 2012; Sousa et al. 2013). Such approaches however, rely on full-likelihood calculations that are computationally unrealistic for large genomic datasets or highly parameterised models representative of complex speciation histories likely experienced by marine populations (Roux et al. 2016). Advances in Bayesian statistical methods, in particular, Approximate Bayesian Computations (ABC) allow tests of alternative divergence models that utilise large genome-wide datasets, but rely on few individuals per population or species for robust inferences of complex divergence histories that avoid full-likelihood computations (Pritchard et al. 1999; Beaumont et al. 2002, 2010; Fagundes et al. 2007; Csilléry et al. 2010; Betorelle et al. 2010; Roux et al. 2013, 2016). Indeed, ABC approaches in biological introductions have been highly informative for reconstructing invasion routes (Estoup and Guillemaud 2010; Lombaert et al. 2011; Barker et al. 2017) and discriminating the contributions of ongoing introgression in contact zones (e.g., Estoup et al. 2001; 2004; Rosenblum et al. 2007; Pascual et al. 2007; Guillemaud et al. 2009; Roux et al. 2013). Coalescent approximations can also strengthen comparative inferences of historical species relationships between weakly differentiated introduced and endemic taxa when taxonomic boundaries are also challenged by invasive introgression.

Smooth-shelled marine mussels in the genus *Mytilus* provide a compelling example of a morphologically cryptic and reproductively semi-isolated group of species that have experienced complex evolutionary histories of past hybridisation and contemporary invasive secondary contact. Despite several well-documented pre- and post-zygotic reproductive isolating mechanisms (e.g., Skibinski et al. 1983; Bierne et al. 2003a,b, 2006), there is strong evidence for semi-permeable barriers to gene flow between several pairs of *Mytilus* congeners; differential genome-wide introgression through historical contact and across present day hybrid zones is supported by population genetic (Rawson and Hilbish 1998; Bierne et al. 2003b; Riginos and Cunningham 2005; Gosset and Bierne 2013; Fraïsse et al. 2015) and coalescent demographic inferences (Fraïsse et al. 2018a; Roux et al. 2014a; Roux et al. 2016). The Mediterranean native *Mytilus galloprovincialis* is

recognised as one of the world's most widespread invasive species and is surprisingly the only *Mytilus* congener known to pose invasion threats globally (McDonald et al. 1991; Lowe et al. 2000). *Mytilus galloprovincialis* has a well-documented history of hybridising with native *Mytilus* congeners where their ranges overlap throughout both its introduced distribution in the northern hemisphere (e.g., Japan, Brannock et al. 2009; California, Rawson et al. 1999; Saarman and Pogson 2015). Intraspecific admixture has also been documented in some parts of its present day native range between genetically divergent lineages from the Mediterranean and the Atlantic coast of Europe (e.g., Fraïsse et al. 2015).

Despite a number of genetic investigations (e.g., McDonald et al. 1991; Daguin and Borsa 2000; Hilbish et al. 2000; Gérard et al. 2008; Gardner et al. 2016; Oyarzún et al. 2015; Larraín et al. 2018; Fraïsse et al. 2018b), relatively less is known about the invasive distribution of *M. galloprovincialis* in the southern hemisphere. The widespread occurrence of northern *M. galloprovincialis* haplotypes along temperate coastlines in Chile, New Zealand and Australia suggests that introduced populations are established in coastal regions (e.g., Westfall and Gardner 2010; Gardner et al. 2016; Larraín et al. 2018). However, the existence of morphologically indistinguishable and taxonomically contentious *Mytilus* lineages endemic to the southern hemisphere has sustained ongoing confusion regarding the invasive status of *M. galloprovincialis* in these regions (Westfall and Gardner 2010; Colgan and Middlefart 2011; Dias et al. 2014; Gardner et al. 2016; Ab Rahim et al. 2016; Larraín et al. 2018; Fraïsse et al. 2018b). In Australia, the occurrence of an endemic taxon is supported by fossil evidence predating European contact (New South Wales, Donner and Jungner 1981; South Australia, Hope et al. 1977; Tasmania, Colhoun et al. 1982; reviewed in McDonald et al. 1991; Hilbish et al. 2000); however resolving the taxonomic affinity of the native Australian species, originally named *Mytilus planulatus* Lamarck 1819, has been complex: Initial genetic studies using size polymorphic nuclear markers suggested high genetic similarity to northern *M. galloprovincialis* and described the native taxon as an endemic southern hemisphere lineage of *M. galloprovincialis* (McDonald et al. 1991; Daguin and Borsa 2000; Borsa et al. 2007, Hilbish et al. 2000). Later phylogenetic comparisons of the mitochondrial marker COI (Gérard et al. 2008), however, dated the origins of southern *Mytilus* to the late Pleistocene approximately 0.84 (0.5-1.3 mya, Gérard et al. 2008) and 1.2 million years ago, implicating deeper historical isolation between northern and southern taxa (Hilbish et al. 2000).

Taxonomic delineation of the endemic taxon has also been confounded by interspecific northern *Mytilus* genotypes at few putatively diagnostic nuclear markers, indicating either possible introgression or close genetic affinities to both northern *M. galloprovincialis* and *M. edulis* (Borsa and Daguin 2000; Borsa et al. 2007; Westfall and Gardner 2010, Ab Rahim et al. 2016). However, high expected levels of incomplete lineage sorting among *Mytilus* species resulting in various gene

topologies (e.g., Fraïsse et al. 2015) are likely to obscure signals of present day introgression and further amplify discordant species relationships when few loci are examined. Given the limitations of traditional approaches, it remains unresolved (i) whether Australian native *Mytilus* comprise a lineage divergent enough from northern *M. galloprovincialis* to warrant species status as *Mytilus planulatus*, and (ii) whether evidence of genetic admixture in Australian mussels reflects a history of recent divergence with or without ongoing hybridisation with introduced congeners. In this study, I use transcriptome-wide data to investigate these two questions: I combine population genomic analyses of species relationships with approximate Bayesian computations to test alternative hypotheses regarding the origins of Australian *Mytilus* mussels (hereafter referred to as its current nomenclature, *Mytilus planulatus*) and the contributions of past and ongoing (post-introduction) gene flow with introduced *M. galloprovincialis* in southeast Australia. This study represents the first transcriptome-wide investigation of demographic history and introgression between introduced and Australian endemic *Mytilus* species. This Chapter also lends towards an improved understanding of marine invasions more generally, including possible genetic consequences for native taxa when introduced species boundaries are not well-defined.

Methods

Sample collection, RNA extraction and sequencing

Mussels were collected from wild populations from the rocky intertidal or subtidal environment (Table 3-1). Outgroup specimens (*Mytilus californianus*, and three blue smooth-shell taxa *Mytilus trossulus*, *M. edulis* and *M. galloprovincialis*) were collected from known contemporary allopatric ranges in order to minimise the possibility of sampling hybrid individuals. Individuals were genotyped for the species diagnostic marker *Glu-5'* (Rawson et al. 1996) to obtain a first clue about species identity. I then performed a principal component analysis (implemented in R; R Development Core Team 2017) of preliminary transcriptome-derived variants as an additional measure to exclude introgressed individuals as outgroup taxa. Preliminary assignment of Australian samples as native *M. planulatus* was based on the F-type (female) mitochondrial marker COIII using primers from Riginos et al. (2004) and phylogenetic topology analyses using neighbor-joining statistics implemented in Geneious 8.1. Total RNA was extracted from 10-20 mg of mantle tissue (preserved in RNAlater) for a total of 47 samples (Table 3-1), which were subsequently checked for quality and sequenced across three lanes of an Illumina Hiseq2000 or across a single lane of an Illumina Hiseq4000.

RNAseq data processing and reference transcriptome assembly

I used Trimmomatic (v0.36) (Bolger et al. 2014) to remove adapter sequences from paired reads. Quality trimming was performed using a 4 bp sliding window, a phred-scale average quality score of 20 and a minimum size filter of 50 bp. I reduced redundancy among high-coverage reads (including rRNA contaminants) and discarded associated sequence errors by digitally normalising each dataset using Trinity's insilico_read_normalization.pl script with a default kmer size of 25 and maximum read coverage of 50 (Grabherr et al. 2011). Overlapping paired reads were merged using FLASH v1.2.11 (Magoč and Salzberg 2011) with a minimum overlap length of 10 bp. Three *M. galloprovincialis* individuals (those with the highest FLASH merging scores: the highest absolute number of reads merged into larger fragments) from three native-range populations were used to make a reference transcriptome assembly intending to capture representative proportions of genetic variation in the *M. galloprovincialis* native range. These samples were used to create three population-specific *de novo* assemblies using Trinity v2.0.6 (Grabherr et al. 2011) with default parameters. The longest isoforms were extracted for each gene group for each assembly. The three reduced Trinity assemblies were then meta-assembled using CAP3 (Huang and Madan 1999) with default parameters into a single high quality reference assembly.

I queried the resulting assembly against the Uniprot-Swissprot protein database using blastx with an e-value threshold of 10^{-3} for significant matches. Subsequently, contigs with significant blast matches to likely environmental contaminants, including bacteria, fungi, viruses, protists (Alveolata), green (Viridiplantae) and red algae (Haptophyceae), and other eukaryotic contaminants (i.e. Euglenozoa) were removed using Biopython v1.68 and R (184,842 contigs). Prior to genomic analyses, transcripts showing high sequence similarity in the *M. galloprovincialis* reference assembly (i.e. likely derived from the same gene) were clustered using Cd-Hit-Est (Li and Godzik 2006; Fu et al. 2012) with a minimum sequence identity threshold of 95% of the shortest sequence. Finally, I removed transcripts with significant blastn matches (e-value 10^{-3}) to the *M. galloprovincialis* male (Genbank reference: FJ890850.1) and female (Genbank reference: FJ890849.1) mitochondrial genomes. The resulting 159,985 nuclear sequences were used as a reference assembly for variant discovery and as input for all downstream analyses.

Genomic data filtering

RNAseq reads from Australian samples and four outgroup species (*M. californianus*, *M. edulis*, *M. trossulus* and *M. galloprovincialis*) were mapped to the reference assembly using Bowtie2 and the

very sensitive local method (Table 3-1; Langmead and Salzberg 2012). PCR duplicates were marked and removed using Picard MarkDuplicates (http://picard.sourceforage.net) and Samtools was used to convert SAM files to indexed BAM files. Single nucleotide polymorphisms (SNPs) were called using Freebayes (https://github.com/ekg/freebayes), and variant filtering was performed in VCFtools (Danecek et al. 2011). Variant sites below a minimum genotype quality threshold of 30 and a minimum mean depth coverage of 10 reads were excluded. For the purpose of phylogenomic analyses (genomic network analysis and topology weighting), I removed singletons, indel variants and all positions with missing data. Genotypes were statistically phased using the program beagle v4.1 (Browning and Browning 2007). I generated a consensus sequence for each individual haplotype using the corresponding VCF file and reference sequences in BCFtools v1.3.1. For analyses investigating population relationships and admixture (ADMIXTURE and TreeMix), I included all Australian (n=23) and *M. galloprovincialis* individuals (n=15). I additionally removed SNPs with a minor allele frequency of less than 5%, but retained positions with up to 20% missing data.

Analyses of population structure and admixture

I performed principal components analysis of three Australian populations and northern *M. galloprovincialis* samples from its native range, to ascertain the presence of the putative endemic *M. planulatus* and introduced northern populations. For this analysis, I used only nuclear SNPs and the same filtering thresholds as described for phylogenomic analyses. To explore population structure and the possibility of admixture between putative native and introduced populations, I estimated individual ancestry proportions using the program ADMIXTURE and *M. edulis* as an outgroup taxon (Alexander et al. 2009). I used VCFtools and PLINK v1.90 (Purcell et al. 2007) to convert the filtered VCF output to BED format files as input, which reduced the original dataset to 34,097 biallelic SNPs across 3945 contigs. I ran ADMIXTURE with the 100 iterations and I used the cross-validation procedure with 50 replicates for K=1 to K=10 genetic clusters.

To formally test the hypothesis of genetic exchange between putative endemic and introduced populations and to identify putative sources of gene flow, I used TreeMix v 1.12 (Pickrell and Pritchard 2012) to estimate the strength (branch weight = *w*) and direction of migration between populations while accounting for historical relationships between groups. TreeMix first uses allele frequency correlations between populations to infer a maximum likelihood population tree that best represents the phylogenetic relationships between groups. Migration edges are subsequently added between population branches to determine whether incorporating admixture events of varying weights and directions improve the likelihood of the population tree given the genetic data. For

TreeMix analyses, I used *M. edulis* as an outgroup to focus inferences on recent admixture events. I accounted for linkage disequilibrium by performing the analysis on windows of 100 variants. I calculated the standard error of migration events with the –se option without sample size correction (option –noss). I examined the residual plot of pairwise population genetic covariances to infer the possibility of gene flow between populations, where negative residual standard error values suggest that populations are more closely related to each other (given the data) relative to the maximum likelihood population tree with no migration events. I then modeled 1-10 migration events sequentially to see whether the addition of migration edges to the phylogeny improved the likelihood fit to the data. I used a stepwise comparison Akaike Information Criterion (AIC) between sequential migration models to determine whether the addition of a migration edge significantly improved the likelihood of the population tree relative to a nested model with n-1 migration edges. I calculated AIC values as (-2(ln(likelihood)) + 2K), where K is the number of free parameters in the model (i.e. K=2; the number of migration edges and the direction of gene flow). I did not consider additional migration events when the difference between nested models was less than two ($\Delta AIC < 2$). The tree graph and residual plots for each analysis were visualised using R.

The three population (f_3) test of admixture (Keinan et al. 2007; Reich et al. 2009) was used to verify evidence of migration inferred by TreeMix (as recommended by Pickrell and Pritchard 2012). I estimated the f_3 statistic using the *threepop* function. The f_3 statistic estimates whether differences in allele frequencies between each population combination deviate more than expected due to incomplete lineage sorting, thereby suggesting recent admixture. Significant migration is inferred if the f_3 statistic has a negative value and a z-score of ≤ -3.8 (equivalent to a P-value < 0.0001), which is determined through a jackknifing procedure over windows of 100 SNPs.

Genomic analysis of species relationships

To visualise genomic relationships between three Australian populations against four northern *Mytilus* outgroup taxa I first performed a genomic network analysis of individual haplotype sequences from all sampled populations (Table 3-1) using the neighbor-net method in SplitsTree4 v4.14.6 (Huson and Bryant 2006). Network phylogenies are useful for visualising species relationships best represented by reticulate evolution, rather than bifurcating splits between sister taxa (Huson and Bryant 2006). The phylogenetic network was generated from a concatenated nucleotide sequence (constructed in R) consisting of 27,343 SNPS from 2620 nuclear contigs, using default settings. I also estimated a phylogenetic network based on 144 SNPs from 12 protein-coding female mitochondrial

genes. The distinctive *Mytilus* male mitotypes were not recovered due to low coverage in the transcriptome data.

I then quantified how species relationships between putative endemic *M. planulatus* and northern hemisphere *Mytilus* species may vary across the nuclear genome; I estimated the relative contributions of three possible alternative topologies (i.e. grouping *M. planulatus* with one of three outgroup species: *M. galloprovincialis* (Mediterranean), *M. edulis* or *M. trossulus*) to the nuclear species tree using a heuristic topology weighting analysis in TWISST (topology weighting by iterative sampling of subtrees; Martin and Van Belleghem 2017). TWISST estimates the weight or contribution of each possible unrooted topology for each locus by resampling a single haplotype per taxon to generate all possible subtrees for that particular locus. To minimise unresolved topologies due to present day introgression in downstream analyses, I only included putative non-introgressed *M. planulatus* samples from Tasmania showing no evidence of admixture in initial analyses. I also excluded the most distant outgroup, *M. californianus,* to limit comparisons to three possible topologies. Consensus haplotypes for each locus were analysed individually; I inferred locus-specific genealogies with the R package 'Ape' using the neighbor-joining method and F84 distances. (Felsenstein and Churchill 1996). To exclude poorly resolved phylogenies (e.g., Martin and Van Belleghem 2017), I performed the analysis on a subset of 343 genealogies with a minimum tree length of 0.025, which is equivalent to 5 SNPs every 200 bp.

Approximate Bayesian Computations (ABC) of demographic history

I used an ABC framework to test the hypothesis that *M. planulatus* samples from Tasmania have experienced an independent evolutionary history from northern *M. galloprovincialis*. Specifically, I tested among six alternative models of Australian *Mytilus* origins representing a spectrum of divergence histories between northern and southern taxa (Figure 3-1): panmixia (*pan*), where populations belong to the same gene pool; divergence in isolation (*div*), isolation with migration (*im*), divergence with ancient gene flow (*divAGF*), divergence with recent (invasive) secondary contact (*divSC*), divergence with ancient gene flow and recent secondary contact (*divAGFSC*). All models assumed constant effective population sizes, and that population pairs diverge from an ancestral panmictic population at time T_{div}. Refer to Figure 3-1 for detailed model descriptions.

Empirical genetic data

I reduced that reference assembly to only contigs containing predicted open reading frames (ORFs), using Transdecoder (Haas et al. 2013); I identified protein-coding sequences greater than 100 amino acids and with significant matches to the Pfam protein database. This resulted in 52,364 transcripts with ORFs, including 16,151 complete protein-coding nuclear loci (in which both start and stop codons were detected), which were used as a reference assembly. I mapped individual reads datasets (as described above) against this reduced complete protein-coding assembly and used to resulting BAM files as input for downstream ABC analyses. I conducted two pairwise population comparisons to calculate summary statistics: I compared the genomic backgrounds of *M. planulatus* sampled in Tasmania (putative endemic) with two divergent *M. galloprovincialis* populations from its native range in the Mediterranean and Atlantic. For each dataset, the *reads2snps* program was used to predict individual genotypes based on a probabilistic maximum-likelihood framework. Genotypes below a minimum read depth of 10 and genotype posterior probability of 95% were removed as well as variants resulting from the misalignment of reads to paralogous contigs. Subsequent analyses were conducted on the output of *reads2snps* using custom scripts in R (available at https://github.com/dinmatias). The R scripts implement an existing pipeline available from the PopPhyl project (https://github.com/popgenomics/popPhylABC; Roux et al. 2016). For each pairwise population comparison, I used PolydNdS to identify synonymous variants and segregating sites for each locus. Sequences were then parsed and filtered to retain only synonymous variants and transcripts above a minimum length of 30 synonymous sites. Following these filtering thresholds, I removed monomorphic loci and loci with missing haplotypes for any individuals. The resulting empirical datasets consisted of 1,362 loci for Mediterranean-Tasmania, and 1,539 loci for Atlantic-Tasmania population pairs.

Coalescent simulations of genetic data

I used msnsam, a modified version of the ms coalescent simulator, to generate one million multilocus simulations under each demographic model, for each population pair (Figure 3-1; Ross-Ibarra et al. 2008). Simulations assumed a neutral mutation rate $\mu=2.763 \times 10^{-8}$ per bp per generation, which was scaled by the number of synonymous sites of each locus to obtain per-locus mutation rates. To account for recombination, I followed the recombination rate implemented by Roux et al. 2016 which is equal to 0.5 of the mutation rate. Initial models assumed equal (i.e. homogenous) effective population size (N_e) among loci and homogeneous migration rate, $M=4 N_{ref} m$ every generation, where N_{ref} is the reference effective population size and m is the proportion of each population consisting of new migrants each generation.

Each simulation was parametrised by model-specific demographic parameters randomly drawn from a uniform prior distribution (Table S3-1, Supplementary Material) generated by a modified version of the *priorgen* software (Ross-Ibarra et al. 2008). Effective population size parameters (N_i, N_j, and $N_{ancestral}$) were randomly drawn from a distribution of 1000-500,000 individuals. To inform the priors for time-related parameters, I capitalised on previous divergence estimates for *Mytilus* species translated to number of generations using a generation time of 2 years (e.g., Roux et al. 2014a). Specifically, divergence time (T_{div}) was sampled from the interval 100,000-1,750,000 generations to capture the earliest estimated time of mitochondrial divergence between southern hemisphere taxa and *M. galloprovincialis* between 0.54-1.31 million years ago (Gérard et al. 2008). The upper bound for T_{div} was informed by the estimated splitting time between *M. trossulus* and the ancestor of *M. edulis* and *M. galloprovincialis* (~3.5 million years) that preceded potential periods of transequatorial migration associated with the late Pliocene about 3.1 million years ago (Lindberg 1991; Hilbish et al. 2000). The prior distribution for bidirectional ancient migration (*ma*) between northern and southern hemispheres was sampled between 0-0.0001 (equivalent to 0-200 migrants per generation when N_{ref} =500,000). I sampled the number of generations since ancient migration seized (T_{nc}) bounded by the interval T_{div} -10,000 generations (corresponding to the last glacial maximum). For the invasive migration parameter (*m*) I explored unidirectional gene flow values sampled on a broader interval *m*=0-0.5 into Tasmania. I sampled the onset of human mediated secondary contact (T_{sc}) on the interval 5-300 generations. A standard set of 39 summary statistics (e.g., Roux et al. 2014a; Fraïsse et al. 2014) of divergence and polymorphism were calculated for each simulation and for the empirical genetic data using *Mscalc* (Ross-Ibarra et al. 2008).

Demographic model selection

To evaluate the posterior support for alternative demographic models, I obtained posterior samples from all simulated data by applying thresholds of 0.001 and 0.01. An acceptance threshold of 0.001 is equivalent to 6000 simulations that generated summary statistics falling closest to the observed empirical values (Blum and François 2009). To estimate the posterior probability of each model, I performed a categorical regression on the model indices and summary statistics of the posterior samples using the feed-forward neural network method (Beaumont 2010). Computations were performed with 50 trained neural networks and a maximum of 2000 iterations while weighing each posterior sample by an Epanechnikov kernel with a maximum value when the simulated values are equal to the observed summary statistics. In comparisons where not all six demographic models had accepted values within the applied threshold, the simple rejection method (i.e. linear regression) was

applied (Beaumont et al. 2010). All these procedures were conducted using the packages `abc` (Csilléry et al. 2012) and `nnet` (Ripley et al. 2016) in R.

To validate the ability to discriminate between alternative models, I simulated an additional 1000 pseudo-observed datasets (PODS) from the prior distribution for each demographic model to perform model checking. Using each POD as the new empirical dataset, I estimated the posterior support for each model given the original simulated summary statistics utilising the model selection procedure (as described above). I then examined the rate by which my approach correctly supported the true model of the PODs (i.e. precision) and the rate of by which incorrect models are supported (i.e. misclassification or Type I error). From this validation procedure, I examined the minimum threshold for model probability that will give a robustness of 0.95 (Roux et al. 2016), that is, a 95% probability to correctly support a model given that its posterior probability is higher than the threshold. I then applied this minimum probability to evaluate if the estimated posterior probability for the empirical data is robust.

Accounting for among-locus variation in genetic drift and migration

For initial ABC comparisons, simulated demographic models assumed genome-wide homogenous N_e and m. Accounting for differential rates of introgression and genomic variance in genetic drift that may result as an outcome of linked selection, however, has been shown to significantly improve the accuracy of demographic inferences in *Mytilus* species (Roux et al. 2014a), as well as other marine taxa (*Ciona sp.,* Roux et al. 2013; sea bass, Tine et al. 2014; *Salmo salar*, Rougemont and Bernatchez 2018). To account for the combined effects of differential migration and selection at linked sites leading to variable among-locus rates of genetic drift, I re-simulated a series of nested models incorporating heterogeneous N_e and/or heterogeneous m under the best demographic scenario (inferred from initial homogeneous model comparisons) to estimate demographic parameters. Specifically, I was interested in whether these models provided an improved model probability (and parameter estimates) when compared to the best inferred model with homogeneous N_e and m.

Models including gene flow were subdivided to allow for heterogeneous migration rate among loci assuming selection against migrants (i.e. below the genomic average; m_{ref}). The N_e parameter was drawn independently for each locus for each simulation, and was either homogeneous (homo), heterogeneous below the genomic average (hetero1) or allowed to be drawn above the genomic average (hetero2) to account for linked selection favoring polymorphism (assuming either homogeneous or heterogeneous m). Heterogeneity in N_e and m values was modeled using a random

parameter drawn from a beta distribution as the hyper-prior for each parameter. Beta distributions were shaped by beta parameters sampled on the interval [1-50] for heterogeneous N_e and [1-20] for m, where the proportion of loci evolving under linked selection in each population was sampled from a uniform distribution [0-1]. I simulated 1 x 10^6 simulations for a total of 13 additional heterogeneous models and calculated summary statistics as described above.

Demographic parameter inference

Demographic parameters were estimated for each population pair using the posterior distribution approximated by accepted simulations under the best demographic model. Parameter values were log transformed prior to regression to ensure that the posterior distribution was contained within the prior bounds (e.g., Estoup et al. 2004; Hamilton et al. 2005). I used 50 neural networks and a maximum of 2000 iterations to obtain weighted nonlinear regressions of the parameters to the summary statistics from 1000 accepted simulations closest to the observed values (acceptance tolerance=0.01%). Parameter inference was based on estimated posterior means and 95% credible intervals when parameters were well differentiated from the uniform prior.

Results

Genomic analysis of species relationships

Haplotype genetic networks of Australian populations and four outgroup taxa constructed from 12 mitochondrial genes and 2620 nuclear contigs revealed strong species tree discordance between nuclear and mitochondrial genomes. Individuals carrying the Australian (female) mitotype form a distinct divergent clade (Figure 3-2). In contrast, the same individuals (independent of mitotype) clustered together in a monophyletic clade alongside *M. galloprovincialis* when the nuclear genome is analysed. TWISST analyses of species relationships corroborated that gene trees grouping *M. planulatus* with *M. galloprovincialis* dominated the nuclear genome (Figure S3-2). The mean weighting for topologies placing *M. planulatus* as a sister species to the invasive taxon was 54%, supporting a close relationship between these two species (Figure S3-2). Only 39/343 loci (11%) had fully resolved topologies (topology weight = 1.0), all of which grouped *M. planulatus* with *M. galloprovincialis.* Visual inspection of topologies revealed that all of these loci are paraphyletic (do not form species-specific clades) with *M. galloprovincialis,* suggesting that high levels of ongoing incomplete lineage sorting in Tasmanian samples or past or recent genetic exchange; although I did

not recover evidence supporting recent introgression in this population in any other gene flow analyses. Topologies grouping *M. planulatus* with *M. edulis* and *M. trossulus* showed mean weightings across contigs of 21% and 23%, respectively (Figure S3-2), suggesting a high degree of incomplete lineage sorting of ancient polymorphisms in the *Mytilus* phylogeny.

Evidence for genome-wide introgression with introduced M. galloprovincialis

Principal component analysis of 20,509 SNP markers revealed separation between Australian samples and northern *M. galloprovincialis* populations, explaining 7.31% (PC1) and 6.55% (PC2) of variance among individual genotypes (Figure 3-3A). Populations from Sydney Harbour and Batemans Bay showed intermediate placement between samples from Tasmania and northern *M. galloprovincialis* populations. Divergence of these populations across the second PC axis points to likely introgression with genetically divergent source lineages of *M. galloprovincialis* from the Mediterranean and Atlantic. ADMIXTURE analyses for K=4 clusters discriminated between genetic groups identified in the PCA (Figure 3-3B). Individuals sampled in Batemans Bay and Sydney Harbour showed greater than 50% shared ancestry proportions with northern *M. galloprovincialis*, suggesting at least two independent introductions into Australia of divergent invasive source lineages: Atlantic *M. galloprovincialis* into Batemans Bay, and Mediterranean *M. galloprovincialis* into Sydney Harbour, and subsequent admixture with these native populations. All sets of analyses (for all K clusters) did not provide evidence of mixed ancestry proportions in Tasmania. Admixture proportions were consistent when analyses were performed using only 1 SNP per contig to account for linkage effects, as nearby SNPs are not independent (Figure S3-1, Supplementary Material).

To formally test for introgression and to validate potential sources of gene flow, I performed joint analyses of migration and historical population relationships inferred from TreeMix. The maximum likelihood population tree without migration explained 95.88% of variation in the allele frequency covariance matrix based on 34,097 SNPs (Figure 3-4A). However, I observed high pairwise residual covariance between Sydney Harbour and Batemans Bay with northern populations suggesting introgression between population pairs. I therefore tested whether the addition of 1-10 to migration edges would sequentially improve the likelihood fit of the population tree to the data. The addition of four admixture events explained more than 99% of the genotypic variance and provided the highest likelihood fit (based on stepwise AIC comparisons) to nested models with fewer migration edges (Figure 3-4B); however only two out of the four migration edges (those into Australian populations) accounted for most of the explained genotypic variance. The model with five migration edges was not a significantly better fit to the data (ΔAIC < 2; Table S3-2, Supplementary Material). I also found

significant p-values (P<<0.001) for individual migration edges, indicating that the corresponding direction and weights of the branches improved the likelihood fit to the data, although the direction of the migration edges should not be interpreted as face value (Figure 3-4B).

Results from TreeMix confirmed evidence for introgression from eastern Mediterranean *M. galloprovincialis* (Montenegro) into Sydney Harbour (*w*=33%), and migration from the Atlantic *M. galloprovincialis* population into Batemans Bay (*w*=40%), suggesting contemporary admixture between native and introduced populations. Results did not provide evidence for migration into Tasmanian mussels. The slight signal of admixture between Batemans Bay and the outgroup taxon suggests shared genetic elements are likely as a result of secondary contact with Atlantic *M. galloprovincialis* populations that share ancestry with *M. edulis* through past and ongoing introgression (Fraïsse et al. 2015). The relatively weak strength of this admixture event (*w*=12%) is consistent with small proportions of *M. edulis* ancestry observed among Atlantic *M. galloprovincialis* individuals in clustering analyses (Figure 3-3B; K=3). Similarly, slight evidence of admixture between the eastern Mediterranean and Batemans Bay likely indicates allele sharing with Mediterranean *M. galloprovincialis* through introgression with Atlantic populations prior to introduction (e.g., Fraïsse et al. 2015); however, this migration edge was not strongly supported (P>0.001; Figure 3-4B). Mixed Atlantic-Mediterranean ancestry was also evidenced in ADMIXTURE analyses (Figure 3-3B; K=4), in which three admixed individuals sampled in Batemans Bay also showed small proportions of shared ancestry with Mediterranean *M. galloprovincialis*. Analyses using only a single variant per contig corroborated significant migration edges, in which the addition of three migration events generated the highest likelihood population tree.

The strongest migration events inferred by TreeMix were supported by f_3 statistics in the three-population test. I found significantly negative values ($P < 0.0001$) for almost all population combinations involving either Sydney Harbour or Batemans Bay as the admixed population and *M. planulatus* (Tasmania) and northern *M. galloprovincialis* populations as putative source ancestral populations (Figure 3-4C). Additionally, I detected signatures of *M. edulis* genetic elements into both Australian (e.g., Batemans Bay) and Mediterranean *M. galloprovincialis* (Mediterranean-West) genetic backgrounds. Three-population tests involving Tasmania as an admixed population did not yield significant values for any population combinations, supporting the hypothesis that samples in this region are representative of the endemic taxon.

Historical demographic inference

Levels of endemic genetic diversity and divergence at synonymous sites

The majority of biallelic polymorphic sites were shared between *M. planulatus* and northern *M. galloprovincialis* (average shared polymorphic sites across loci = 5.36-5.83) compared to polymorphic private alleles in *M. planulatus* (average private polymorphic sites = 2.87-4.07). Levels of nucleotide diversity at synonymous sites (averaged across loci) were similar for all populations based on Tajima's pi (Tajima 1983) ranging between 0.023-0.024 for *M. planulatus* sampled in Tasmanian and 0.021 (Mediterranean) and 0.026 (Atlantic) for northern *M. galloprovincialis* populations. Pairwise population comparisons indicated low population differentiation evidenced by few fixed variants (0-0.02 averaged across contigs) and low F_{ST} or absolute (d_{xy}) and net (D_a) divergence values between Tasmania and Mediterranean (mean F_{ST}=0.052; d_{xy}=0.25; D_a=0.003) or Atlantic (mean F_{ST}=0.087; d_{xy}=0.30; D_a=0.005) *M. galloprovincialis*, indicating that population differentiation is largely driven by the presence of private alleles. Departures of the site frequency spectrum measured as mean Tajima's D values (Tajima 1989) varied between populations; I observed negative values close to neutrality in Tasmania (average D= -0.151), and greater negative values in northern *M. galloprovincialis* (Mediterranean D= -0.475; Atlantic= -0.445), indicating an excess of rare alleles due to population expansions, gene flow from unsampled populations or signatures of directional selective processes.

Historical isolation between northern and southern hemispheres

Using an ABC framework, I tested the hypothesis that Tasmanian samples are representative of an endemic taxon by comparing six alternative divergence models with and without past or ongoing migration between *M. planulatus* and *M. galloprovincialis*. I first inferred the best demographic model by comparing models with homogeneous among-locus parameters. For both population pairs, the model of divergence with ancient gene flow (*divAGF*) received the highest posterior support (> 78% posterior probability using neural network inference and acceptance threshold 0.001) when compared to other candidate models assuming homogeneous N_e and m (Table 3-2). The divergence model (*div*) provided the second highest posterior probability given the data in most comparisons. I also observed consistent rejection of the isolation-with-migration (*im*) model suggesting that historical gene flow was followed by a period of divergence in isolation. Indeed, I predicted that this scenario would be unlikely (although possible) given known periods of historical isolation (i.e. deep divergence of mitochondrial genome between northern hemisphere taxa and *M. planulatus*) and

equatorial barriers to migration following throughout the Pleistocene (Vermeij 1992). Model comparisons returned weak support for models including very recent secondary contact associated with invasive introductions <600 years ago and no support for demographic model of panmixia.

Model choice validation indicated that I could discriminate between the best inferred model (*divAGF*) model and models including recent genetic exchange (*im, divSC, divAGFSC*). The *divAGF* model had the highest posterior model probability in 59% (Atlantic) and 61% (Mediterranean) of model comparisons using PODs generated under the same model (i.e. precision; Figure S3-3 Table S3-3). However, I found that 41% (Atlantic) and 39% (Mediterranean) of PODs simulated under the *divAGF* model were misclassified as divergence in isolation (*div)*. Measures of robustness in the accuracy of model discrimination indicated a minimum threshold for model probability ≥ 83% required to yield a robustness of 0.95% or greater for the *divAGF* model, corroborating initial model choice inference (Table 3-2). Overall, there was clear discrimination between models excluding (*divAGF, div*) and including (*im, divSC, divAGFSC*) ongoing gene flow since divergence, supporting the hypothesis that *M. planulatus* and *M. galloprovincialis* have experienced a period of historical isolation.

Genome-wide heterogeneous genetic drift and migration

I tested whether including among-locus variation in genetic drift and migration provided an improved model fit to the observed genetic data. Comparisons of the heterogeneous models under the best demographic scenario (*divAGF*; inferred from initial homogeneous model comparisons) accounting for heterogeneity in N_e and m provided an improved probability when compared to the original *divAGF* model with homogeneous parameters (Table S3-4). Specifically, I found that models allowing N_e to vary above the genomic background (hetero2) with either homogeneous or heterogeneous m, outperformed all other models with a cumulative posterior probability >90% for both population comparisons.

Demographic parameter inference from the best models

I estimated posterior density distributions for divergence time parameters corresponding to the *divAGF* model with the highest estimated probabilities in which N_e parameter varied above the genomic background (hetero2) and m was either homogeneous (Mediterranean-Tasmania) or heterogeneous (Atlantic-Tasmania). Divergence time parameters were contained within the prior distribution and well differentiated in both population comparisons (Mediterranean-Tasmania Mode: 67,842 generations, CI 95% [52,456-107,355]; Atlantic-Tasmania: Mode: 177,661 generations, CI

95% [124,963-360,622]; Figure 3-5). Divergence time estimates suggest that Australian *M. planulatus* started diverging in allopatry between 100,000 and up to 800,000 years ago, and is likely to have experienced periods of low historical gene flow promoted by interglacial periods of trans-equatorial migration throughout the late Pleistocene. However, estimates of ancient migration (*ma*), including the time of onset of ancient gene flow since the present (T_{nc}), were poorly resolved and could not be estimated with accuracy.

Discussion

In this study, I document ongoing (post-introduction) introgression between closely-related native and introduced species whose genomes are also shaped by recent divergence histories and historical gene flow. Using transcriptome-wide data, I uncover multiple recent introductions of *M. galloprovincialis* into southeastern Australia are derived from genetically divergent Mediterranean and Atlantic lineages. Accounting for species relationships, I also provide evidence that high rates of genome-wide introgression between introduced and endemic populations are associated with both introductions. Increased resolution of transcriptome-wide comparisons of *M. planulatus* sampled in Tasmania against northern hemisphere species validated this taxon's strong genetic affinity to *M. galloprovincialis*. However, high variance in gene tree topologies implied extensive levels of incomplete lineage sorting in Tasmanian mussels. Demographic inferences indicated late-Pleistocene divergence times and low levels of historical gene flow between northern *M. galloprovincialis* and the Tasmanian endemic lineage, suggesting that native and introduced taxa have experienced a period of historical isolation of at least 100,000 years prior to present day human-mediated contact. Taken together, this study demonstrates the utility of genomic data for detangling the contributions of contemporary invasive hybridisation from signatures left by historical gene flow and recent divergence histories in high dispersal marine species.

Population relationships and introgression with multiple introduced source populations

Documenting the invasive spread of *M. galloprovincialis* in Australia has been the subject of a number of previous genetic investigations; however, taxonomic resolution of introduced and native taxa has been hampered by low genetic differentiation between populations and high levels of marker discordance (Hilbish et al. 2000; Gérard et al. 2008; Westfall and Gardner 2010). Representing a much larger proportion of genetic variation than previous approaches, the findings of this Chapter confirm strong species tree discordance between mitochondrial and nuclear loci (Hilbish et al. 2000; Gérard et al. 2008), even when genome-wide data are analysed. Network analysis of 12 protein-

coding mitochondrial genes from three Australian populations placed southern lineage mitochondrial haplotypes in a divergent clade distinct from northern *M. galloprovincialis* (Figure 3-2). Additionally, haplotypes across all genes remained paraphyletic for northern *M. galloprovincialis* and *M. edulis* sister taxa despite 2.5 million years of divergence (Roux et al. 2014a); these results further implicate historical isolation between southern and northern lineages (e.g., Gérard et al. 2008). In contrast to mitochondrial loci, variation across the nuclear genomic background of Australian mussels alongside northern hemisphere taxa validated strong genetic affinities to *M. galloprovincialis*, suggesting a much closer genetic relationship to the invasive taxon (e.g., Hilbish et al. 2000; Fraïsse et al. 2018b).

Comparisons of genotypic variance across a larger panel of nuclear SNP markers revealed genetic separation between three Australian populations and northern *M. galloprovincialis.* Specifically, populations from Sydney Harbour and Batemans Bay were genetically differentiated from a more divergent Tasmanian lineage and northern *M. galloprovincialis* populations sampled from its native range (Figure 3-3). Within these two populations, all individuals displayed mixed ancestry with high genomic contributions (i.e. 33-82% ancestry proportions; Figure 3-3B) from northern *M. galloprovincialis*, pointing to likely introductions of northern genotypes into mainland Australia (i.e. Westfall and Gardner 2010). Indeed, formal tests of migration in TreeMix revealed the strongest evidence for introgression from eastern Mediterranean *M. galloprovincialis* into Sydney Harbour, and migration from Atlantic *M. galloprovincialis* into Batemans Bay (Figure 3-4B), confirming post-introduction admixture between native and introduced populations. Additionally, I did not identify any non-introgressed *M. planulatus* individuals in these populations (i.e. with the exception of Tasmania), suggesting that introgression is extensive and that both introductions are accompanied by high rates of introgression likely from the native into the introduced genetic backgrounds. Taken together, these results provide strong evidence for genome-wide introgression between *M. planulatus* and introduced *M. galloprovincialis* as a result of two contemporary introductions from divergent Mediterranean and Atlantic source lineages. However, additional sampling from the native range across Europe and northern Africa will be essential to provide finer resolution regarding the exact sources of introduced populations.

For many non-native marine species, multiple introductions are a common feature of biological invasions (Rinquet et al. 2013; Rius et al. 2015a; Viard et al. 2016). High fecundity and dispersal capacities displayed by marine taxa appear to encourage successive introductions of large numbers larvae through human-mediated vectors (e.g., ballast water discharge and aquaculture transplantation; Carlton 1996). Repeated introductions are therefore likely to promote secondary contact between genetically differentiated lineages (e.g., Simon-Bouhet et al. 2006; Keller and Taylor 2010; Rius and

Darling 2014; Roman and Darling 2007; Uller and Leimu 2011; Jeffery et al. 2018). For example, differentiated lineages of the invasive European green crab, *Carcinus maenas* have been independently introduced into the eastern coast of Northern America (Darling et al. 2008; Tepolt and Palumbi 2015), and post-introduction admixture between warm-adapted and cold-adapted introduced lineages has been posited as a potential factor in the success of invasive populations (Tepolt and Somero 2014; Jeffery et al. 2018). There is also evidence that hybridisation with native taxa or divergent introduced lineages may benefit founding populations through demographic processes by reducing the severity of Allee effects expected for marine introduced populations (Mesgaran et al. 2016; Gagnaire et al. 2018). Overall, the findings of this Chapter are in line with general perceptions that successful marine introductions are likely to involve propagules from multiple and potentially diverse source populations (Lockwood et al. 2005; Riquet et al. 2013; Rius et al. 2015; Viard et al. 2016). However, whether post-introduction admixture of divergent *M. galloprovincialis* lineages has enhanced the success of introduced populations will require additional investigation examining these synergies (Rius et al. 2015).

Hybridisation between introduced and native genotypes may also increase genetic diversity available for adaptation through the introgression of diverse genetic backgrounds into native genetic diversity (e.g., Fitzpatrick et al. 2010). In Australia, a particular concern is that one admixed population retained signatures of divergent northern *M. edulis* genetic elements likely obtained through post-introduction admixture with introgressed Atlantic *M. galloprovincialis* (rather than direct gene flow with northern *M. edulis*) (e.g., Fraïsse et al. 2015). Occurrences of gene flow evidenced by three-population tests revealed significantly higher proportions of shared ancestry between Batemans Bay and both Atlantic *M. galloprovincialis* and *M. edulis* than expected by incomplete lineage sorting. These findings are consistent with previous investigations implicating asymmetric introgression of *M. edulis* genes into Atlantic *M. galloprovincialis* populations in southwestern Europe (where *M. galloprovincialis* is native) (Rawson and Hilbish 1998; Bierne et al. 2003b; Boon et al. 2009; Gosset and Bierne 2013; Roux et al. 2014; Fraïsse et al. 2014). Additionally, a recent study (A. Simon unpublished results) has shown that Mediterranean lineage mussels introduced into five Atlantic shipping ports show mixed *M. edulis* ancestry leading to genetic separation from local genomic backgrounds. Interestingly, introduced populations also displayed continuity in genetic composition among ports (even when surrounding native populations were of Atlantic *M. galloprovincialis* decent), suggesting that admixture with *M. edulis* has occurred prior to regional introductions. The findings in the present study therefore raise important considerations regarding the possibility of adaptive introgression of *M. edulis* genetic variation into endemic Australian genetic diversity and

the consequences of continued dispersal of these admixed genotypes throughout introduced ranges (e.g., Keller and Taylor 2010; Rius and Darling 2014; Saarman and Pogson 2015).

Analyses of introgression in TreeMix did not provide support for significant contemporary gene flow from northern *M. galloprovincialis* into Tasmanian mussels (Figure 3-4), suggesting that samples in this region are largely representative of the divergent endemic taxon (Figure 3-2; Hilbish et al. 2000; Gérard et al. 2008; Westfall and Gardner 2010; Fraïsse et al. 2018b). Consistently, ABC inferences recovered weak support for all demographic models including recent gene flow associated with human-mediated introductions (<600 years ago) in this region; I also found little evidence that Tasmanian mussels are introgressed with *M. edulis* through past admixture as suggested by previous authors (e.g., Borsa et al. 2007). Instead, high variance in gene tree topologies in TWISST indicated extensive levels of shared incomplete lineage sorting with both northern *M. edulis* and *M. trossulus*, which may account for the presence of outgroup alleles in some populations (Figure S3-2; e.g., Westfall and Gardner 2013). Furthermore, pervasive paraphyly among *M. planulatus* and *M. galloprovincialis* haplotypes in locus-specific topologies (analysed in TWISST), suggest ongoing incomplete lineage sorting (i.e. shared ancestral polymorphism). I also observed the presence of a Mediterranean *M. galloprovincialis* mitochondrial haplotype in one Tasmanian individual (Figure 3-3), indicating that we cannot entirely exclude the possibility of recent (post-introduction) introgression in this region.

Late-Pleistocene divergence between native and introduced Australian blue mussels

To clarify whether *M. planulatus* sampled in Tasmania has experienced a unique demographic history from northern *M. galloprovincialis* prior to human-mediated secondary contact, I used a comparative ABC approach to leverage signals of past isolation from the genomic data. Consistent with my hypothesis, demographic inferences provided the strongest support for a model of divergence with ancient gene flow, suggesting that northern and southern *Mytilus* species have experienced a substantial period of historical isolation predating the end of the last glacial maximum. Divergence time estimates according to this model (*divAGF*) were relatively recent, indicating that *M. planulatus* began to diverge from *M. galloprovincialis* around 100,000 and 800,000 years ago; while estimates of divergence time are dependent on generation time assumptions of 2 years, these approximated values are compatible with the fossil record indicating the presence of endemic *Mytilus* in Australia since the end of the late glacial retreat (>10,000 years bp). These estimates and are also in line with previous studies proposing separation times between northern and southern hemisphere *Mytilus* species ~0.5 to 1.3 million years ago based on mitochondrial loci (Gérard et al. 2008). I therefore

posit that both relatively recent divergence times and periods of historical contact have contributed to modest signals of contemporary genome-wide differentiation observed between northern and southern lineages.

Transequatorial migration is hypothesised as a primary source for the origins of antitropical distributions for many marine taxa (e.g., Lindberg 1991; Hilbish et al. 2000; Gerrard et al. 2008). Our data suggest that southward migrations leading to the establishment of the Australian *Mytilus* lineage took place during the late Pleistocene, when gene flow was likely facilitated by glacial melting and the formation of cool-water refugia across the equator (Lindberg 1991). Although ABC inference did not provide robust conclusions regarding the rate or symmetry of historical gene flow, including the time of the onset of allopatric isolation (i.e. termination of ancient gene flow), I found significant improvements in model fit to our genetic data (compared to strict isolation) when low levels of historical gene flow (m=0.0001 per generation) were accounted for in our inferences. However, whether *M. planulatus* experienced low levels of ancient gene flow during a single bout of migration or intermittent transequatorial migrations will require further investigations of explicit models incorporating periodic migration (e.g., Fraïsse et al. 2018a). Nevertheless, consistent rejection of all other models in which gene flow was ongoing, including panmixia, isolation-with-migration, and models including very recent (post-introduction) secondary contact, supports the hypothesis that Tasmanian mussels are largely representative of the endemic Australian taxon.

Modelling the effects of selection on genome-wide variation

Modelling the effects of selection on genome-wide variation has been shown to significantly improve model discrimination between shared ancestral polymorphism and gene flow in semi-isolated species (Roux et al. 2016). I found that incorporating genome-wide heterogeneity in the rate of introgression and genetic drift as a result of linked selection provided improvements in model fit compared to models assuming fixed among-locus values for demographic parameters (Table S3-4). Specifically, substantial support for models incorporating higher than average N_e values among loci, suggests that linked selection may not only act to reduce genetic variation, but may also increase variation by favouring polymorphisms at some loci. For example, elevated N_e may reflect locus-specific selection or polygenic balancing selection (reviewed in Charlesworth 2006; Guerrero and Hahn 2017), unaccounted introgression from unsampled ghost taxa (e.g., Butlin et al. 2014; Vernot and Akey 2015; Frantz et al. 2015), or intrinsic processes affecting variation in the genome differentially, such as recombination. Alternatively, increased diversity at linked neutral variation can result from asymmetrical sorting of ancestral polymorphisms between descendent taxa (Guerrero and Hahn

2017). Importantly, such processes can mimic signatures of migration, potentially leading to false inferences of gene flow based on differentiation statistics that rely on symmetrical distributions of ancestral polymorphism (i.e. three-population statistics; Reich et al. 2009). Strong support for genome-wide heterogeneity in genetic drift and introgression in this study corroborates that processes leading to both low and high levels of variation at synonymous loci (including levels of ancestral diversity) are important factors to account for when reconstructing population histories and estimating demographic parameters (Guerrero and Hahn 2017; e.g., *Anopheles gambiae* species complex, Fontaine et al. 2015).

Implications for species delimitation and endemic diversity

This study builds upon previous genetic studies investigating *M. galloprovincialis* introductions and its interactions with endemic southern hemisphere lineages (e.g., Borsa et al. 2007; Gérard et al. 2008; Westfall and Gardner 2010). The findings of this Chapter overcome several limitations in interpreting species identification and signatures of hybridisation when conclusions are based on few loci assumed to be diagnostic of northern and southern lineages. For example, the small number of *M. galloprovincialis* and *M. edulis* hybrids identified in Australia in previous studies (Westfall and Gardner 2010; 2013) is difficult to interpret when introduced populations are likely carrying introgressed variation from outgroup taxa or when shared polymorphism is expected to be extensive.

Many loci are required for adequate resolution of endemic and introduced populations. Based on the genomic data presented here, I infer that shared polymorphisms between *M. planulatus* (Tasmania) and *M. galloprovincialis* is mostly due to recent divergence and possibly associated with historical gene flow, but I found little evidence for contemporary introgression from human-mediated introduction. Demographic inferences supported a substantial period of isolation between *M. planulatus* and *M. galloprovincialis* of at least 100,000 years. This result supports previous proposals to assign the endemic Australian lineage with a regional subspecies status (e.g., Borsa et al. 2007, Borsa and Daguin 2000, Gérard et al. 2008, Westfall and Gardner 2010, Hilbish et al. 2000). Further clarification, however, will be required on whether *M. planulatus* has experienced historical introgression with southern clades in New Zealand and South America (e.g., Larraín et al. 2018). Genetic investigations to date have used either insufficient loci to resolve species relationships and migration (e.g., Westfall and Gardner 2010) or have sampled only a single landmass (e.g., Gardner et al. 2016). A recent genetic survey has confirmed genomic differentiation between these southern clades (e.g., Gérard et al. 2008), including the Kergulen islands (using ancestry-informative SNPs); however, introgression with *M. planulatus* was not investigated (Fraïsse et al. 2018b). The

relationship between two southern *Mytilus* taxa, that is, *M. planulatus* in Australasia that is more related to *M. galloprovincialis* (this study) and *M. platensis* in South America and the Kergulen islands that is more related to *M. edulis* (Fraïsse et al. 2018b) will require additional investigation.

The findings of this Chapter also demonstrate that diagnostic methods used to identify hemisphere origins based on mitochondrial loci are not a reliable method for differentiating native and introduced individuals, especially if hybridisation is pervasive (e.g., Colgan and Middlefart 2011; Dias et al. 2014; Ab Rahim et al. 2016). Interestingly, I observed asymmetry in the frequency of northern mitochondria in the southern lineage genomic background (Figure 3-3A). Samples from Sydney Harbour were fixed for the *M. galloprovincialis* mitochondrial marker COIII, compared to northern mitochondria in only three individuals in Batemans Bay and a single individual with a Tasmanian nuclear genetic background. Significant overrepresentation of northern mitochondria has been previously identified among samples in New Zealand, implicating possible selection for northern maternal and southern paternal parents (Westfall and Gardner 2013; also see Rawson and Hilbish 1998; Śmietanka et al. 2014). The possibility of asymmetrical introgression of northern mitochondria has significant implications for the genetic integrity of endemic populations, including extensive alterations to native genetic diversity through introgressive swamping of native genotypes (Fitzpatrick et al. 2010; Todesco et al. 2016). However, further investigation into this hypothesis is necessary to evaluate whether hybridisation and introgression may play a role in colonisation success of *M. galloprovincialis*.

Monitoring M. galloprovincialis introductions in Australia

Mytilus mussels are notorious ecosystem engineers and could impose significant alterations to rocky intertidal communities (Braby and Somero 2006b). From an applied perspective, severe ecological impacts of *M. galloprovincialis* in other parts of the world (Rawson et al. 1999; Branch and Steffani 2004; Bownes and McQuaid 2006), including negative effects on aquaculture populations (Dias et al. 2014; Michalek et al. 2016; Jones and Creeper 2006) warrant caution against *M. galloprovincialis* introductions into Australian industries. For the case of endemic *Mytilus*, however, it is evident that the use of many markers presents a significant impediment for rapid and efficient identification of invasive and endemic populations. Additionally, eradication of introduced populations will be unlikely for a marine species with such high dispersal potential (Larraín et al. 2018).

The findings of this Chapter raise ethical and practical questions regarding the value of recognising and protecting native genetic diversity when populations are challenged with extensive invasive

hybridisation (e.g., Fitzpatrick et al. 2010). For taxa with large effective population sizes, theory suggests that small numbers of migrants per generation (i.e. $N_e m$ > 1-25) are sufficient to homogenise populations (Waples 1998; Crandall et al. 2018) and eliminate evolutionary independence between lineages (Waples and Gaggiotti 2006). While genomic studies (e.g., Tepolt and Palumbi 2015), including the present investigation, illustrate the power of high-throughput approaches for distinguishing native diversity from different sources (i.e. lineages) of introduced populations, it is evident that even genome-scale surveys may not overcome the combined effects high migration rates and large effective population sizes characteristic of high gene flow marine taxa (e.g., Gagnaire et al. 2015). In this study, I show that multiple introduced sources are only detected when significant structure exists in the native range. It is therefore likely that even genome-scale approaches will not be sufficiently robust to resolve exact source populations in the absence of strong differentiation, despite adequate sampling of native range variation (i.e. eastern and western Mediterranean *M. galloprovincialis*). In such cases, surveys for marine invasive species using environmental DNA (Thomsen et al. 2012; Kelly et al. 2014) may not be well suited to dissect intraspecific non-native diversity or even interspecific variation, despite holding promise for detecting marine invasive taxa in rare or unrecognised taxonomic groups (Kelley et al. 2014; reviewed in Bourne et al. 2018).

Knowledge regarding the potential ecological consequences associated with introduced populations (i.e. niche displacement; California, Rawson et al. 1999; South Africa, Branch and Steffani 2004) will therefore be essential for monitoring and minimising *M. galloprovincialis* spread in sympatric ranges. Future research should focus on temporal and spatial sampling to assess the stability of hybrid populations (Glotzbecker et al. 2016; Ayres et al. 2008; Strong and Ayres 2013) and whether potential expansions may have a role in marine invasive spread. Indeed, there is growing evidence that interspecific hybridisation and introgression between native and introduced taxa is pervasive, and this could be an important source of adaptive variation for some marine invasive species (Le Roux and Wieczorek 2009; Sloop et al. 2009; Roux et al. 2013). A recent study found low levels of asymmetric introgression from native *M. trossulus* into the invasive *M. galloprovincialis* nuclear genomic background in its introduced range in California (Saarman and Pogson 2015), consistent with predictions that introduced populations furthest from the source of introduction should experience stronger introgression into the invading species (Currat et al. 2008). These findings suggest that the direction of introgression and the potential for introgression swamping towards endemic taxa are likely driven by the demography of the invasion (i.e. relative population sizes of native and non-native populations) rather than by selection processes. I hypothesise that an absence of pure *M. galloprovincialis* individuals in this study indicates either few present-day introductions or significant and rapid contributions of introgression to the genomic composition of Australian mussels during the

early stages of introduction. However, sampling of greater numbers of individuals and additional populations will be required to investigate the direction of introgression and to assess whether introgression swamping can lead to the elimination of pure native genomes on ecologically-relevant timescales (Riley et al. 2003; Todesco et al. 2016; Glotzbecker et al. 2016). Such data will be critical for informing our understanding of the scope and potential of long-term (evolutionary) impacts of biological introductions on receiving marine communities.

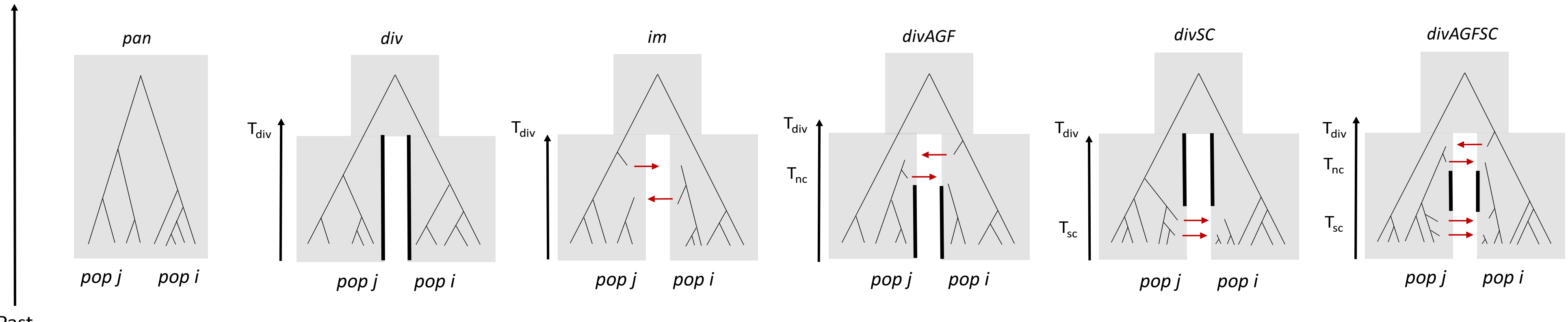


Figure 3-1. Competing models of divergence between northern *M. galloprovincialis* (*pop j*) and native *M. planulatus* (*pop i*). The *pan* model, assumes that populations belong to the same gene pool. The *div* model assumes populations evolve independently with no gene flow since their divergence. In the *im* model, populations diverge with ongoing gene flow to the present day. The *divAGF* model assumes bidirectional migration is restricted to the early stages of speciation from T_{div} to a more recent time (T_{nc}) up to the last glacial maximum (20,000 years ago), after which populations evolve independently with no migration. This scenario implicates historical transequatorial migration between the northern and southern hemispheres facilitated by cyclical glacial cooling of the oceans during the late Pleistocene. In the *divSC* model, populations evolve in allopatry until a more recent time of human-mediated secondary contact (T_{sc}), when populations begin to exchange genes. This scenario tests explicitly for the presence of post-introduction gene flow from northern *M. galloprovincialis* into Australian populations assuming that the onset of migration occurs after the earliest record of European contact (<600 years ago). Finally, the *divAGFSC* model assumes that populations diverged with ancient migration for a period of time, after which they evolved in allopatry; genetic exchange is re-established at T_{sc} following recent secondary contact as a result of human-mediated introductions.

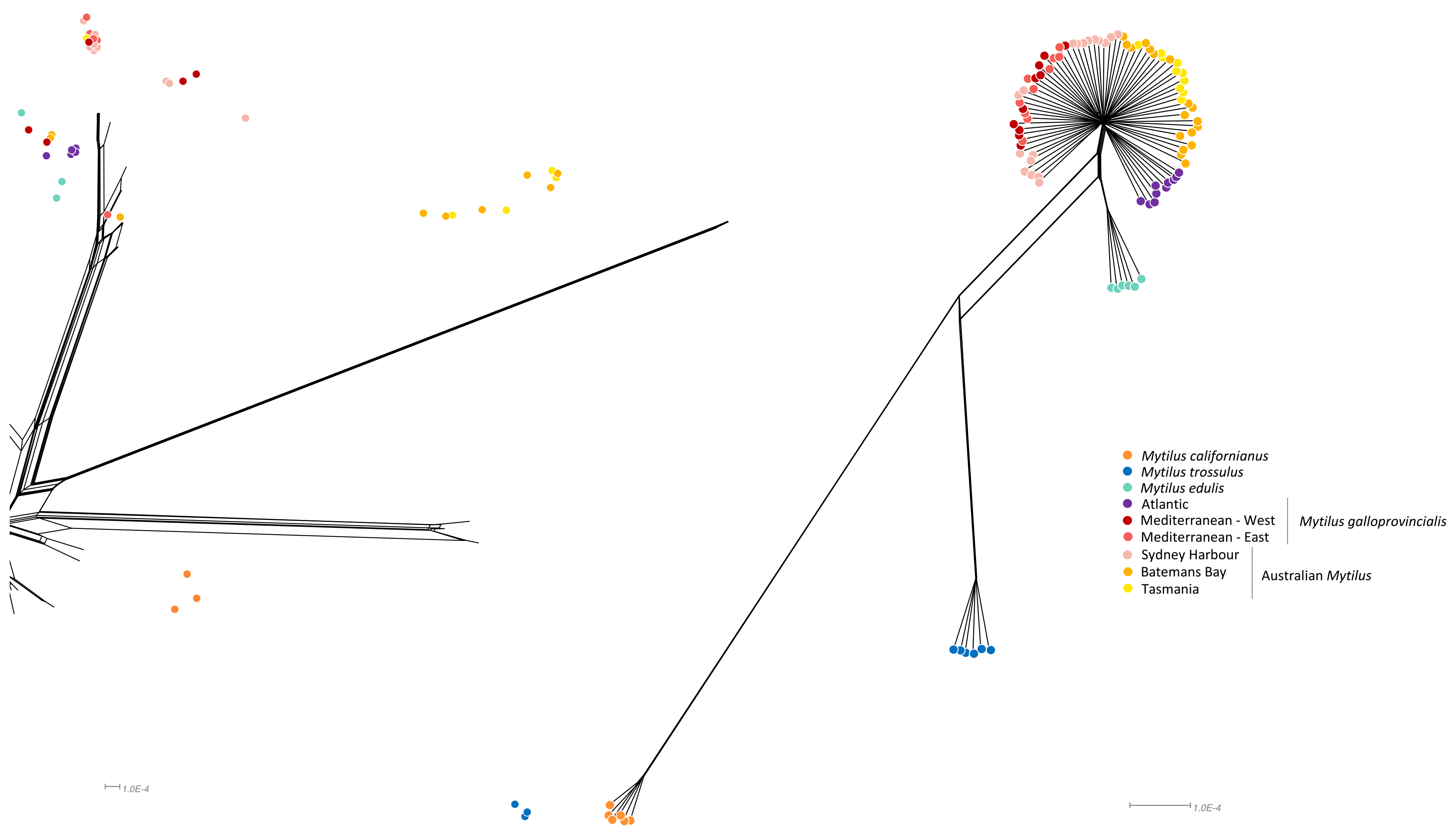


Figure 3-2. Haplotype genetic network of Australian samples (*M. planulatus* samples from Tasmania are shown in yellow) and four outgroup taxa. Network phylogenies are constructed from 12 mitochondrial genes (left) and 2620 nuclear contigs (right).

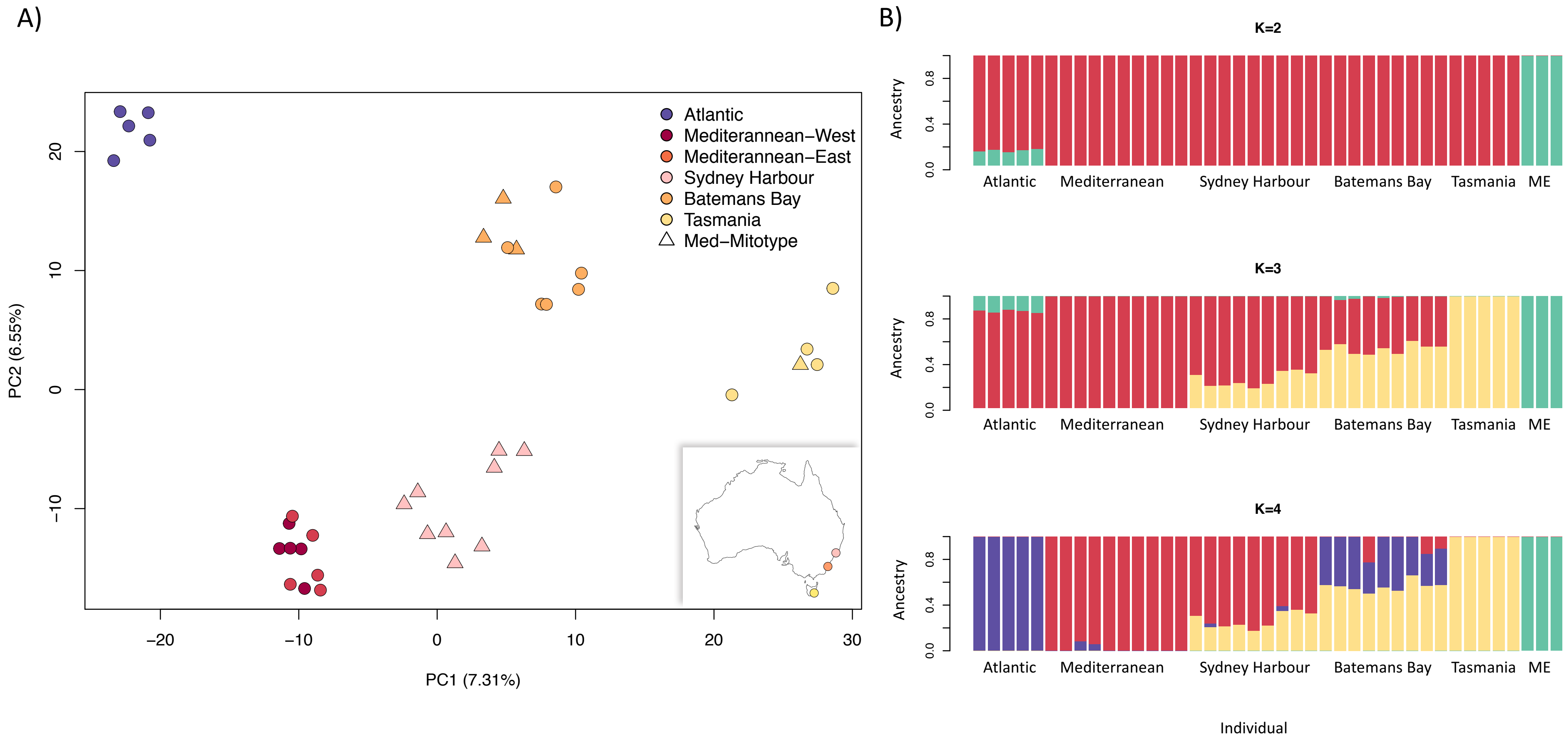
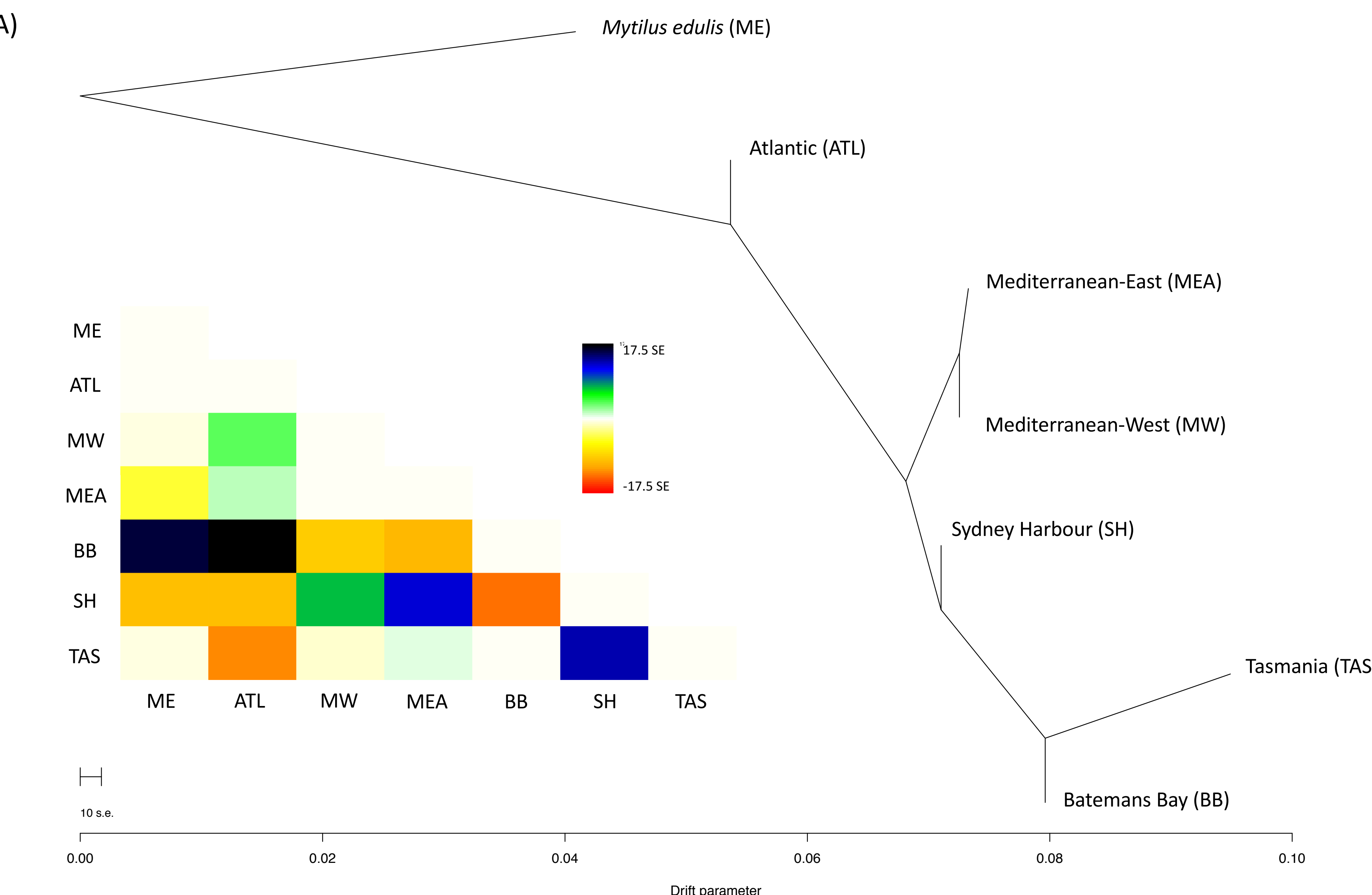
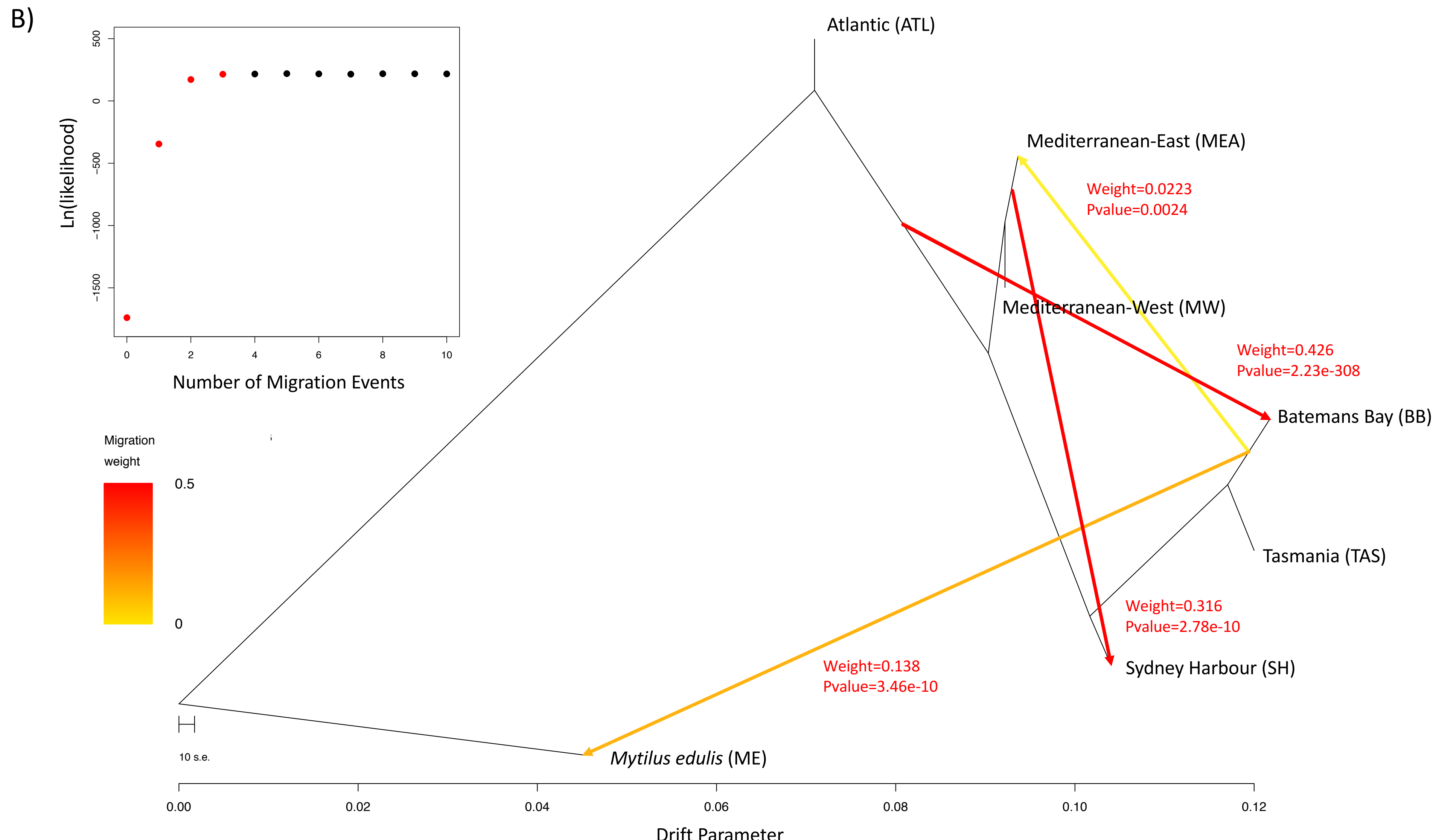


Figure 3-3. A) Principal component analysis of three populations sampled in Australia (shown in inset map) and northern *M, galloprovincialis* sampled from its native range in the Atlantic and Mediterranean Sea. Colours correspond to populations and individuals marked with a triangle indicate Australian samples carrying a Mediterranean *M. galloprovincialis* mitochondria based on the COIII marker. **B)** ADMIXTURE analyses for K=2 to K=4 genetic clusters. Each bar represents an individual belonging to one or more ancestral clusters, corresponding to different colours.


A)
Mytilus edulis (ME)
Atlantic (ATL)
Mediterranean-East (MEA)
Mediterranean-West (MW)
Sydney Harbour (SH)
Tasmania (TAS)
Batemans Bay (BB)
ME
ATL
MW
MEA
BB
SH
TAS
17.5 SE
-17.5 SE
10 s.e.
0.00
0.02
0.04
0.06
0.08
0.10
Drift parameter


B)
Ln(likelihood)
500
0
-500
-1000
-1500
0
2
4
6
8
10
Number of Migration Events
Migration
weight
0.5
0
10 s.e.
Atlantic (ATL)
Mediterranean-East (MEA)
Weight=0.0223
Pvalue=0.0024
Mediterranean-West (MW)
Weight=0.426
Pvalue=2.23e-308
Batemans Bay (BB)
Tasmania (TAS)
Weight=0.316
Pvalue=2.78e-10
Sydney Harbour (SH)
Weight=0.138
Pvalue=3.46e-10
Mytilus edulis (ME)
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Drift Parameter

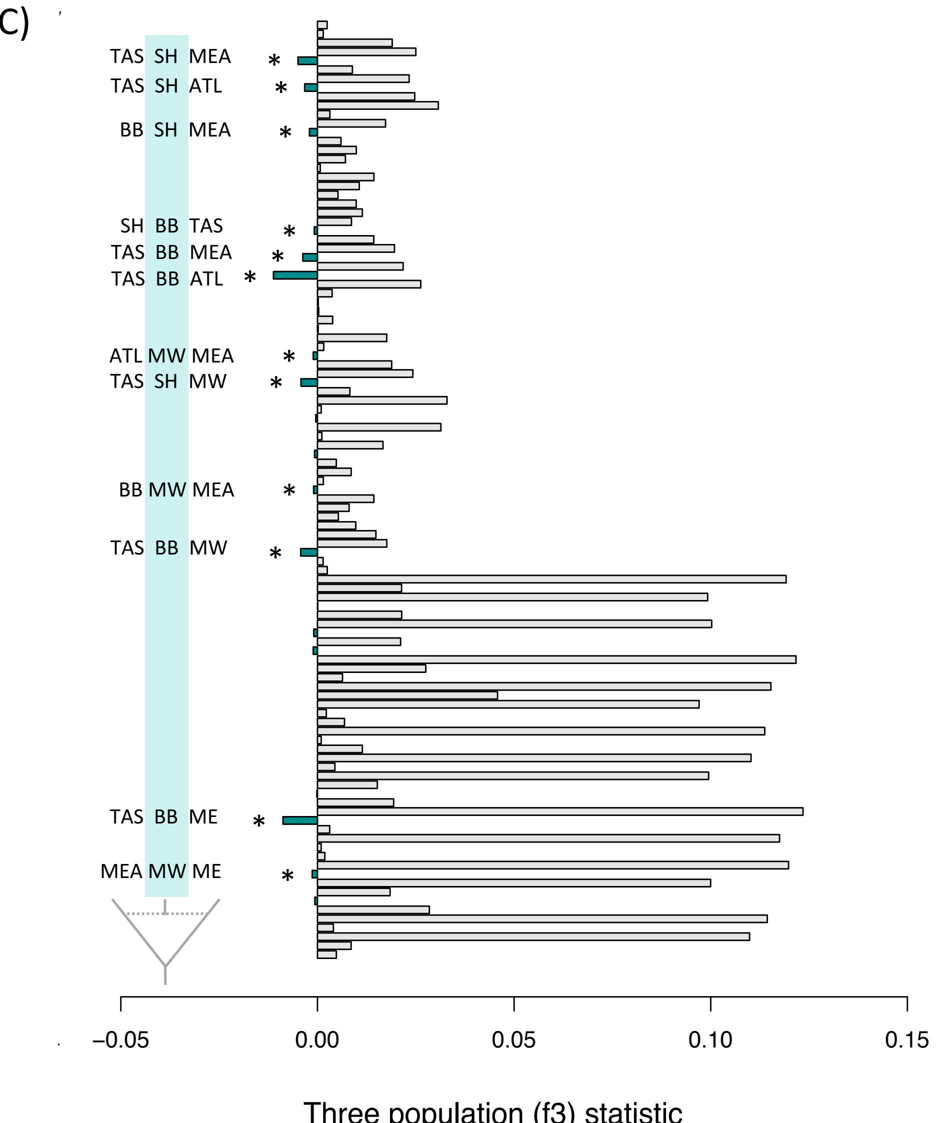


Figure 3-4. Maximum likelihood trees with Australian *M. planulatus* and *M. galloprovincialis* populations using only *M. edulis* as an outgroup inferred by TreeMix **A)** Maximum likelihood population tree without migration. The drift parameter indicates the amount of genetic drift that separates groups. Under a model of zero migration, the heat colours indicate pairwise population residual genotypic covariance of allele frequencies. The darkest boxes in this residual matrix indicate high genotypic covariance between Sydney Harbour or Batemans Bay populations with northern taxa. **B)** Maximum likelihood population tree including four migration events. The addition of four admixture events significantly improved the fit of the population tree to the genetic data compared to a model with no migration. Significant p-values indicate support for individual migration edges and the corresponding direction and weight; **C)** Summary of results from the three-population (f_3) test. The right panel shows f_3 statistics for all population combinations including *M. edulis* as an outgroup, with significant and negative f_3 values are indicated with an asterisk. The corresponding three-population combination is shown to the left of the asterisk. The middle population marked with green shows strong evidence of being admixed with putative ancestral populations indicated on either side.

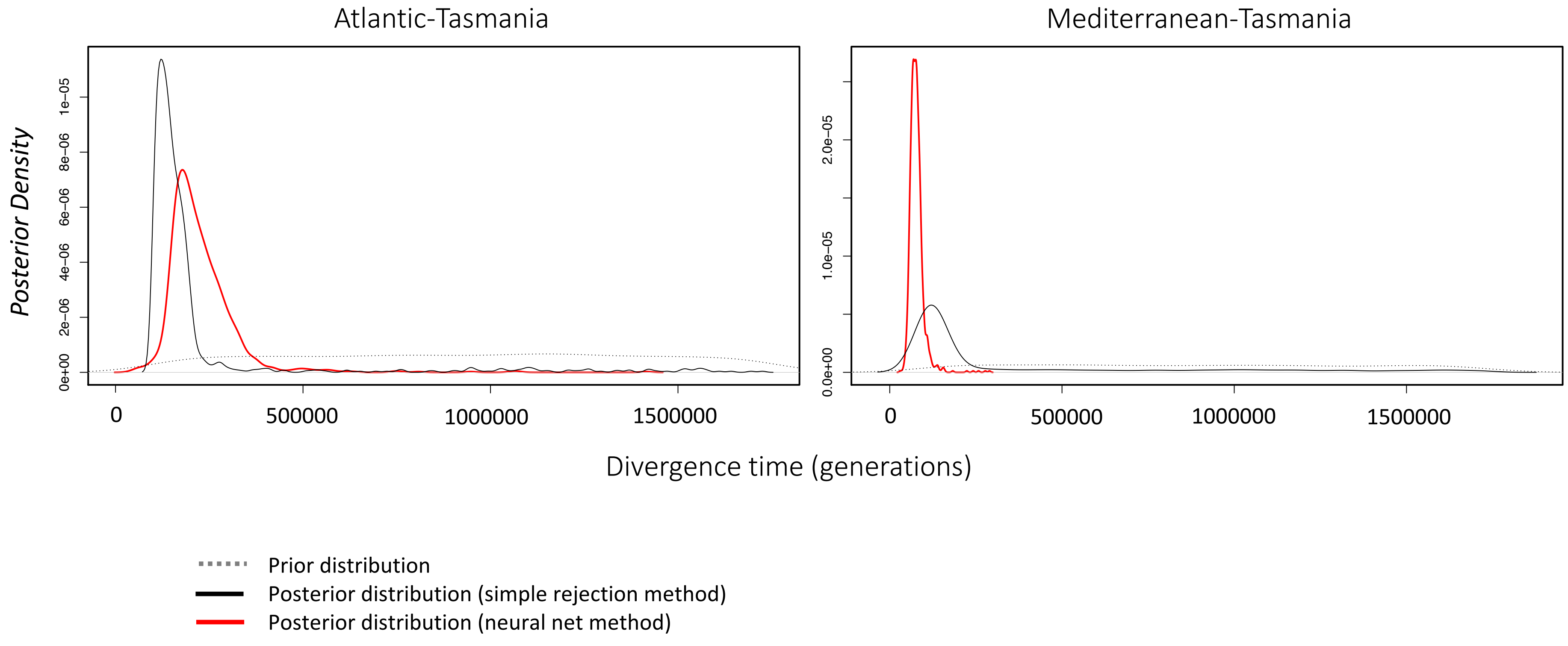


Figure 3-5. Posterior distributions for divergence time parameters inferred from ABC analyses for two population pairs based on highest probability demographic model (*divAGF*) that account for heterogeneous N_e and m parameters. The parameter plots correspond to these models with the highest estimated probability in which the N_e parameter was hetero2 and m was either homogeneous (Mediterranean-Tasmania) or hetero1 (Atlantic-Tasmanian).

Table 3-1. Details of samples and collection locations. Samples in bold indicate putative species identity (based on mitochondrial genome) and range.

Taxon	Sampling Location	Range	Samples Sequenced (RNAseq)
Mytilus californianus	Scripps Oceanographic Institute, California, USA	Native	3
Mytilus trossulus	Light House Park, British Columbia, Canada	Native	3
Mytilus edulis	Darling Marine Station, Maine, USA	Native	3
Mytilus galloprovincialis	Primel, France (Atlantic)	Native	5
	Crique des Issambles, France (Mediterranean)	Native	5
	Herceg Novi, Montenegro (Mediterranean)	Native	5
Mytilus galloprovincialis	**Sydney Harbour, New South Wales, Australia**	**Introduced**	**9**
	Batemans Bay, New South Wales, Australia	**Introduced**	**3**
Mytilus planulatus	**Batemans Bay, New South Wales, Australia**	**Native**	**6**
	Spring Bay, Tasmania, Australia	**Native**	**5**

Table 3-2. Summary of demographic model selection under an approximate Bayesian computation (ABC) framework. Models posterior probabilities for models assuming homogeneous population size and migration parameters for: **A)** the Mediterranean-Tasmania population pair; and **B)** the Atlantic-Tasmania population pair. The simple rejection method was applied in cases where not all models had accepted values at the specified tolerance rate. **Bold** indicates the highest probability model for each comparison.

A)

		Demographic Model Probability: Proportion of accepted simulations					
Tolerance	**Method**	***pan***	***div***	***im***	***divSC***	***divAGF***	***divAGFSC***
0.001	Rejection	0.0000	0.1320	0.0417	0.0188	**0.7835**	0.0240
0.01	Rejection	0.0000	0.1955	0.0708	0.0546	**0.6130**	0.0660
0.001	Rejection	-	0.1588	0.0266	0.0100	**0.7894**	0.0152
0.01	Rejection	-	0.1972	0.0664	0.0398	**0.6421**	0.0545
0.001	Neural Net	-	0.0328	0.0375	0.0056	**0.9094**	0.0142
0.01	Neural Net	-	0.0618	0.0051	0.0026	**0.9267**	0.0038

B)

		Demographic Model Probability: Proportion of accepted simulations					
Tolerance	**Method**	***pan***	***div***	***im***	***divSC***	***divAGF***	***divAGFSC***
0.001	Rejection	0.0000	0.3868	0.0037	0.0015	**0.6047**	0.0033
0.01	Rejection	0.0000	0.3052	0.0135	0.0069	**0.6654**	0.0090
0.001	Rejection	-	0.4252	0.0036	0.0028	**0.5654**	0.0030
0.01	Rejection	-	0.3129	0.0116	0.0062	**0.6611**	0.0081
0.001	Neural Net	-	0.0527	0.0019	0.0059	**0.8273**	0.1123
0.01	Neural Net	-	0.0408	0.0010	0.0009	**0.9557**	0.0015

Chapter 4. Parallel genetic differentiation in multiple introduced populations of the marine mussel *Mytilus galloprovincialis* sheds light on long non-coding RNA evolution.

Abstract

Non-native species are likely to experience new ecological conditions under which they did not evolve and there is growing evidence that non-native species are frequently exposed to strong selection following introduction. Here, I investigate the repeatability of post-introduction evolution in replicated worldwide introductions of the marine mussel, *Mytilus galloprovincialis*. I test the hypothesis that introduced *M. galloprovincialis* populations in Australia share strong deviations in genetic differentiation with replicated introductions in California and South Africa originating from the same native source lineages. I also identify major non-coding elements of the transcriptome, long non-coding RNAs (lncRNA), to estimate their relative contributions to population structure and parallel genetic differentiation in introduced populations. Comparisons of transcriptome-wide variation between introduced populations revealed parallel genetic differentiation from their native genetic backgrounds at the level of individual contigs; however, repeated genetic changes did not involve the same nucleotide polymorphisms. Inconsistent evidence for repeated selective sweep patterns leading to reduced nucleotide diversity at shared outlier contigs suggested that parallel genetic differentiation is likely influenced by both selective and non-selective processes. Analyses of variation among lncRNA transcripts indicated significantly lower contributions to intraspecific genetic variation compared to all expressed transcripts, suggesting that lncRNAs evolve under selective constraints and are unlikely to underlie population-specific differentiation on timescales relevant for species introductions. These findings build on existing insights regarding the genomic scale at which differentiation may be repeatable in introduced *M. galloprovincialis* populations and shed light on the genomic architecture of post-introduction evolution in high gene flow marine non-native species.

Introduction

Biological invasions are among the leading threats to global biodiversity and are considered a central component of global change (Simberloff 2013; Chown et al. 2015). The impacts of introduced species (including invasive species) on native communities and ecosystems can be diverse and complex, and subsequently difficult to predict (Odour et al. 2013). Yet, a common factor among successful non-native species is their initial ability to establish populations and become widespread across new introduced environments (Prentis et al. 2008; Colautti and Lau 2015). Since the earliest studies on biological invasions (e.g., Baker 1965; Baker 1974), the literature now contains numerous examples of shared ecological traits contributing to increased survival and strong competitive abilities in introduced species (e.g., including rapid growth rates and high fecundity through asexual reproduction; reviewed in Richardson 2011). There is also growing evidence that local adaptation is a critical feature of species occupying novel habitats and a significant factor in the success of introduced populations (reviewed in Dlugosch and Parker 2008; Prentis et al. 2008; Colautti and Lau 2015). Indeed, differentiation in ecologically important traits related to life history strategies, reproduction, growth and competition have been documented among introduced and native populations in both plants and animals (e.g., Phillips et al. 2006; Lee 2016; Turner et al. 2014; Vandepitte et al. 2014; Ferrero et al. 2015; Colautti and Lau 2015). We know relatively less, however, about the shared genetic traits and the sources of heritable genetic variation that may underlie common adaptive responses in successful invasions (Bock et al. 2015).

Genetic and genomic studies examining parallel evolutionary change (i.e. the extent to which the same genetic mechanisms underlie the evolution of similar phenotypes in independent lineages; Elmer and Meyer 2011) have yielded evidence that adaptation can be highly constrained and repeatable in natural populations (Conte et al. 2012). Studies have demonstrated replicated genetic changes across distinct lineages at various genomic scales, ranging from single amino acid polymorphisms (e.g., Hoekstra et al. 2006) to individual genes (e.g., Stern and Orgogozo 2009) and large genomic regions (e.g., Renaut et al. 2014). Comparative evidence for parallel evolution in response to selection has also been demonstrated among closely related populations (e.g., Turner et al. 2010; Jones et al. 2012), suggesting that adaptation may be predicable even on timescales relevant for invasive species research. However, most studies of genetic repeatability are predicated on the assumptions that populations have evolved under similar selective pressures (Arendt and Reznick 2008; Stern 2013). Invasive species, however, are likely to experience new ecological conditions under which they did not evolve (Colautti and Lau 2015) and have been shown to occupy vastly different abiotic conditions and habitats in their introduced ranges (Tepolt 2015; Guzinski et al. 2018).

The rate of adaptation to novel environments depends largely on the amount of genetic variation available to selection during the course of introduction (Prentis et al. 2008; Colautti and Lau 2015; Estoup et al. 2016). Given the recent timescales for most species introductions, selection on pre-existing standing genetic variation is expected to be the primary genetic mechanism of adaptation in introduced populations (Barret and Schluter 2008; Prentis et al. 2008). Alternatively, if favoured novel mutations arise independently in introduced populations, selection on new alleles can lead to strong differentiation between native and introduced populations, particularly when effective population sizes are large (Kimura 1983; Bock et al. 2015). The demographic history of introduced species, however, poses several challenges for detangling the effects of selection from genetic differentiation due to neutral processes (Dlugosch et al. 2015; reviewed in Sherman et al. 2016). Strong founder events or genetic bottlenecks can affect the distribution of both neutrally evolving and selected genetic variants in introduced populations (Dlugosch and Parker 2008; Excoffier et al. 2009; Estoup et al. 2016). For example, rare alleles that pass through genetic bottlenecks may become lost or drastically increase in frequency by genetic drift when introduced population sizes are small (Nei et al. 1975; Excoffier et al. 2009). Strong genetic drift and weak purifying selection may also allow deleterious mutations to accumulate at the margins of invading populations (Klopfstein et al. 2005; Peischl and Excoffier 2015; Plough 2016), leading to greater genetic loads in introduced populations relative to those in the native range (Peischl et al. 2013). Both processes may lead to pronounced genetic differentiation when introduced and native populations are compared. Such processes can also inflate the neutral variance in measures of population differentiation (i.e. underestimate variance in the estimated null F_{ST} distribution; Nei and Maruyama 1975) against which outlier loci are inferred, potentially confounding signatures of adaptive evolution and selectively favoured variation in non-native populations (Bierne et al. 2013; Bock et al. 2015; Campagna et al. 2015).

In the marine realm, the processes limiting the ability of invading populations to adapt and become widespread appear to be less severe than theory would predict (Nei et al. 1975; Dlugosch and Parker 2008; Viard et al. 2016). In general, high fecundity and highly dispersive larval stages are likely to promote successive introductions of large numbers of individuals through human-mediated vectors (e.g., ballast water discharge; Carlton 1996). Multiple and repeated introductions and high propagule pressures are therefore key mechanisms suspected to help founding populations avoid the negative impacts of genetic drift and associated reductions in genetic diversity (Lockwood et al. 2005; Roman and Darling 2007; Blackburn et al. 2015; Viard et al. 2016; Narum et al. 2017). To date, empirical studies in marine invasive species have supported these predictions, rarely demonstrating strong signals of genetic bottlenecks or reductions in genetic diversity in introduced populations compared

to native ranges (Riquet et al. 2013; Rius et al. 2015a; Bernardi et al. 2016; Viard et al. 2016). It is also posited that successful marine invasions are unlikely to occur without sufficiently high numbers of propagules; demographic bottlenecks and Allee effects, for example, are expected to be strong for even the most notorious marine invasive species with bentho-pelagic life cycles, as mating opportunities may be limited by larval dispersal away from founding populations (Johannesson 1998; Rius et al. 2015a; Vaird et al. 2016; Gagnaire et al. 2018).

The combination of frequent and dense introductions that likely underpin successful marine invasions provides exceptional opportunities to gain a better understanding of the role of selection and adaptation in species introductions. Multiple worldwide introductions of marine invasive species provide naturally replicated experiments of independently introduced populations occupying diverse coastal habitats. Additionally, introductions seeded with high numbers of propagules (potentially from differentiated source populations; Rius et al. 2015a) are expected to provide ample standing genetic variation on which selection may act to increase the probability of successful establishment (Roman and Darling 2007; Riquet et al. 2013; Viard et al. 2016, but see Bierne 2016). Despite these advantages, few studies have investigated genome-wide variation and population differentiation in marine non-native species across both native and introduced ranges (exceptions include: Riquet et al. 2013; Rohfritsch et al. 2013; Saarman and Pogson 2015; Tepolt and Palumbi 2015; Bernardi et al. 2016; Guzinski et al. 2018; Gagnaire et al. 2018). Inadequate resolution of population differentiation (e.g., Riquet et al. 2013) or complex demographic introduction histories (e.g., Rohfritsch et al. 2013), however, have precluded evidence for shared genetic responses in introduced populations. Additionally, genomic studies of marine non-native species have (i) largely focused on genome scans of random genetic loci (Riquet et al. 2013; Saarman and Pogson 2015; Bernardi et al. 2016; Guzinski et al. 2018; but see Tepolt and Palumbi 2015), or (ii) have not discriminated among putative sources of functional variation (i.e. protein-coding and non-coding variation) driving population genetic structure (Tepolt and Palumbi 2015). In turn, such studies have provided little insight regarding the genomic components of selectively favoured variation that may underlie population differentiation in introduced environments.

Major non-protein coding elements of the transcriptome have been recently implicated as important sources of genetic variation during invasion (Stapley et al. 2015). While the functional roles of transcribed, but non-protein coding transcripts are generally poorly understood, major components of the transcriptome, namely, long non-coding RNAs (lncRNAs) act as regulatory molecules controlling the transcription protein-coding genes and other RNAs (i.e. small RNAs) during all stages of development (reviewed in Ulitsky 2016; Gaiti et al. 2015, 2017). There is also evidence that lncRNAs

interact with other RNA molecules to form functionally important secondary and tertiary structures (Ulitsky 2016). Secondary and tertiary RNA structures have been implicated in adaptive responses to temperature changes in bacteria and there is growing evidence for similar roles in higher order species (de la Fuente et al. 2012; Somero 2018). These findings suggest that lncRNA molecules are likely to experience selective pressures on ecological timescales that are relevant for studying populations or closely related species (Somero 2018), but have yet to be used explicitly to study biological invasions. Indeed, studies in both animals and plants suggest faster sequence evolution in lncRNAs relative to protein-coding genes (Pang et al. 2006), but higher sequence conservation compared to introns and random intergenic sequences, indicating functional significance (reviewed in Ulitsky 2016; Quinn and Chang 2016; Quinn et al. 2016). Furthermore, lncRNA loci frequently harbor transposable elements which have been recently implicated as important genetic novelties contributing to rapid adaptation in invasive species (Barrón et al. 2014; Schrader et al. 2014; Stapley et al. 2015; Goubert et al. 2017; Schrader and Schmitz 2018).

Here, I take advantage of multiple worldwide introductions of the invasive marine mussel *Mytilus galloprovincialis* to investigate whether geographically independent introduced populations have undergone parallel genetic changes following introduction. I use transcriptome-wide data (i.e. both protein-coding and non-coding elements) to compare the most differentiated loci in four introduced *M. galloprovincialis* populations from California, South Africa and two locations in southeastern Australia, against their respective native source populations from the Mediterranean Sea and the Atlantic coast of Europe. As a primary objective, I searched for genomic outlier loci displaying atypical patterns of differentiation between populations to test whether introduced populations share strong deviations (i.e. outliers) in genetic differentiation relative to their native genetic backgrounds. Because resolving the sources of selectively favoured genetic variation can improve our understanding of the genetic mechanisms underlying successful invasions (Prentis et al. 2008), I also investigated whether differentiation occurs predominantly on expressed protein-coding genes or whether major non-coding elements of the transcriptome involved in gene regulatory networks are significant sources of genetic diversity for introduced populations (Stapley et al. 2015). To test the hypothesis that lncRNA molecules have a role in adaptations to selective pressures following introduction, I first identified putative lncRNA transcripts from the transcriptome data and investigated the relative contributions of lncRNAs to population structure; I predicted that lncRNAs would contribute to elevated overall differentiation among populations or individual genotypes when this subset of loci is analysed relative to variants from the full transcriptome. Alternatively, if lncRNAs are conserved within the *M. galloprovincialis* genome, I expected that transcripts would contribute significantly less to population structure compared to all expressed transcripts. I then asked

whether lncRNAs show atypical patterns of genetic differentiation between native and introduced populations as putative differentiation outliers following introduction. The genetic data presented here represent the first genome-wide investigation of *M. galloprovincialis* from multiple introduced and native populations, and the first comparative investigation of lncRNA evolution in an invasive species to date.

Methods

Study system

As one of the world most widespread marine invasive species, *M. galloprovincialis* has established introduced populations in multiple regions of the globe, spanning several continents outside of its native range in the Mediterranean Sea and the Atlantic coast of Europe (McDonald et al. 1991; Daguin and Borsa 2000). Introductions have been documented in California (McDonald and Koehn 1988), Chile (Daguin and Borsa 2000; Westfall and Gardner 2010, 2013; Oyarzún et al. 2015; Larraín et al. 2018), South Africa (Grant and Cherry 1985; Branch and Steffani 2004), New Zealand and Australia (Westfall and Gardner 2010; Gardner et al. 2016), as well as parts of temperate East Asia (Japan, Wilkins et al. 1983; Hong Kong, Lee and Morton 1985; Korea, McDonald et al. 1991; Daguin and Borsa 2000). Invasive populations have been particularly problematic (and thus well documented) in intertidal communities throughout California and South Africa. Since its introduction in the early 1900s (Geller et al. 1994), *M. galloprovincialis* (Mediterranean lineage) has spread over 1100 km of the Californian coastline and populations have entirely displaced native smooth-shelled congener, *Mytilus trossulus,* from the southern portions of its natural distribution up to Monterey Bay, California (McDonald and Koehn 1988; Geller et al. 1994; Rawson et al. 1999; Geller 1999; Hilbish et al. 2010). In Africa, *M. galloprovincialis* (Atlantic lineage) has formed dense mussel beds throughout the southern and western coasts of South Africa and south Namibia (Branch and Steffani 2004), where it was accidentally introduced in the 1970s (Grant and Cherry 1985; Branch and Steffani 2004). Its present-day distribution in South Africa continues to expand annually and significant alterations to receiving marine communities have been documented, including local displacement of indigenous bivalves (*e.g., Aulacomya ater* and *Perna perna*) and damage to local habitats (Branch and Steffani 2004; Robinson et al. 2007).

In contrast, documenting the non-native distribution of *M. galloprovincialis* in the southern hemisphere has been complicated by low genetic differentiation and high levels of traditional marker discordance between introduced populations and morphologically similar endemic *Mytilus* lineages

(e.g., McDonald et al. 1991; Daguin and Borsa 2000, Hilbish et al. 2000; Gérard et al. 2008; Gardner et al. 2016; Oyarzún et al. 2015; Larraín et al. 2018). Representing a much larger proportion of genetic variation than previous surveys, transcriptome-wide investigations (Chapter 3) indicated that contemporary *M. galloprovincialis* introductions on the southeastern coast of Australia originate from genetically divergent Mediterranean and Atlantic native source lineages. Furthermore, analyses of introgression provided strong evidence that both introductions are associated with high rates of admixture with the endemic Australian *Mytilus planulatus* lineage. In the present study, I sampled two, novel introduced *M. galloprovincialis* populations from California and South Africa and combined these data with previously analysed introduced populations from Sydney Harbour and Batemans Bay, Australia (Chapter 3). Previous genetic surveys of allozymes (e.g., McDonald and Koehn 1988; McDonald et al. 1991; Sanjuan et al. 1997) and nuclear markers (*Glu 5'*, *mac-1*, Daguin and Borsa 2000) suggest independent origins of introduced *M. galloprovincialis* populations in California and South Africa: *M. galloprovincialis* from California are introduced from the Mediterranean Sea while individuals sampled in South Africa show genetic similarity to Atlantic *M. galloprovincialis* (Daguin and Borsa 2000). Consistently, another study based on 1337 Single Nucleotide Polymorphisms (SNPs; Saarman and Pogson 2015) between invasive *M. galloprovincialis* populations in California and six native Mediterranean individuals confirmed that introduced populations were likely derived from the Mediterranean native range. Combining these new data with two contact zones in Australia allowed me to compare the outcomes of genetic differentiation in introduced populations in Australia with replicated introductions in California and South Africa originating from the same native genetic sources. I accounted for genetic structure in the native range by sampling native *M. galloprovincialis* populations belonging to two divergent lineages separated by the Almeria-Oran front (between the Atlantic and Mediterranean; Fraïsse et al. 2015). I also included a third population east of the Siculo-Tunisian Strait, which divides the eastern and western Mediterranean basins (Fraïsse et al. 2015; Table 4-1; Figure 4-1)

Sample collection, RNA extraction and sequencing

Mussels were collected from wild populations from rocky intertidal or subtidal environments (Table 4-1). All individuals sampling in the northern hemisphere were initially genotyped for the species diagnostic marker *Glu-5'* (Rawson et al. 1996) to confirm species identity. To minimize the possibility of sampling hybrids, outgroup specimens belonging to the blue smooth-shell sister taxon *Mytilus edulis* were collected from a contemporary allopatric range. I also performed a Principal Component Analysis (PCA) of preliminary transcriptome-derived variants as an additional measure to exclude individuals introgressed with outgroup taxa (implemented in R; R Development Core

Team 2017). Assignment of Australian samples as native *M. planulatus* was based on mitochondrial genotyping and genome-wide analyses of species relationships performed in Chapter 3. Total RNA was extracted from 10-20 mg of mantle tissue (preserved in RNAlater) for a total of 10 new samples from South Africa and California (Table 4-1); samples were checked for quality and sequenced across three lanes of an Illumina Hiseq2000 or across a single lane of an Illumina Hiseq4000.

Reference transcriptome assembly: Additional filtering

I used the high-quality *M. galloprovincialis* reference transcriptome, assembled *de novo* in Chapter 3, for all downstream analyses (refer to *Chapter 3*: *RNAseq Data Processing and Reference Transcriptome Assembly* for details). The reference assembly consisted of 184,842 transcripts and was used for lncRNA identification (see *Identification of lncRNA transcripts*). Prior to population genomic analyses, I performed additional filtering of the reference assembly: Transcripts showing high sequence similarity were clustered using Cd-Hit-Est (Li and Godzik 2006; Fu et al. 2012) with a minimum sequence identity threshold of 95% of the shortest sequence. I removed contigs without a significant blastn hit to the *M. galloprovincialis* draft genome (e-value 10^{-4}; Murgarella et al. 2016). I also removed transcripts with significant blastn matches (e-value 10^{-3}) to the *M. galloprovincialis* male (Genbank reference: FJ890850.1) and female (Genbank reference: FJ890849.1) mitochondrial genomes. The resulting 143,100 nuclear sequences were used as a 'full' reference transcriptome assembly for Single Nucleotide Polymorphism (SNP) variant discovery.

Identification of lncRNA transcripts

Putative lncRNA transcripts were identified following the computational pipeline for lncRNA discovery from RNAseq data, as outlined in Gaiti et al. (2015). This bioinformatics pipeline involves stringent classification of transcripts as either protein-coding or non-protein coding; it follows stepwise filtering, annotation and isolation of only transcripts satisfying lncRNA criteria originally developed for the demosponge *Amphimedon queenslandica* (Gaiti et al. 2015). The *M. galloprovincialis* reference assembly was first queried against the Uniprot-Swissprot protein database (e-value 10^{-4}) to remove transcripts with significant blastx hits to known proteins; transcripts below a minimum length of 300 nucleotides were also removed. I used Transdecoder (Haas et al. 2013) to identify the longest open reading frame (ORF) for each transcript and retained only those with predicted ORFs shorter than 50 amino acids from start to stop codon. The resulting assembly was queried against SignalP and HMMER and to remove putative signal peptides and transcripts with significant homology matches to the Pfam protein database. All similarity comparisons were

conducted using an e-value 10^{-4}. Functional potential was further assessed using the Coding Potential Calculator (CPC; Kong et al. 2007) to generate an assembly of 46,872 putative lncRNA loci. Finally, I retained only lncRNA loci that had a significant blastn hit to the *M. galloprovincialis* draft genome (e-value 10^{-4}; Murgarella et al. 2016) to minimise the inclusion of erroneously assembled transcripts due to technical artifacts.

Mapping and identification of SNPs

I mapped RNAseq reads from four introduced and three native-range *M. galloprovincialis* populations and a single outgroup taxon (*M. edulis*) to the full and lncRNA reference assemblies (Table 4-1). I used Bowtie2 and the *very sensitive local* method (Table 4-1; Langmead and Salzberg 2012). PCR duplicates were marked and removed using Picard MarkDuplicates (http://picard.sourceforage.net) and Samtools was used to convert SAM files to indexed BAM files. SNPs were called using Freebayes (https://github.com/ekg/freebayes), and variant filtering was performed in VCFtools (Danecek et al. 2011). Variant sites below a minimum genotype quality threshold of 30 and a minimum mean depth coverage of 10 reads were excluded. I also removed singletons and indel variants from all datasets. Additional filtering of variants was applied to individual datasets prior to statistical analyses as outlined in each section below.

Analyses of overall population structure

I visualised variation among individual genotypes using PCA on SNPs derived from both the full transcriptome assembly and the lncRNA dataset. Because I previously determined that introduced *M. galloprovincialis* populations from Sydney Harbour and Batemans Bay are introgressed with the native Australian taxon, *M. planulatus* (Chapter 3), I also included samples from Tasmania as a reference population. For these analyses, I removed all variants with missing genotype data, which resulted in 16,779 biallelic SNPs for the full dataset and 471 biallelic SNPs for the lncRNA dataset. I then calculated pairwise F_{ST} (Weir and Cockerham 1984) using the R package *hierfstat* (Goudet 2005) to explore the contributions of each dataset to overall population genetic structure among native and introduced populations. I assessed whether observed values were significantly different from zero for both datasets using the *boot.ppfst* function in *hierfstat*. I applied 10,000 bootstrap replicates and calculated the 2.5% and 97.5% confidence intervals of the resulting distribution to determine significance at $P \leq 0.05$.

To assess whether differences in F_{ST} were significantly different between the full transcriptome and lncRNA datasets (rather than an outcome of the drastically smaller size of the lncRNA SNP dataset), I randomly resampled 471 variants (i.e. number of SNPs equivalent to the lncRNA dataset) from the full dataset 1000 times and calculated pairwise F_{ST} for each iteration to obtain a null distribution of estimates. I then assessed whether the observed lncRNA F_{ST} values fell within the 5th or 95th quantiles of the null distribution to determine whether the lncRNA variants contributed significantly to elevated or reduced population differentiation compared to the random subsets at a p-value threshold of $P \leq 0.05$. I used the '*qvalue*' R package to correct p-values for multiple population comparisons at qvalue≤0.05. Because F_{ST} differentiation estimates are influenced by high within population genetic diversity (Jost 2008; Cruickshank and Hahn 2014), and assume that populations are in mutation-drift equilibrium (Wright 1951), I also calculated the total variance explained across all individual genotypes for 1000 resampled datasets using principal components. I performed PCA on each iteration to obtain the proportion of variance explained across the first and second PC axes. I inferred the difference in the proportion of variance between datasets as statistically significant if the values for PC1, PC2 or the sum of PC1 and PC2 fell within the 5th or 95th quantiles of the null distribution.

Assessment of introduced population independence

Prior to genomic outlier detection, I investigated whether introduced populations were established independently from the *M. galloprovincialis* native range or whether populations are likely to represent stepping stone introductions between introduced ranges. For subsequent analyses (ADMIXTURE and TreeMix), I removed SNPs with a minor allele frequency of less than 5%, but retained positions with up to 20% missing data. I estimated individual ancestry proportions with the program ADMIXTURE (Alexander et al. 2009); I included samples from Tasmania as a reference population and used *M. edulis* as an outgroup taxon. I used VCFtools and PLINK v1.90 (Purcell et al. 2007) to convert the filtered VCF output to BED format files as input, which reduced the original dataset to 37,784 biallelic SNPs across 3946 contigs. I ran ADMIXTURE with 100 iterations and used the cross-validation procedure with 50 replicates for K=1 to K=7 genetic clusters.

To test the hypothesis of non-independence (i.e. gene flow) between introduced *M. galloprovincialis* populations, I used TreeMix v 1.12 (Pickrell and Pritchard 2012) to estimate the strength (branch weight = *w*) and direction of migration between populations while accounting for relationships between groups. TreeMix uses allele frequency correlations between populations to infer a maximum likelihood population tree of the phylogenetic relationships between groups. Migration edges are subsequently added between population branches to determine whether incorporating admixture

events of varying weights and directions improve the likelihood of the population tree given the genetic data. For TreeMix analyses, I used Atlantic *M. galloprovincialis* as an outgroup to focus inferences on admixture events between *M. galloprovincialis* populations only. I accounted for linkage disequilibrium by performing analyses on windows of 100 variants. I calculated the standard error of migration events with the –se option without sample size correction (option –noss). I examined the residual plot of pairwise population genetic covariances to infer the possibility of gene flow between populations, where negative residual standard error values suggest that populations are more closely related to each other (given the data) relative to the population tree with no migration events. I then modeled 1-10 migration events sequentially to see whether adding migration edges to the phylogeny improved the likelihood fit to the data. I used a stepwise comparison Akaike Information Criterion (AIC) between sequential migration models to determine whether additional migration edges significantly improved the likelihood of the population tree relative to a nested model with n-1 migration edges. I calculated AIC values as (-2(ln(likelihood)) + 2K), where K is the number of free parameters in the model (i.e. K=2; the number of migration edges and the direction of gene flow). I did not consider additional migration events when the difference between nested models was less than two ($\Delta AIC < 2$). The tree graph and residual plots for each analysis were visualised using R.

I used the three population (f_3) test of admixture (Keinan et al. 2007; Reich et al. 2009) to verify evidence of migration inferred by TreeMix (Pickrell and Pritchard 2012). I estimated the f_3 statistic with the *threepop* function implemented in TreeMix. The f_3 statistic estimates whether differences in allele frequencies between each population combination deviate more than expected due to incomplete lineage sorting, thereby suggesting recent admixture. Significant migration is inferred if the f_3 statistic has a negative value and a z-score of ≤ -3.8 (equivalent to a P-value< 0.0001), which is determined through a jackknifing procedure over windows of 100 SNPs.

Genomic outlier detection

I used the full transcriptome dataset to identify putative loci under selection following introduction. For this analysis, I first divided the full dataset into each introduced vs. putative source population pair, then removed columns with more than 20% missing data in the resulting VCF file. Given the low differentiation between the eastern and western Mediterranean populations, I combined these populations for all relevant analyses. To detect genomic outliers showing statistically greater differentiation relative to the average genomic background, I used the 'pcadapt' package in R (Luu et al. 2017). The pcadapt method uses individual genotype distances (without prior information about

the population groupings or the model of population structure) to create a model-free null distribution of genetic variation against which outlier loci are identified (Duforet-Frebourg et al. 2014; 2015). This method therefore takes into account existing population structure among groups, controlling for both shared population history and potential genome-wide effects of genetic drift experienced during the introduction of invasive populations (Sherman et al. 2016).

PCA was performed over a pairwise (introduced versus source) population SNP matrix; for each analysis, I retained only the first two PC axes to maximise the variance explained by the underlying population structure. I also applied a minimum allele frequency of 5% and a 'SNP thinning' interval of 200 SNPs (default $r^2 = 0.1$) to account for physical linkage if multiple variants are found nearby on the same contig. Each SNP was regressed onto the retained PCs and assigned a z-score proportional to the strength of the correlation between the SNP to each PC axis. Z-scores were converted to a Mahalanobis distance statistic for each variant; p-values were computed based on the scaled Mahalanobis distance which should follow a chi-squared distribution representative of the genome-wide null distribution of population differentiation. Individual variants with the largest scores (i.e. variants disproportionately contributing to population structure, relative to the null expectation) were considered as putative outlier loci. I converted p-values to q-values using the '*qvalue*' R package (Dabney et al. 2010). SNPs with q-values below 0.01 were classified as genomic outliers. I compared significant outliers between populations pairs derived from same native genetic background to assess if populations (i) shared outlier SNP loci; (ii) shared contigs containing at least one outlier SNPs (i.e. outlier contigs); and (iii) if shared outlier contigs were identified as putative lncRNA transcripts.

Nucleotide diversity

To test the hypothesis that shared outlier contigs have experienced selective sweeps following introduction (rather than selection within the native source population), I compared within population nucleotide diversity for each paired dataset (which I also analysed for genomic outliers). I calculated nucleotide diversity based on equation 12.66 from Nei and Kumar 2000 for each introduced (*di*) and corresponding source (*ds*) population. Because introduced populations are likely to experience genome-wide genetic drift, I expected that joint nucleotide diversity estimates for most contigs would fall along the slope of the linear regression between *di* and *ds*. I estimated the slope of the joint distribution of *di* and *ds* and tested if it was significantly different from 1 using the SMATR R package (Warton et al. 2012). If shared outlier contigs have experienced parallel selective sweeps in independent introduced populations, I would expect shared outlier contigs to exhibit reduced gene-

wide nucleotide diversity relative to the source population (i.e. negative residual values) in both population comparisons.

For each paired population subset, genotypes were statistically phased using the program beagle v4.1 (Browning and Browning 2007). I generated a consensus sequence for each individual haplotype using the corresponding VCF file and reference sequences in BCFtools v1.3.1. Within population nucleotide diversity was calculated using custom scripts in R (available at https://github.com/dinmatias). Incomplete and unordered scaffolds in the *M. galloprovincialis* draft genome assembly of the (Murgarella et al. 2016), however, precluded inferences regarding the genomic positions and the possibility of linkage between shared outlier contigs.

Results

Contributions to genetic differentiation: Full and lncRNA datasets

Principal component analysis of 16,779 biallelic variants derived from the full transcriptome assembly revealed genetic separation between introduced populations in California and South Africa (Figure 4-1). The first and second PC axes explained 6.54% (PC1) and 5.72% (PC2) variance among individual genotypes (Figure 4-1). Genetic similarities were evident between introduced *M. galloprovincialis* in California and native Mediterranean populations; individuals sampled in South Africa clustered closely with Atlantic mussels, confirming previous inferences that these introductions originate predominantly from Mediterranean and Atlantic source lineages respectively. This analysis also confirmed genetic structure previously identified between two introduced populations sampled in Australia (Chapter 3).

Applying a conservative pipeline developed for lncRNA identification (Gaiti et al. 2015), I predicted 44,096 putative lncRNA loci, representing 23.8% of the full *de novo* assembled *M. galloprovincialis* transcriptome. This value is greater than the proportion of putative lncRNAs identified in the *M. galloprovincialis* digestive gland transcriptome (Gerdol et al. 2014), in which only 14.6% of transcripts were identified as lncRNA candidates. Despite stringent filtering, the final lncRNA assembly in the present study may include untranslated regions (UTRs), other RNA molecules belonging to polyadenylated RNA transcripts from other major RNA families (i.e. small regulatory RNAs, tRNAs and rRNAs), assembly artifacts or novel *Mytilus* coding peptides with no similarity matched to protein databases. Variant identification and filtering resulting in only 200 transcripts with 471 high quality SNPs; lncRNAs have been reported to be expressed at lower levels relative to coding

transcripts (Quinn and Chang 2016), which may have subsequently resulted in the removal of many variants due to low sequence coverage.

The PCA of 471 SNPs derived from the lncRNA dataset explained a lower proportion of variance across the first (5.92%) and second (5.10%) PC axes. For this dataset, only *M. planulatus* sampled in Tasmania and Batemans Bay showed genetic separation from all other populations which did not reflect any clustering based on sampling region (Figure 4-1). The proportion of variance explained by the lncRNA dataset was significantly lower than expected based on 1000 random subsamples from the full transcriptome dataset for PC1 ($P<0.0001$), PC2 ($P=0.002$) and their cumulative variance ($P=0.031$) (Figure S4-1). Pairwise comparisons of population differentiation for full dataset indicated F_{ST} values significantly different than zero for all population pairs except for comparisons between the eastern and western Mediterranean and California (Table 4-2). Compared to the full dataset, pairwise comparisons for the lncRNA dataset returned lower F_{ST} values for all populations pairs, with the exception of California and the eastern or western Mediterranean (Table 4-2). Reductions in lncRNA F_{ST} were significant in some population pairs when compared the null distribution corresponding to 1000 randomly chosen SNP subsets from the full dataset (Table 4-2). Significant reductions in population structure in the lncRNA dataset, however, were only evident for population comparisons displaying high F_{ST} values in the full transcriptome dataset (i.e. population comparisons with Tasmania).

Analyses of population independence

Low population structure between native and introduced populations in the northern hemisphere presented several difficulties for validating independent introductions into South Africa, California and Australia. ADMIXTURE analysis were able to discriminate between divergent Mediterranean and Atlantic *M. galloprovincialis* populations (K=4; Figure 4-2); however, analyses with greater numbers of genetic groups (K5-7; Figure S4-2) could not differentiate between related native and introduced populations. Specifically, differentiated clusters were not recovered for Atlantic and South African populations, or populations from the Mediterranean and California.

TreeMix did not recover strong evidence suggesting migration between introduced populations. I observed high pairwise residual covariance in allele frequencies between Batemans Bay and Atlantic *M. galloprovincialis* (Figure 4-3A), confirming previous signatures of admixture between these populations (Chapter 3). However, allele frequency residuals are inferred assuming that the maximum likelihood population tree (with no migration) accurately represents the phylogenetic relationships

between groups. Consistent grouping of Sydney Harbour with Mediterranean populations in the base topology resulted in high allelic covariance between Sydney Harbour and Tasmania (than expected by the base tree; Figure 4-3A); this result is consistent with introgression between these populations (Chapter 3). I found that the addition of five migration edges provided the highest improvement in the likelihood fit of the population tree (Figure 4-3B), compared to nested models with fewer migration edges (ΔAIC > 2). The strongest migration edges were evident between the Atlantic and Batemans Bay (w=41%; P<<0.001). Migration edges connecting Sydney Harbour with both Tasmania and the eastern Mediterranean were also significant (w=24-48%; P<<0.001). Weaker evidence of gene flow from South Africa into the western Mediterranean (w=3%; P>0.001) likely reflects allele sharing with Mediterranean *M. galloprovincialis* through introgression with Atlantic source populations prior to their introduction into South Africa (e.g., Fraïsse et al. 2015). Mixed Atlantic-Mediterranean ancestry in South African samples was also evidenced in ADMIXTURE analyses (i.e. K=4; Figure 4-2). TreeMix analyses suggested weak migration from South Africa into California, indicating low levels of shared variation between these populations; however, this migration edge was not statistically significant (P=0.137; Figure 4-3B).

A key consideration for these analyses, however, is that the close grouping of Sydney Harbour with California (Figure 4-3B; but see Figure 4-3A) suggests a close genetic relationship between these populations. Accounting for this genetic similarity in the base maximum likelihood tree precludes the ability to estimate whether adding a migration edge would account further for allelic covariance between these populations, thereby supporting a stepping stone introduction into Sydney Harbour via California.

Without substantial structure between native and introduced populations (i.e. California and Mediterranean populations) it is difficult to infer the exact sources introduced genetic variation in Australian populations. Indeed, several three population tests of population combinations involving two introduced populations (i.e. one as admixed and the other as a putative ancestral population) had significantly negative values (P < 0.0001; Table 4-3; Figure 4-4). I found that the most significant f_3 statistics (in which the z-score fell below the lowest 5% of the distribution of all values; z-score < -20.25) were associated with population combinations consistent with evidence of admixture in focal populations (Figure 4-4; Table 4-3; PopA: Sydney Harbour, Batemans Bay and South Africa). However, for Sydney Harbour there was no discrimination in whether the second ancestral population (Table 4-3, PopC) was California, Mediterranean-East or Mediterranean-West and all combinations were significant. Moreover, analyses with Batemans Bay as the focal admixed population (PopA)

returned highly significant values for all potential northern hemisphere sources (i.e. South Africa, Atlantic, California, Mediterranean-East or Mediterranean-West).

Repeatability of post-introduction evolution

To evaluate the degree of parallel evolution among introduced populations with the same native genetic origins, I searched for common genomic outliers between four introduced-source population pairs (Figure 4-5). Using a conservative qvalue threshold of 0.01, no outlier SNPs were shared among populations pair comparisons. For introduced populations derived predominantly from Atlantic *M. galloprovincialis* (South Africa and Batemans Bay), I identified 24 shared outlier contigs (i.e. containing a least one SNPs classified as an outlier variant; Figure 4-5). For introduced populations derived from Mediterranean *M. galloprovincialis* (California and Sydney Harbour), I identified four shared outlier contigs, suggesting common evolutionary responses in both introduced populations. I did not identify any lncRNA transcripts as shared outlier contigs in paired comparisons (Figure 4-5); however, five lncRNA transcripts had a single outlier SNP between the Atlantic and South Africa, and two lncRNA transcripts had an outlier SNP between the Mediterranean and Sydney Harbour. Additionally, there was no overlap among shared outlier contigs between comparisons with different source populations using a qvalue threshold of 0.01.

To investigate whether shared outlier contigs have experienced selective sweeps following introduction (rather than selection within the source population), I tested the hypothesis that shared outlier contigs in both introduced populations would display reduced nucleotide diversity (*di*) relative to their respective source (*ds*) population. Linear regressions of the joint diversity distribution were significant ($P<<0.0001$) for all population combinations (Figure 4-6). The slope of each regression was significantly different from 1 all population pairs ($P<0.01$) expect for the Atlantic and Batemans Bay population pair ($P=0.794$). Out of 24 shared outlier contigs, 8 showed negative residual values indicating reduced diversity in both South Africa and Batemans Bay, compared to the Atlantic *M. galloprovincialis* population (Figure 4-6). These genes had diverse functions, including stress response, protein transport, but there were no evident functional patterns among candidates (Table 4-4). Shared outlier contigs did not display reduced diversity in both Sydney Harbour and California compared to the Mediterranean *M. galloprovincialis* population (Figure 4-6). More broadly, I did not identify specific genes or functional groups previously identified as candidates for thermal adaptation in *M. galloprovincialis* (i.e. Chapter 2) among shared outlier contigs; however, two outlier SNPs were present in candidate genes Dipeptidyl peptidase (cathepsin C) in South Africa, (qvalue=0.01), and small heat shock protein 25 of the alpha-crystallin family in Sydney Harbour (qvalue=0.1) suggesting

that loci under selection for species divergence (i.e. between *Mytilus* congeners; Chapter 2) may also experience intraspecific selection following introduction in some populations.

Discussion

Multiple worldwide introductions of marine invasive species provide exceptional opportunities to study the role of selection and adaptive responses in species introductions (Rius et al. 2015a). Such data can shed light on shared genetic differentiation underlying post-introduction evolution and the genomic traits contributing to invasive success. In this study, I tested whether introduced populations of the marine mussel *M. galloprovincialis* display parallel genetic differentiation relative to their native genetic backgrounds. I also identified major non-coding elements of the transcriptome to estimate their relative contributions to population structure and parallel genetic differentiation. Comparisons of transcriptome-wide variation between pairs of introduced populations (originating from the same source lineage) revealed evidence for parallel genetic differentiation at the level of individual contigs; however, I did not find repeated genetic changes at individual polymorphisms between introduced populations. There was no consistent reduction in gene-wide nucleotide diversity at outlier contigs shared between introduced populations, suggesting that repeatable patterns of genetic differentiation may be influenced by both selective and non-selective evolutionary processes during introduction. However, because genetic structure between northern hemisphere populations was not significantly different than zero, I could not exclude shared demographic introduction histories as causative factors underlying genetic commonalities between northern and southern hemisphere introduced populations. Analyses of variation from the all expressed transcripts and only the lncRNA dataset indicated a significantly lower contribution of these non-coding transcripts to intraspecific genetic variation. This result suggests that lncRNAs likely evolve under purifying selection or other selective constraints in a lineage specific manner and should be considered as largely non-neutral components of expressed genetic variation in transcriptome-based surveys of differentiation. Consistent with this hypothesis, lncRNA transcripts were not present among shared outlier contigs, pointing to a dominant contribution of protein-coding loci to replicated signatures of differentiation between introductions.

Repeatable patterns of post-introduction evolution

The results of this Chapter demonstrate that independent introductions can lead to repeatable patterns of genetic differentiation, despite high levels of standing genetic diversity predicted for introduced marine populations (Gagnaire et al. 2015, 2018). Outlier analyses of strongly differentiated loci in

four introduced *M. galloprovincialis* populations revealed that the same contigs were repeatedly involved in high levels of differentiation between introduced populations and their relative native genetic backgrounds. For introduced populations derived predominantly from Atlantic *M. galloprovincialis* (South Africa and Batemans Bay), I identified 24 shared outlier contigs containing at least one SNP classified as an outlier variant (Figure 4-5). While there were a number of genomic outliers in both California and Sydney Harbour (derived from Mediterranean *M. galloprovincialis*) independently, there was little overlap in outlier contigs when these populations were compared. Consistently, I found no evidence for genetic parallelism at individual SNPs among both sets of population comparisons (Figure 4-5). A lack of substantial genetic differentiation between native and introduced populations in the northern hemisphere, however, precluded my ability to robustly rule out the possibility that geographically independent populations have acquired differentiated alleles through gene flow between introduced ranges (i.e. stepping stone introduction routes from South Africa and California into Australia). However, because shared outlier contigs were associated with different SNP variants in different populations, it is less likely that shared patterns of differentiation are a result of common introduction histories. Overall, these findings suggest that while repeated changes at the same nucleotide substitutions are unlikely, common genetic responses to introduction may occur at higher genomic levels in *M. galloprovincialis*; although the extent of repeatability may vary between introduced populations.

While there are many examples of replicated evolution in natural populations (reviewed in Arendt and Reznick 2008; Elmer and Meyer 2011; Stern 2013; Conte et al. 2012), the findings in this Chapter are concordant with few studies examining the repeatability of genetic changes across non-native lineages. Evidence for common evolutionary shifts among introduced populations has been largely limited to weedy and invasive plant species (e.g., Prentis et al. 2008; Lai et al. 2008; Hodgins et al. 2015; Colautti and Lau 2015). These investigations, however, have yielded inconsistent conclusions regarding the frequency of shared genetic mechanisms underlying the evolution of similar phenotypes associated with invasions. For example, Hodgins et al. (2014) found little evidence that positive selection has acted on the same amino acids or protein-coding genes across six invasive lineages of flowering plants, even among loci previously associated with independent domestication events in introduced lineages (Kane et al. 2011). Evidence for adaptive genetic changes was also inconsistent when individual native and introduced genotypes were compared, although positive selection on similar gene functional groups across species suggested that selective pressures in introduced environments may broadly operate on common traits through divergent molecular pathways (Elmer and Meyer 2011; Hodgins et al. 2014). In the present study, I found cases of shared differentiation in the same individual contigs, however, shared genes had no obvious functional consistency (Table 4-

4). Taken together, these findings suggest that the genomic scale of repeated evolution during introduction may vary across divergent non-native taxa (i.e. terrestrial and marine) that are likely to also differ in effective populations sizes and the amount of standing variation that is available for adaptation following introduction.

Genomic regions where multiple introduced populations share strong deviations in differentiation may underlie important sources of variation involved in adaptive responses in introduced environments (Jones et al. 2012). To investigate whether observed patterns of repeated differentiation may be the result of selective pressures, I predicted that shared outlier contigs in pairs of introduced populations would display reduced nucleotide diversity relative to their respective source populations. However, I found only modest evidence for repeated gene-wide selective sweeps in shared outliers contigs: Comparisons of nucleotide diversity between introduced populations derived from Atlantic *M. galloprovincialis* (South Africa and Batemans Bay) indicated relative reductions in gene-wide diversity in only a fraction of transcripts (8 out of 24) identified as shared outlier contigs (Figure 4-6). Similarly, I did not identify any consistent diversity reductions among introduced populations derived from Mediterranean *M. galloprovincialis* (California and Sydney Harbour). These findings suggest shared deviations in differentiation outliers may be driven by selection within both native range and introduced populations. This interpretation is contingent, however, on the assumption that pronounced reductions in gene-wide variation (that are indicative of selection) are a result of classical hard selective sweeps experienced by focal loci (Bierne et al. 2013; Sherman et al. 2016). Common patterns of reduced diversity may not be evident if selective sweeps at shared outlier contigs are 'soft' and produce only small changes in intragenic allele frequencies between populations. For example, soft sweep patterns may be expected if (i) novel beneficial mutations arise in parallel in introduced populations but on different haplotypes, or if (ii) different variants (either new mutations or standing genetic variants) are favoured within the same gene or genomic region (reviewed in Messer and Petrov 2013). Because selection on standing genetic variation is predicted to be the most likely adaptive mechanism by which non-native populations rapidly evolve (Prentis et al. 2008; Bock et al. 2015), I posit that the latter scenario may explain parallel differentiation in response to post-introduction selection despite inconsistent selective sweep patterns among shared outlier contigs.

Several other molecular or demographic processes that are not necessarily related to natural selection may determine genetic diversity within introduced and native ranges and may additionally lead to shared signals of reduced variation in introduced populations. Genetic repeatability can be highly dependent on evolutionary constraints associated with species-specific genetic architecture (Jones et al. 2012; Renaut et al. 2014). For example, the highest rates of repeatable evolution were observed

across large chromosomal regions among sunflowers species, compared to individual polymorphisms, suggesting that genomic traits, namely mutation rate and recombination, largely determine which genes will respond to selection (Lynch 2007; Renuat et al. 2014). Importantly, because genomic architecture is most likely to be conserved between closely related lineages (Conte et al. 2012), these features can lead to predictable evolutionary outcomes independently of selective pressures implying that in some cases, the position of a gene may be equally as relevant as gene functionality in predicting patterns of differentiation across a genome (Renuat et al. 2014; Cruickshank & Hahn 2014). Such genomic features may be particularly relevant for invasive species research, as introduced populations are expected to experience a wide range of selective pressures in introduced environments that are imposed by variable abiotic conditions (i.e. large latitudinal gradients), diverse community assemblages and the presence of closely related native lineages (Riquet et al. 2013; Guzinski et al. 2018; Chapter 3). However, validating how genomic architecture imposes evolutionary constraints on post-introduction evolution and whether some genomic features may enhance opportunities for adaptation in non-native species will require additional theoretical and empirical investigations to test these focused hypotheses (Chown et al. 2015).

Interpretations of genomic outliers can additionally be improved by accounting for the influences post-introduction admixture with closely related lineages in generating signals of differentiation associated with multiple introductions (Fraïsse et al. 2015 Viard et al. 2016; Rougemont and Bernatchez 2018). Strong evidence of introgression from endemic Australian *M. planulatus* into populations sampled in Sydney Harbour and Batemans Bay (Chapter 3) has contributed to stronger differentiation between these introduced populations and native source populations, compared to northern hemisphere introduced populations (i.e. Sydney Harbour and Mediterranean populations display F_{ST} values significantly different than zero, compared to no differentiation between populations from California and the eastern or western Mediterranean). In the northern hemisphere, local introgression from *M. edulis* into Atlantic *M. galloprovincialis* contributes significantly to differentiation at intraspecific outliers between *M. galloprovincialis* populations from the Atlantic and Mediterranean, and between Mediterranean populations from the western and eastern basins (Fraïsse et al. 2015). Furthermore, the same loci were strongly differentiated in some populations, suggesting that adaptive introgression from outgroup taxa can promote repeated patterns of intraspecific differentiation in specific areas of the genome (Fraïsse et al. 2015). While there is evidence for weak introgression between *M. galloprovincialis* and *M. trossulus* at the northern range boundary where their ranges overlap in California (Rawson et al. 1999; Saarman and Pogson 2015), individuals sampled in the present study did not display outstanding differentiation indicative of introgression with *M. trossulus*. It is therefore unlikely that shared outlier contigs between Sydney

Harbour and California are a result of similar patterns of interspecific introgression into the introduced genetic backgrounds. Similarly, replicated differentiation due to interspecific introgression would not be expected for population pairs with Atlantic *M. galloprovincialis* origins, as there are no *Mytilus* congeners native to South Africa.

An additional consideration is that highly differentiated loci in introduced *M. galloprovincialis* populations in this study are likely to include introgressed variants from divergent native range *M. galloprovincialis* lineages. The large number of SNP outliers showing extreme differentiation in South Africa relative to Atlantic *M. galloprovincialis* likely represent introgressed genetic variation originating from the Mediterranean native range in at least some individuals. Specifically, admixture analyses revealed mixed ancestry proportions in two individuals sampled in Cape Columbine, South Africa while three samples from Yzerfontein (less than 90 km south) indicated genetic clustering with only Atlantic genotypes (Figure 4-2). Evidence to date, based on few size polymorphic markers, suggests that South African populations are exclusively derived from Atlantic *M. galloprovincialis* (e.g., Daguin and Borsa 2000). The possibility that more recent secondary introductions of Mediterranean *M. galloprovincialis* have occurred in this region is not an unlikely scenario. Interestingly, all individuals from South Africa showed genetic separation from Atlantic samples, suggesting that significant differentiation has occurred despite recent introductions (<50 years) into this region (Figure 4-1); however, whether intraspecific admixture of alleles from divergent parts of the native range has implications for increased fitness in introduced *M. galloprovincialis* populations remains to be explored and represents a significant knowledge gap in marine invasive species research more broadly (Rius and Darling 2014).

In the present study, the main objective was to quantify the degree to which outlier loci are shared between introduced populations rather than to identify highly differentiation loci most generally. While the strength of a paired population approach allowed me to examine a large number of loci and to identify regions of the genome undergoing parallel differentiation between introduced populations, several additional caveats in my analyses should be considered. First, although outliers of differentiation were associated with different polymorphisms in different populations in my analyses, shared patterns of differentiation between northern and southern hemisphere introduced populations owing to shared introduction histories could not be excluded. Specifically, I could not distinguish between the sources introduced genetic variation from South Africa into Batemans Bay and from California into Sydney Harbour. Evidence of admixture across all possible population combinations including Batemans Bay and Sydney Harbour as focal admixed populations in three population tests returned significant signals of admixture with multiple or all potential northern hemisphere

populations as putative sources of gene flow, such that stepping stone introduction histories could not be fully rejected. Determining the native origins of introduced populations appears to be a ubiquitous issue in marine invasion studies (e.g., Riquet et al 2013), and raises significant challenges for interpreting signals of differentiation associated with introductions (Geller et al 2010; Viard et al 2016). *M. galloprovincialis* stepping stone introductions into the southern hemisphere may have involved multiple translocations through human objects, such as offshore oil rigs migrating between Australia, New Zealand and South Africa (Gardner et al. 2016). Bayesian statistical approaches, such as Approximate Bayesian Computations may provide additional power to leverage signatures of intraspecific differentiation to reconstruct and test hypotheses about likely routes of introduction (i.e., Estoup and Guillemaud 2010; Lombaert et al. 2011; Barker et al. 2017).

A second potential caveat is that inconsistent patterns of reduced nucleotide diversity among shared outlier contigs could be due to high variance in gene-specific diversity estimates when small sample sizes are used. Small samples sizes within individual populations also precluded explicit inferences regarding the importance of novel mutations versus standing genetic variation to observed shifts in allele frequencies (Elmer and Meyer 2011), as alleles absent in native populations may simply be unsampled low frequency variants. The results of this analysis should therefore be compared against results from other outlier detection methods that do not make specific assumptions about demographic models (e.g., OutFLANK, Whitlock and Lotterhos 2015; Lotterhos and Whitlock 2015), or scans for shared selective sweeps (reviewed in Vatsiou et al. 2016) to validate commonalities in differentiation. On the other hand, it is likely that outlier scans for highly differentiated loci may also fail to detect commonalities in genetic responses if selection in the introduced environment is weak or if evolution occurs in polygenic traits controlled by large numbers of small effect loci (Le Corre and Kremer 2012; Gagnaire et al. 2015). Because adaptation in invasion is likely to involve subtle allele frequencies shifts at a large number of loci underlying complex traits under selection (Dlugosch et al. 2015), such weaker signatures of differentiation may be overlooked by genome scans searching for strong signals of locus-specific differentiation (Bourret et al. 2014; Stephan 2016; Bierne et al. 2016).

Consistent with this premise, genome scans investigating differentiation between introduced and native populations in the marine environment have not recovered strong evidence for adaptation post-introduction (e.g., Riquet et al. 2013; Rohfritsch et al. 2013; reviewed in Tepolt 2015; Viard et al. 2016). For instance, a genetic survey of size polymorphic markers in the slipper limpet, *Crepidula fornicata,* identified a small number of outliers confined to native range populations, but found no evidence for strong differentiation between invasive and native populations (Riquet et al. 2013). In another example, Rohfritsch et al. (2013) identified shared genetic outliers in two invasive

populations of the Pacific oyster *Crassostrea gigas*, but it could not be determined whether the observed patterns reflected parallel responses to selection or genetic differentiation resulting from shared demographic histories of introduction. Most recently, Guzinski et al. (2018) did not recover evidence for local adaptation among several introduced populations of the Pacific kelp, *Undaria pinnatifida* from both natural and cultivated populations. In a notable exception, however, Bernardi et al. (2016) found 46 outlier loci between native and colonising populations of the bluespotted cornet fish, *Fistularia commerconii*, across a salinity gradient in the Mediterranean Sea, suggesting that rapid evolution of large-effect loci implicated in salinity adaptation has occurred since the initial invasions into the Mediterranean. In another comprehensive study, Tepolt and Palumbi (2015) compared genetic differentiation within and between native and introduced populations of the of invasive European green crab, *Carcinus maenas*. Comparing a SNP panel derived from a complete transcriptome assembly against a putatively neutral subset of variants they demonstrated reduced F_{ST} differentiation when putatively selected variants were removed. While these findings point to an important role for selective processes in shaping population structure during invasion, the genomic regions underlying population genetic structure or the sources of functional variation underlying outliers of differentiation in introduced populations were not investigated. For introduced *M. galloprovincialis,* however, warm-temperature tolerance appears to have facilitated initial invasions into California (Lockwood and Somero 2011a; Lockwood et al. 2015). Thus, a lack of consistent patterns indicating selection at shared outlier loci in the present study may suggest that preadaptation is the predominant evolutionary mechanism explaining successful *M. galloprovincialis* introductions globally. Evidence supporting alternative post-introduction processes, such as introgression with native congeners (i.e., Chapter 3), further suggests that successful introduced populations are likely to evolve under a combination of pre- and post-introduction evolutionary processes.

Contributions to genetic differentiation: Long non-coding RNAs

Clarifying the major components of genetic variation underlying intraspecific differentiation allowed me to explore the relative contributions of non-coding elements of the transcriptome on patterns of differentiation between native and introduced populations. Analyses of variation revealed that lncRNA transcripts contributed significantly less to intraspecific variation among individual genotypes compared to variants derived from the full transcriptome (Figure 4-1). Primary lncRNA sequences also showed reduced differentiation between populations, more so than expected by the background level of genomic divergence across a random subset of markers; however, significant reductions in population structure were only apparent for population pairs showing strong differentiation (i.e. high F_{ST} values) across all expressed loci (Table 4-2). Despite notable progress in

understanding the role of lncRNAs molecules in development and interspecific diversity (Gaiti et al. 2018), inconsistent patterns in both sequence conservation and rapid evolution among lineages has reinforced major gaps in our understanding of lncRNA evolution. It has been shown in previous studies that lncRNAs frequently evolve faster than protein coding sequences, as the primary sequence is not constrained by protein function (Pang et al. 2006; Hezroni et al. 2015; Quinn and Chang 2016; Ulitsky 2016). Instead, selective constraints appear to be limited only to genomic architecture (i.e. genomic position; Hezroni et al. 2015), and short promoter regions and motifs underlying secondary structure (Washietl et al. 2014). The main feature of lncRNA transcripts in this study, however, is constrained evolution through either purifying selection or other selective constraints that do not appear to operate in a population-specific way, or at least not within the timescale relevant for biological invasions.

Consistent with this hypothesis, lncRNA transcripts were not present among shared outlier contigs, suggesting that protein-coding transcripts are a primary driver of genetic differentiation within introduced ranges. These findings also raise important considerations that large numbers of expressed transcripts likely evolve under constrained evolution. Large non-neutral components of expressed genetic variation may effectively mimic population demographic effects on some parts of the genome. Not accounting for these conserved elements may lead to underestimates of the mean neutral genomic expectations, which may confound interpretations of neutral population structure from transcriptome scans (e.g., Tepolt and Palumbi 2015), and thus, the contributions of selection outliers to differentiation; however, this hypothesis has not been explicitly tested in the present study. Purifying selection has also been shown to act on intraspecific lncRNA polymorphism in *Drosophila* (Haerty and Ponting 2013), which is consistent with suspected lncRNA functions in conserved regulatory networks involved in development (Perry and Ulitsky 2016; Gaiti et al. 2018b). While most molecular mechanisms underlying these functions remain unclear (Mattick and Makunin 2006; Quinn and Chang 2016), it is evident that selection on functional properties of lncRNAs may promote rapid evolution in some lineages. Non-coding RNAs have been identified as targets of positive selection in human populations (e.g., Pollard et al. 2006) and have been implicated in local adaptation during historical migrations out of Africa into Europe and Asia (Duforet-Frebourg et al. 2015). Secondary and tertiary RNA structures have also been associated with biochemical adaptations to temperature (de la Fuente et al. 2012; Somero 2018), as well as host pathogen coevolution (Amaral et al. 2013), suggesting that lncRNA molecules may experience selective pressures on ecological timescales.

For invasive species research, the regulatory role of expressed non-coding RNA transcripts and the non-coding genome most generally, is largely unexplored, despite growing evidence that selection on

heritable regulatory changes play pivotal roles in adaptation in invasive species (Pérez et al. 2006; reviewed in Sherman et al. 2016; Wellband and Heath 2017). Parallel shifts in gene expression in weedy plants (Lai et al. 2008) and the evolution of phenotypic plasticity in both terrestrial and marine animals have been shown to facilitate the spread of introduced and colonising populations, with notable examples documented in cane toads (*Rhinella marina*) (Winwood-Smith 2015; Rollins et al. 2015) and marine copepods (reviewed in Lee 2016). Importantly, stress-induced genomic modifications may provide genetic novelties responding to selective pressures in non-native species (Prentis et al. 2008). lncRNAs frequently harbour transposable elements, which are shown to play an adaptive role in heritable genetic diversification for invading populations (e.g., invasive ant *Cardiocondyla obscurior*, Schrader et al. 2014; reviewed in Stapley et al. 2015). Interestingly, I identified five lncRNA transcripts with outlier SNPs between the Atlantic and South Africa, and two outlier SNPs between the Mediterranean and Sydney Harbour; however, conclusions regarding selection on lncRNA transcripts cannot be inferred without further investigation. Additionally, because selection may act only on secondary or tertiary structures (Washietl et al. 2014) rather than primary sequences, lncRNA diversification may occur through evolutionary mechanisms other than single nucleotide polymorphisms that cannot be detected by genomic outlier scans or standard measures of differentiation.

Conclusions

Here, I provide comparative evidence for repeatable genetic changes in multiple introduced *M. galloprovincialis* populations across its present-day distribution in the northern and southern hemispheres. These findings suggest that shared genetic responses to introduction can occur at the level of contigs, but are not likely to involve the same nucleotide polymorphisms. Outliers of differentiation were dominated by protein coding transcripts, although inconsistent repeatability in selective sweep patterns suggests that post-introduction differentiation is due to selective and non-selective processes. Strong conclusions regarding the repeatability of genetic changes and adaptive responses are contingent on resolving the independence and the true sources of introduced populations (Viard et al. 2016). Despite the increased power of high-throughput approaches for detecting population structure in high gene flow marine species, this study reinforces key challenges posed by marine populations for distinguishing accurate native genetic backgrounds (Viard et al. 2016) against which genetic commonalities in introduced populations can be inferred. Focusing on the contribution of lncRNAs to population structure, I can conclude that non-coding transcriptomic elements are unlikely to underlie population-specific adaptations associated with introductions; however alternative modes of evolution, such as adaptive changes to secondary and tertiary molecular

phenotypes should be considered as potent topics for future investigations in invasive species evolution. Taken together, the main findings of this Chapter further our understanding of the genomic scale at which genetic differentiation may be repeatable and shed light on the genomic architecture of post-introduction differentiation in *M. galloprovincialis.* Additionally, this study emphasises the importance of detangling the sources of highly differentiated genomic variation for gaining deeper insight into both the evolutionary constraints and opportunities for adaptation in introduced populations of high gene flow marine non-native species.

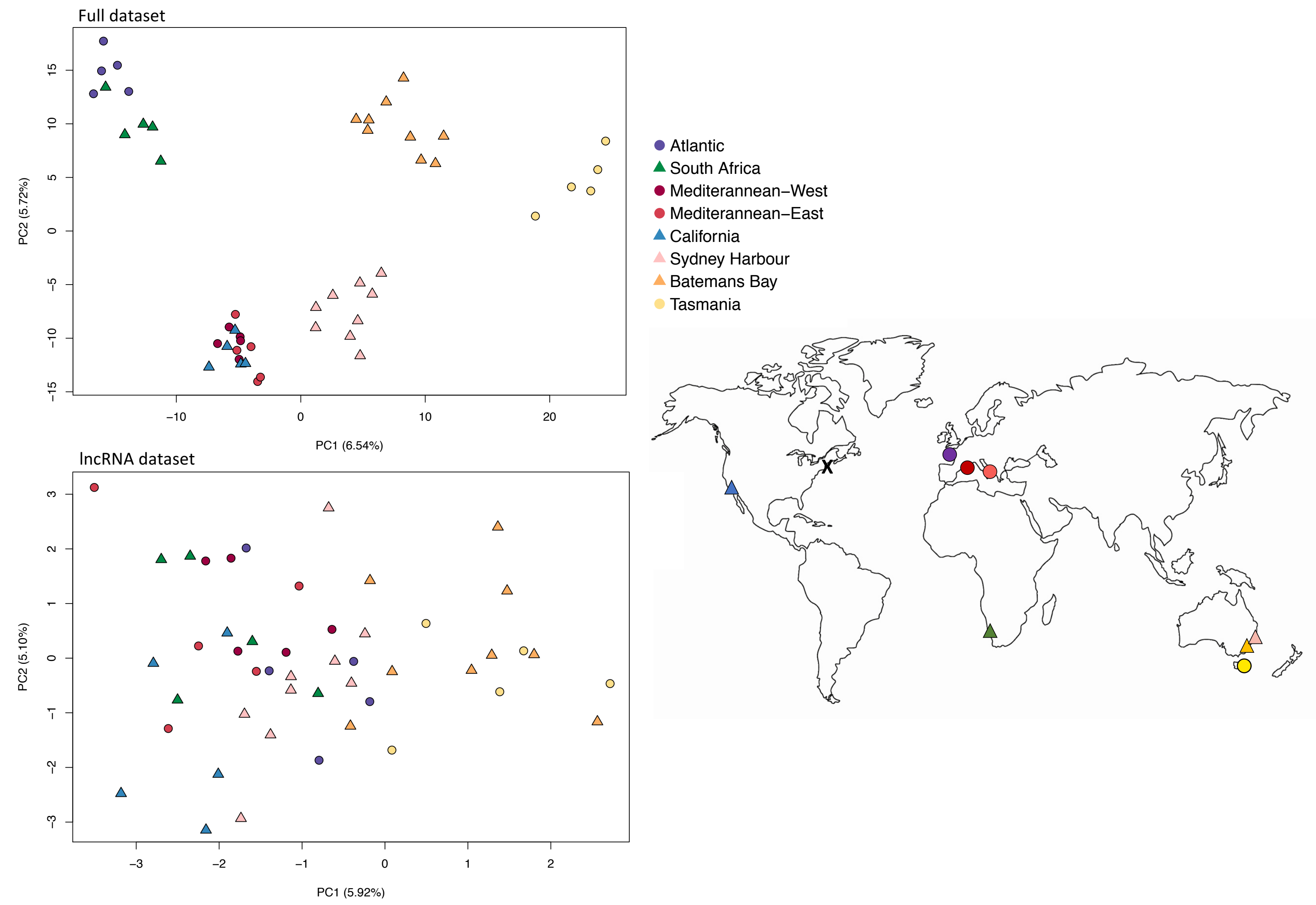


Figure 4-1. Principal components analysis (PCA) of three native (circles) and four introduced (triangles) *M. galloprovincialis* populations sampled from geographic locations shown on the map. *Mytilus edulis* sampling location is indicated with an **X**. Top panel indicates a PCA of 16,779 variants derived from the full transcriptome assembly. The bottom panel shows a PCA across 471 SNPs derived from the lncRNA reference dataset. Tasmanian *M. planulatus* samples are included as a reference population.

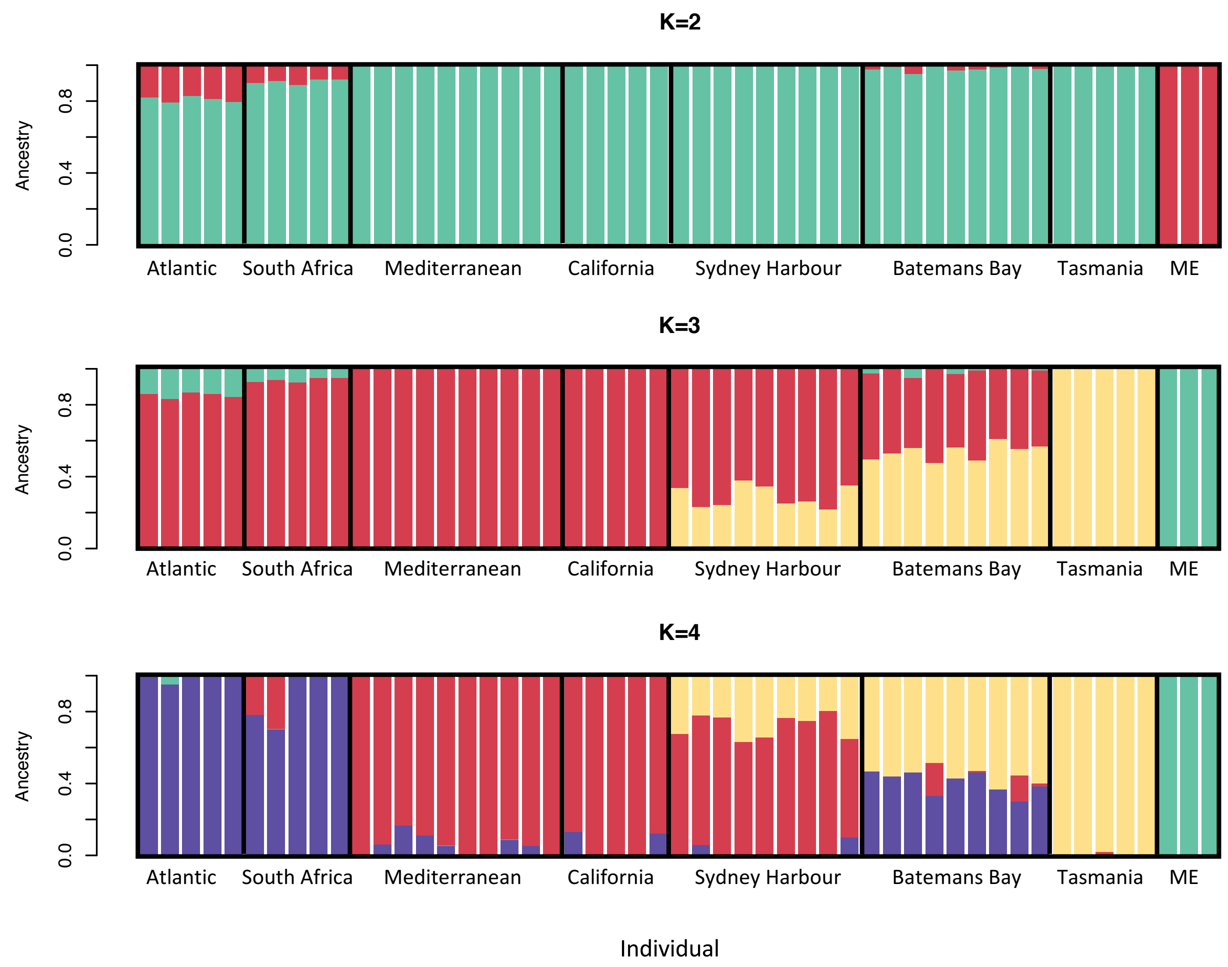
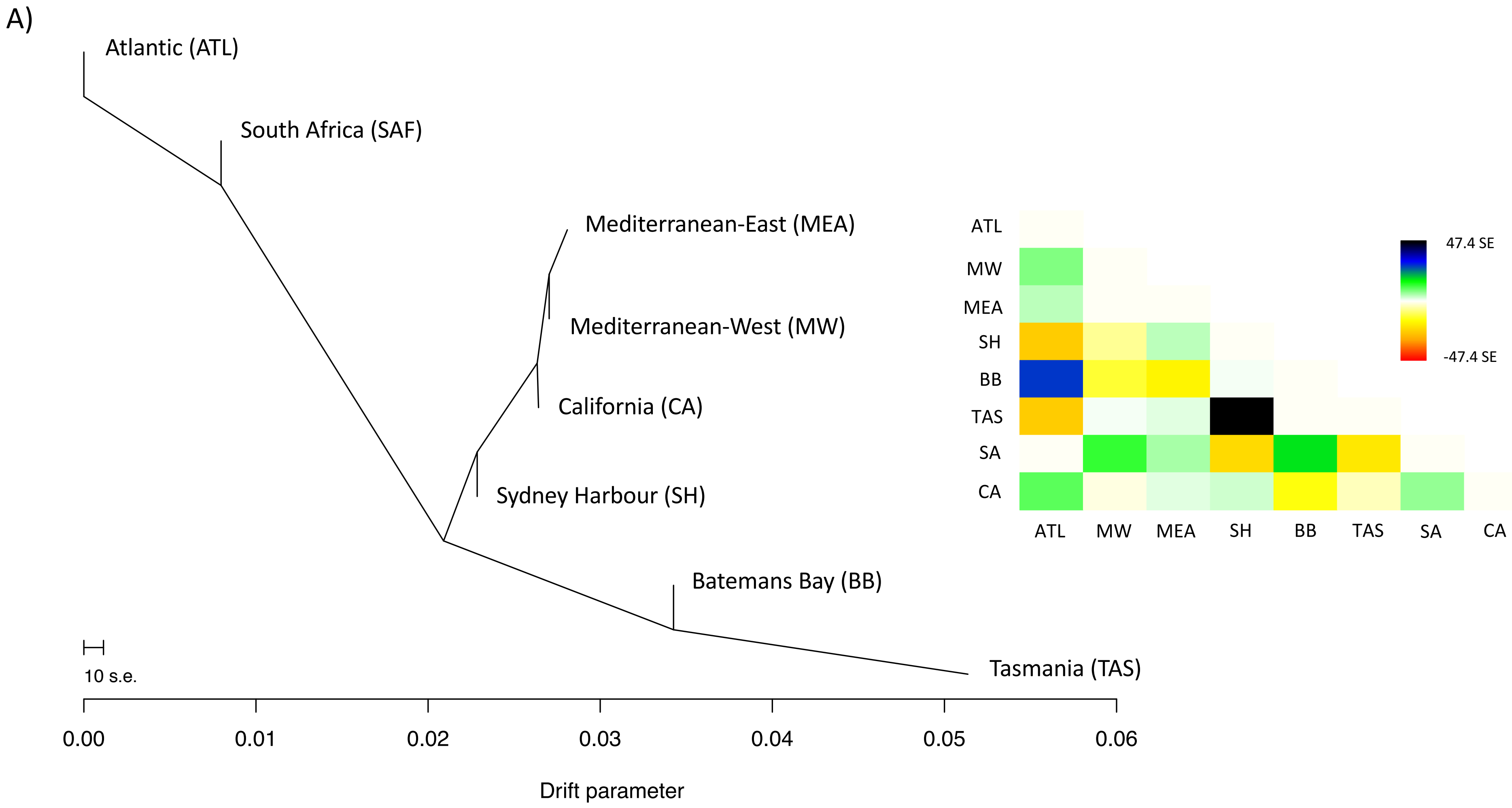


Figure 4-2. ADMIXTURE analyses of introduced and native *M. galloprovincialis* populations, including Australian populations and *M. planulatus* (Tasmania) previously sampled in Chapter 3 and *M. edulis* (ME) as an outgroup taxon. Results for K=2-4 indicate shared ancestry between introduced *M. galloprovincialis* in California and native Mediterranean populations; individuals sampled in South Africa clustered with Atlantic mussels.


A)
Atlantic (ATL)
South Africa (SAF)
Mediterranean-East (MEA)
Mediterranean-West (MW)
California (CA)
Sydney Harbour (SH)
Batemans Bay (BB)
Tasmania (TAS)
10 s.e.
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Drift parameter
ATL
MW
MEA
SH
BB
TAS
SA
CA
47.4 SE
-47.4 SE

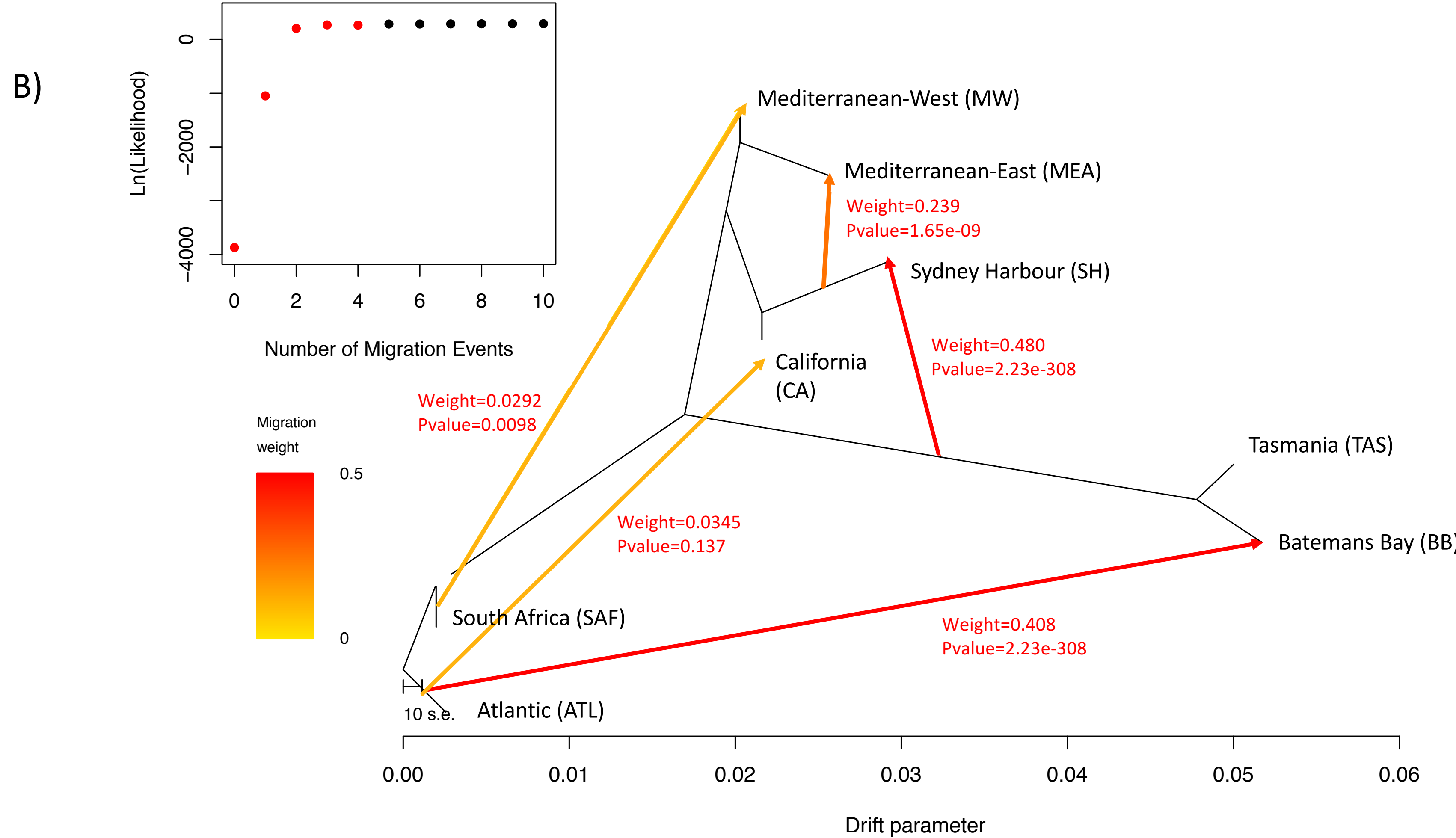


Figure 4-3. Maximum likelihood trees of *M. galloprovincialis* populations using Atlantic *M. galloprovincialis* as an outgroup inferred by TreeMix. Populations trees also include Tasmanian *M. planulatus* as a reference population. **A)** Maximum likelihood population tree without migration. The drift parameter indicates the amount of genetic drift that separates groups. Under a model of zero migration, the heat colours indicate pairwise population residual genotypic covariance of allele frequencies, with the darkest boxes indicating high genotypic covariance between populations. **B)** Maximum likelihood population tree including five migration events, showing the highest significant improvement in the fit of the population tree to the genetic data compared to a no migration model. Significant p-values indicate support for individual migration edges and their corresponding direction and weight.

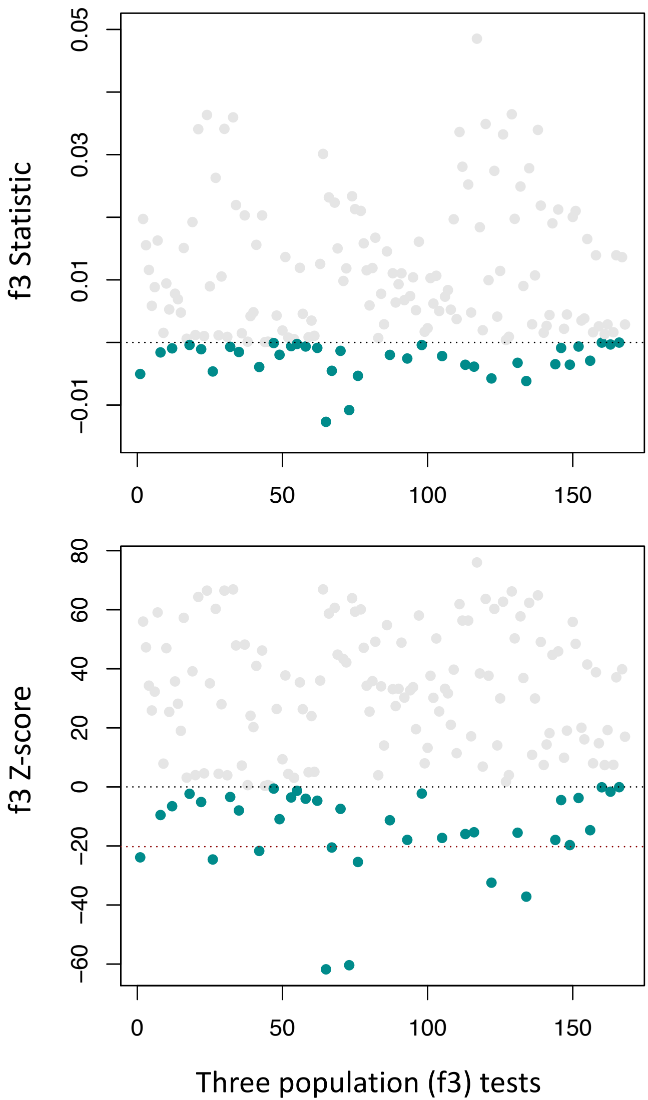


Figure 4-4. Summary of results from three-population (f_3) tests for all population combinations. The top panel shows f_3 statistics for all population combinations, with significant and negative f_3 values indicated in cyan (P < 0.0001; refer to Table 4-3). The bottom panel shows the z-scores corresponding to each test. Points below to dashed red line indicate significant f3 statistics in which the z-score fell below the lowest 5% of the distribution of all values (z-score < -20.25).

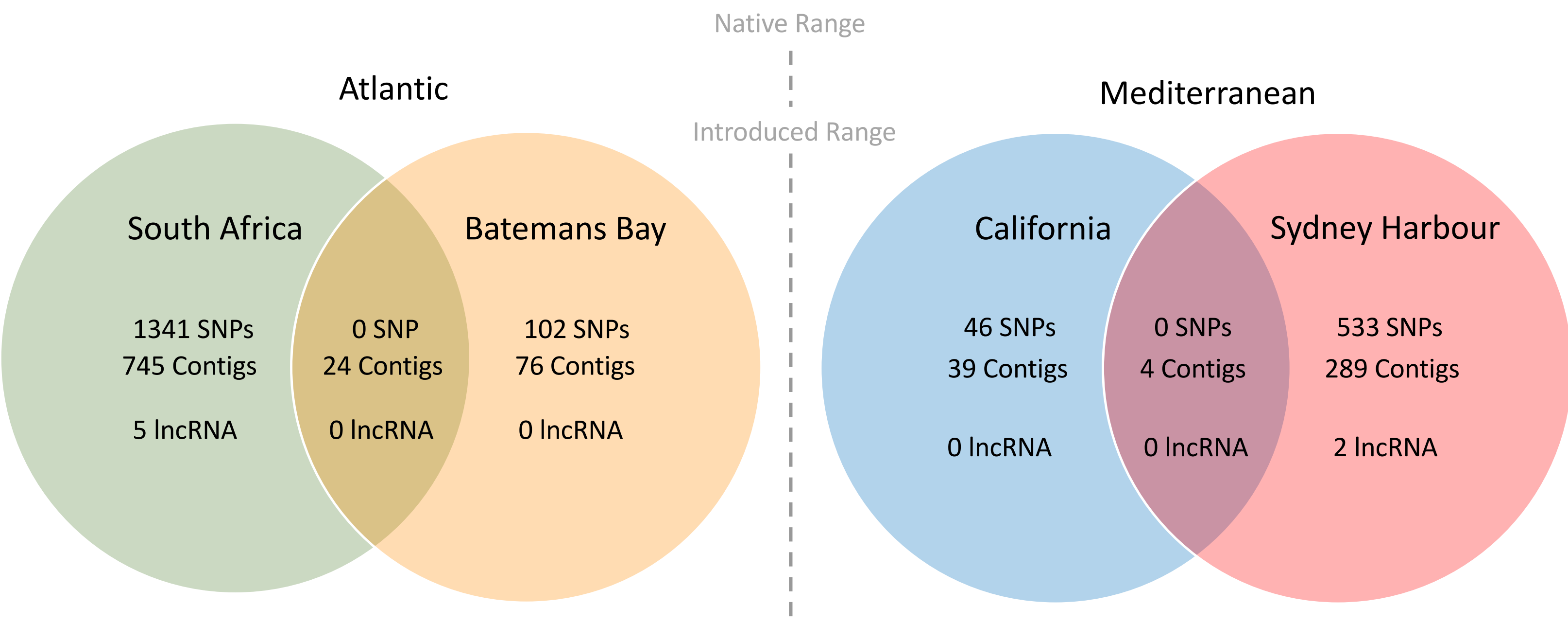


Figure 4-5. Summary of results from genomic outlier analyses conducted between two pairs of introduced populations relative to their native genetic source population. Outlier analyses revealed evidence for parallel genetic differentiation between introduced populations at the level of individual contigs (qvalue= 0.01); although no outlier SNPs were shared among populations pair comparisons. I did not identify any lncRNA transcripts as shared outlier contigs in paired comparisons.

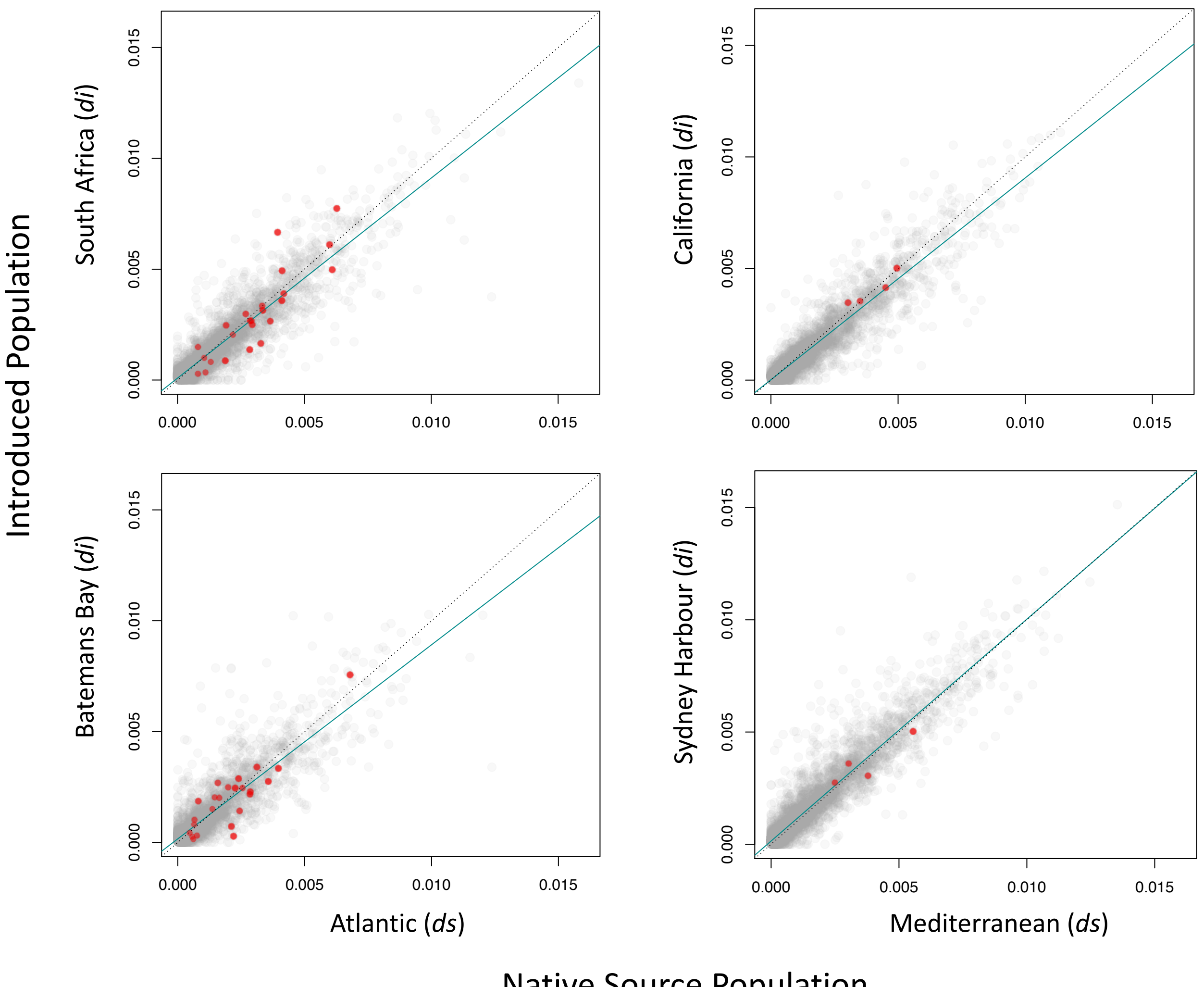


Figure 4-6. Nucleotide diversity comparisons between introduced populations (*di*) and their respective source (*ds*) population. Significant linear regressions of the joint diversity distributions (P<<0.0001) are shown in solid cyan. The slope of each regression was significantly different from 1 (dashed black) in all populations pairs (P<0.01) expect for the Atlantic and Batemans Bay population pair (P=0.794). Out of 24 shared outlier contigs, 8 showed negative residual values indicating reduced diversity in both South Africa and Batemans Bay, compared to the Atlantic *M. galloprovincialis* population. No shared outlier contigs displayed reduced diversity in both Sydney Harbour and California compared to the Mediterranean *M. galloprovincialis* population.

Table 4-1. Summary of *M. galloprovincialis* sampling locations. Species identity and range for samples marked with an asterisk* was based on genome-wide analyses of species relationships (Chapter 3).

Taxon	Sampling Location	Range	Samples sequenced (RNAseq)
Mytilus edulis	Darling Marine Station, Maine, USA	Native	3
Mytilus galloprovincialis	Primel, France (Atlantic)	Native	5
	Crique des Issambles, France (Western Mediterranean)	Native	5
	Herceg Novi, Montenegro (Eastern Mediterranean)	Native	5
	Scripps Oceanographic Institute, California, USA	Introduced	5
	Cape Columbine, South Africa	Introduced	3
	Yzerfontein, South Africa		2
	Sydney Harbour, New South Wales, Australia*[†]	Introduced	9
	Batemans Bay, New South Wales, Australia*[†]	Introduced	9
Mytilus planulatus	Spring Bay, Tasmania, Australia*	Native	5

[†] *M. galloprovincialis* samples are introgressed with *M. planulatus*

Table 4-2. Summary of population differentiation and p-values for full and lncRNA datasets. **(i)** Pairwise F_{ST} values for the full assembly (lower triangle) and for the lncRNA dataset (upper triangle). Values significantly different than zero are marked with an asterisk**; **(ii)** P-values marked with an asterisk indicate significant reductions in F_{ST} in the lncRNA dataset compared to F_{ST} distribution corresponding to 1000 randomly chosen SNP subsets from the full dataset. Significance level is P≤0.05 following corrections for multiple comparison.

(i)	Pairwise Population Comparison: F_{ST} values							
	Atlantic	**Batemans Bay**	**California**	**Mediterranean-East**	**Mediterranean-West**	**South Africa**	**Sydney Harbour**	**Tasmania**
Atlantic	-	0.0296**	0.0518**	0.0499**	0.0314**	0.00687**	0.0276**	0.0585**
Batemans Bay	0.0524**	-	0.0516**	0.0465**	0.0300**	0.0301**	0.0194**	0.000513
California	0.0666**	0.0555**	-	0.0232**	0.0326**	0.0324**	0.0174**	0.0594**
Mediterranean-East	0.0720**	0.0529**	0.00243	-	0.000594	0.0250**	0.0184**	0.0550**
Mediterranean-West	0.0648**	0.0495**	0.00322	-0.00335	-	0.0193**	0.0174**	0.0376**
South Africa	0.00763**	0.0416**	0.0490**	0.0512**	0.0414**	-	0.0174**	0.0727**
Sydney Harbour	0.0765**	0.0289**	0.00967**	0.0116**	0.0146**	0.0553**	-	0.0261**
Tasmania	0.133**	0.0264**	0.103**	0.0969**	0.0965**	0.118**	0.0513**	-

(ii)	Pairwise Population Comparison: P-values							
	Atlantic	**Batemans Bay**	**California**	**Mediterranean-East**	**Mediterranean-West**	**South Africa**	**Sydney Harbour**	**Tasmania**
Atlantic	-							
Batemans Bay	0.022	-						
California	0.240	0.389	-					
Mediterranean-East	0.112	0.347	0.983	-				
Mediterranean-West	0.024	0.048	0.995	0.664	-			
South Africa	0.485	0.143	0.138	0.038	0.059	-		
Sydney Harbour	0.000**	0.139	0.823	0.824	0.629	0.002**	-	
Tasmania	0.000**	0.003**	0.006**	0.007**	0.001**	0.006**	0.023	-

Table 4-3. Summary of three population (f_3) tests. Values are shown for population comparisons yielding significant and negative f_3 statistics (P-value< 0.0001). Values in bold indicate z-scores within the lower 5th quantile (z-score < -20.25) based on all observations.

PopA (Admixed)	PopB (Ancestor 1)	PopC (Ancestor 2)	f_3 Statistic	z-score
Batemans Bay	**Mediterranean-West**	**Tasmania**	**-0.00504**	**-23.9**
Mediterranean-West	Batemans Bay	Mediterranean-East	-0.00159	-9.53
Sydney Harbour	Mediterranean-West	Batemans Bay	-0.000938	-6.55
Mediterranean-West	Tasmania	Mediterranean-East	-0.00108	-5.12
Sydney Harbour	**Mediterranean-West**	**Tasmania**	**-0.00463**	**-24.6**
Mediterranean-West	Atlantic	Mediterranean-East	-0.00153	-7.99
South Africa	**Mediterranean-West**	**Atlantic**	**-0.00392**	**-21.7**
Mediterranean-West	Mediterranean-East	South Africa	-0.00197	-10.9
California	Mediterranean-West	Sydney Harbour	-0.000643	-4.01
Mediterranean-West	South Africa	California	-0.000870	-4.68
Batemans Bay	**Tasmania**	**Atlantic**	**-0.0127**	**-61.8**
Batemans Bay	**Tasmania**	**Mediterranean-East**	**-0.00453**	**-20.5**
Batemans Bay	Tasmania	Sydney Harbour	-0.00134	-7.45
Batemans Bay	**Tasmania**	**South Africa**	**-0.0108**	**-60.4**
Batemans Bay	**Tasmania**	**California**	**-0.00533**	**-25.4**
South Africa	Batemans Bay	Atlantic	-0.00199	-11.3
Sydney Harbour	Batemans Bay	Mediterranean-East	-0.00256	-17.9
Sydney Harbour	Batemans Bay	California	-0.00217	-17.3
Sydney Harbour	Tasmania	Atlantic	-0.00357	-15.9
South Africa	Tasmania	Atlantic	-0.00387	-15.4
Sydney Harbour	**Tasmania**	**Mediterranean-East**	**-0.00575**	**-32.4**
Sydney Harbour	Tasmania	South Africa	-0.00325	-15.5
Sydney Harbour	**Tasmania**	**California**	**-0.00617**	**-37.2**
South Africa	Atlantic	Mediterranean-East	-0.00348	-17.9
California	Atlantic	Mediterranean-East	-0.000876	-4.47
South Africa	Atlantic	Sydney Harbour	-0.00355	-19.7
South Africa	Atlantic	California	-0.00292	-14.7

Table 4-4. Summary of Uniprot annotations for 8 intraspecific outliers contigs that were shared between Atlantic *M. galloprovincialis* derived populations (South Africa and Batemans Bay) and also showed evidence of reduced gene-wide nucleotide diversity (i.e. negative residuals) when compared to the native genetic background (also refer to Figure 4-6).

Outlier	Length	UNIPROT ID | Annotation
Contig15574	2687	Q9BLG0 | TNNC_TODPA Troponin C
Contig21706	2919	O94903 | PROSC_HUMAN Proline synthase co-transcribed bacterial homolog protein
Contig2717	3816	Q9HD45 | TM9S3_HUMAN Transmembrane 9 superfamily member 3
Contig4256	3916	Q9H987 | SYP2L_HUMAN Synaptopodin 2-like protein
Contig44229	7213	Q68FD5 | CLH1_MOUSE Clathrin heavy chain 1
Contig44591	3047	Q5T848 | GP158_HUMAN Probable G-protein coupled receptor 158
Contig44596	5385	P07742 | RIR1_MOUSE Ribonucleoside-diphosphate reductase large subunit
TR33645bc1_g1_i1	2605	P17133 | RU17_DROME U1 small nuclear ribonucleoprotein 70 kDa

Chapter 5. General Discussion

Overview

It has been exactly 20 years since the first application of high-throughput sequencing to study biological invasions (American bullfrog; Ficetola et al. 2008; Rius et al. 2015b). Two decades of invasion genomics have allowed deeper insight into a range of evolutionary questions regarding the success of non-native species. Empirical advances have improved our understanding of the influence of genetic drift (e.g., Dlugosch and Parker 2008) and admixture (e.g., Mesgaran et al. 2016) in resolving the paradox of species invasions (Estoup et al. 2016), and we now have key examples of molecular and phenotypic traits underpinning adaptive evolution in non-native species (e.g., Vindepitte et al. 2014; Schrader et al. 2014; reviewed in Bock et al. 2015). In the ocean, genomic tools have been essential for detecting introduced marine taxa (i.e. through metabarcoding approaches; reviewed in Viard et al. 2016) and for distinguishing introduced populations from endemic congeners (Geller et al. 2010; Bouchemousse et al. 2016). The ability to genotype individuals at thousands of markers has also provided greater power to distinguish weakly differentiated populations (Gagnaire et al. 2015; Guzinski et al. 2018), resolve introduction sources (e.g., Tepolt and Palumbi 2015), and generate predictions about loci influenced by selection in both native (Galindo et al. 2010; Riquet et al. 2013) and introduced ranges (Bernardi et al. 2016). Despite considerable progress, empirical work to date has also uncovered inherent complexities and idiosyncrasies accompanying marine introductions, highlighting challenges for resolving the basic features of marine invasions even when genome-wide data are analysed (Viard et al. 2016). Recent review articles have also raised open evolutionary questions that are either understudied or yet to be explored in the field of marine invasion genomics: Do adaptive differences between closely-related species promote invasive potential (Tepolt 2015)? How variable are the outcomes of interspecific and intraspecific admixture in post-introduction evolution (Bock et al. 2015)? How often does local adaptation play a role in successful marine introductions (Viard et al. 2016)? What influence does the regulatory genome have on rapid evolutionary responses to introduced environments (Sherman et al. 2016)?

This thesis contributes new empirical work towards an improved understanding of these unresolved questions and provides new perspectives regarding the contributions to pre- and post-introduction evolution in the success of marine non-native species. Chapter 2 furthers our understanding of the genomic features that are unique to invasive taxa and sheds light on the molecular traits that may

underlie species-specific differences in invasive potential. Chapter 3 and Chapter 4 contribute to improved knowledge on how introduced and native populations become genetically differentiated following introduction in new environments. Below, I discuss the major findings of each Chapter and how they advance our understanding of *M. galloprovincialis* evolution and the processes accompanying present day introductions. I also draw upon the broad contributions of this body of work to the field of marine invasion genomics, and discuss important questions for future investigations raised from this research.

'Pre-adaptation'

Physiological studies in the *Mytilus* system have collectively suggested that physiological pre-adaptations in temperature tolerance are a primary factor explaining the ability of introduced *M. galloprovincialis* to outcompete native congeners in warm habitats where their distributions overlap (Lockwood and Somero 2011a; Fields et al. 2012). These investigations have also provided insight into the genetic basis of these interspecific differences that may also underlie apparent variation in invasive capacity (Lockwood et al. 2015). In Chapter 2, I explored whether the physiological traits that make invasive *M. galloprovincialis* distinct from non-invasive congeners are paralleled by divergence at the molecular level. I utilised this comprehensive body of physiological and biochemical literature to test the hypothesis that genomic functions experimentally associated with adaptations to temperature-stress have also experienced accelerated evolutionary divergence in warm-tolerant and invasive *M. galloprovincialis*, compared to three cold-tolerant, non-invasive congeners. The results of this Chapter provide several lines of molecular evidence for lineage-specific accelerated divergence in protein-coding genes that are consistent with genomic functions (primarily oxidative stress) associated with the strongest species-specific responses to thermal stress in comparative physiological studies. These findings point to functional genomic variation differentiating warm-adapted *M. galloprovincialis* from cold-adapted congeners and implicate a contribution of temperature-dependent selection in the divergence of invasive and non-invasive species in this genus. Importantly, these data provide corroborative evidence that genomic functions linked to temperature adaptation and invasive success at the whole organism level may also underlie molecular pre-adaptations contributing to the evolution of invasiveness in *M. galloprovincialis*.

A significant consideration of these interpretations is that comparative inferences based on a single invasive taxon against three non-invasive congeners do not offer a statistical test of the pre-adaptation hypothesis; rather, such data provide only corroborative evidence of interspecific genetic variation consistent with physiological studies. Additionally, the conclusions of this Chapter are contingent on

the assumption that selected candidate genes in *M. galloprovincialis* are indicative of the type of selection involved in species divergence (i.e. temperature-dependent selection). In the present study, however, striking functional concordance between positively selected thermotolerance candidates (loci linked to thermal adaptation in physiological studies) and temperature-related loci with the strongest signatures of selection implicate a common selective pressure acting the same genomic functions in the *M. galloprovincialis* lineage. These findings demonstrate how independent lines of experimental evidence from comparative physiological studies of environmental adaptation can be leveraged to strengthen evolutionary hypotheses about adaptation in non-model species. Useful future endeavours should examine broader phylogenomic datasets that exploit diversity across multiple invasive and non-invasive taxa (e.g., Lai et al. 2012; Hodgins et al. 2014), rather than limiting comparisons to a single genus. Such studies will be essential for making stronger inferences about the influence of selection on various levels of molecular phenotypes in invasive taxa, including the evolution of individual amino acids, genes, functional groups and entire metabolic pathways. Shared evolutionary targets may, in turn, reveal genomic features predisposing certain lineages to becoming successful invaders (Tepolt 2015) or may point to shared evolutionary trade-offs contributing to invasive success (Hodgins et al. 2014; Chown et al. 2015). The candidate genes identified in this Chapter should be of notable interest for functional genetic assays of protein functions and their role in temperature adaptation in this system. Such data may therefore form the basis of predictions for future comparative tests of pre-adaptation hypotheses in other non-model marine invasive taxa.

Post-introduction evolution: Hybridisation and introgression

Genomic datasets have greatly enhanced our ability to leverage signals of isolation and migration from molecular data, allowing sophisticated model-based inferences of demography history that do not require large numbers of sampled individuals (because of increased power afforded by considering thousands of loci) (Roux et al. 2016; Gagnaire et al. 2018). In Chapter 3, I utilise the power of demographic modelling together with genomic analyses of migration to document strong evidence for secondary contact and subsequent introgression between introduced *M. galloprovincialis* and the morphologically cryptic native Australian taxon, *M. planulatus*. The major findings of this Chapter advance our understanding of the potential genetic outcomes of human-mediated secondary contact and include three important results: First, I establish that two contemporary introductions of *M. galloprovincialis* originate from genetically divergent northern source lineages from the Mediterranean Sea and the Atlantic coast of Europe. These findings are in line with general perceptions that successful marine introductions are likely to involve propagules from multiple and potentially diverse source populations (Lockwood et al. 2005; Tepolt and Palumbi

2015). Second, I discover that both introductions are associated with high rates of introgression between introduced and endemic populations. Additionally, one admixed population retained signatures of divergent northern *M. edulis* genetic variation likely obtained through post-introduction admixture with introgressed Atlantic *M. galloprovincialis* (rather than direct gene flow with northern *M. edulis*). These results reinforce the notion that patterns of genomic differentiation and signatures of migration must be interpreted within the context of the demographic history of the lineages studied (Hoban et al. 2016). Specifically, sufficient sampling of closely related sister lineages in both sympatric (i.e. endemic *M. planulatus*) and allopatric ranges (i.e. *M. edulis* outgroup) are imperative for accurate interpretations of introgression and long-term impacts on native genetic diversity. These findings also demonstrate that adequate sampling in the native range will be essential for (i) resolving the exact genetic sources of introduced populations, and (ii) for stronger inferences of post-introduction evolution through hybridisation. Full representation of native genetic diversity will be critical to exclude alternative hypotheses that genetic differentiation in introduced populations represents unsampled variation from the native range or that genetic differentiation is an outcome of local adaptation following introduction (Dlugosch et al. 2015).

The third major finding of this Chapter is that demographic modelling supported historical isolation between *M. planulatus* and *M. galloprovincialis* of at least 100,000 years, despite close genomic similarities. *Mytilus planulatus* is currently recognised as valid nomenclature in the World Register of Marine Species (WORMS) database. Strong sequence similarities and permeable barriers to gene flow with northern *M. galloprovincialis,* however, support previous proposals to assign the endemic Australian lineage with a regional subspecies status (e.g., Borsa et al. 2007, Borsa and Daguin 2000, Gerard et al. 2008, Westfall and Gardner 2010, Hilbish et al. 2000). From an applied perspective, this presents imminent issues for standardising nomenclature of native Australian *Mytilus* among government and industry bodies. To date, inconsistent classification of the endemic species as 'native southern *Mytilus galloprovincialis*' (Department of Primary Industries, NSW) or '*Mytilus edulis*' (Department of Primary Industries and Resources, SA), has sustained ongoing confusion regarding the species identity of Australian endemic mussels. While this empirical Chapter fills a significant gap in our systematic knowledge of native Australian biodiversity, these findings raise important ethical considerations for the conservation and management of native genetic diversity. How much divergence is enough for endemic genetic diversity to be intrinsically valuable? How do we treat native species challenged with extensive invasive hybridisation? For invasive species research, these findings also echo major challenges for delineating closely-related endemic and introduced species and for monitoring cryptic marine introductions (Viard et al. 2016). For the case of *Mytilus* congeners, it is evident that large numbers of loci are required to adequately resolve endemic and introduced

populations. Moving forward, the genomic data presented here form the basis of ongoing SNP marker development for distinguishing introduced and endemic *Mytilus* species and their hybrids. Indeed, protecting Australia's natural resources and biodiversity from invasive disease and pests is a high priority research area and a national goal for the Australian government (Hayes et al. 2005). The present research also offers future prospects to extend genomic tools to aquaculture industry and offer industry scientists methods to determine the extent to which introduced *M. galloprovincialis* is being cultivated in Australia.

From a molecular perspective, this research Chapter raises important considerations regarding the possibility of adaptive introgression of diverse genetic backgrounds into native genetic diversity (e.g., Keller and Taylor 2010; Rius and Darling 2014; Saarman and Pogson 2015). Indeed, there is growing evidence that hybridisation between introduced and native genotypes may enhance the spread of some non-native species. Importantly, hybridisation may accelerate successful introductions by increasing genetic diversity through adaptive introgression of favoured genetic variants into the non-native genomic background (Currat et al. 2008; Schierenbeck and Ellstrand 2009; Hovick and Whitney 2014); or conversely, into the native genetic background (e.g., Fitzpatrick et al. 2010). Experimental studies testing links between introgression and colonisation success will greatly advance our understanding of these synergies (Rius and Darling 2014; Hovick and Whitney 2014; Bock et al. 2015). An obvious next step in the *Mytilus* system is to resolve whether introgression between *M. galloprovincialis* and *M. planulatus* is a causal factor in the success of introduced populations in Australia. Future research should focus on temporal and spatial sampling to assess the stability of hybrid zones and the rate of dispersal of admixed genotypes throughout introduced ranges. For example, whether hybrid mussels experience higher physiological tolerance than parental lineages is unknown and represents a fruitful future avenue for understanding the mechanisms promoting successful marine introductions. Delineating the physiological limits of hybrid genotypes may additionally facilitate unexplored predictions about how the fitness outcomes of hybridisation may change with density dependence in expanding introduced populations (Dlugosch et al. 2015). At the genomic level, investigations resolving which loci are experiencing variable rates of gene flow will also be important for understanding the role of introgression and adaptation in successful *M. galloprovincialis* introductions. On the other end, identifying genomic regions that experience reduced levels of gene flow across hybrids zones will inform our understanding of the rate at which species barriers accumulate between incipient native and introduced taxa. Parallel introductions and invasive-native hybrid zones along the Australian coastline hold great potential to explore these replicated effects of introgression on the genome, and underline several intriguing questions for future investigation about how hybridisation may alter genomic architecture in species introductions: Which

genes are more likely to cross species boundaries and in which direction does introgression occur? Are the same genes involved in replicated introductions and are these the same loci previously implicated in adaptive introgression in other contact zones (e.g., *Mytilus* hybrid zone, Fraisse et al. 2015)? Do variable introgression rates across the genome promote the evolution of co-adapted complexes of genes and how do they evolve to promote adaptation in invasions?

Post-introduction evolution: Repeatability of genetic differentiation

Chapter 4 draws upon major results from Chapter 3 to investigate the repeatability of post-introduction evolution in multiple introduced *M. galloprovincialis* populations in the northern and southern hemispheres. I also take advantage of the unique features of RNAseq sequencing technology to investigate the putative sources of selectively constrained and favoured variation contributing to differentiation between introduced and native populations. The main findings in this Chapter reveal comparative evidence that parallel genetic differentiation has occurred in two introduced *M. galloprovincialis* populations sampled in Australia (Chapter 3) and replicated introductions in California and South Africa that originate from the same native genetic backgrounds. Specifically, I find that the same individual contigs were involved in shared deviations in genetic differentiation; however repeated evolution was not driven by the same nucleotide substitutions. I also discover that long non-coding RNA (lncRNA) transcripts contributed significantly less to population structure compared to all expressed transcripts; this result suggests that sequence variation in these regulatory molecules evolves under selective constraints and is unlikely to be implicated in population-specific adaptations on a timescale applicable to species introductions. Taken together, the main findings of this Chapter further our understanding of the genomic scale at which genetic differentiation may be repeatable and shed light on the genomic architecture of post-introduction differentiation in *M. galloprovincialis.* Indeed, previous authors have concluded that the form and nature of genomic variation (i.e. gene position and recombination landscape) may be more relevant for both the predictability of post-introduction evolution (Renaut et al. 2014) and the success of non-native populations (Dlugosch et al. 2015), rather than the quantity or function of introduced genetic variation. A key outstanding question for *M. galloprovincialis*, however, will be to determine whether shared outliers of differentiation are relevant for adaptation in the introduced range.

Three key results of this Chapter bring to light important perspectives for interpreting the actions of selection in *M. galloprovincialis* introductions and have broad applications for the study of adaptation in other marine non-native taxa. First, inconsistent repeatability in selective sweep patterns in shared outlier contigs suggests that differentiation of introduced variation is due to both selective and non-

selective processes. An alternative hypothesis, however, is that patterns of increased diversity at genomic outliers in the introduced range (relative to native populations) is the result of changes in allele frequencies in response to a relaxation of selection. Such a scenario would be expected if introduced populations are released from strong selective pressures that are present within the native range. To my knowledge this hypothesis, from a genomic perspective, remains to be tested in marine (or animal) invasive species and represents an interesting topic for future investigations. Comparative molecular evolutionary approaches that apply codon models to search for signatures of relaxed selection across lineages (Wertheim et al. 2014) may hold promise for detecting the evolution of non-native populations or taxa under released selective pressures in introduced ranges.

Second, the results of this Chapter suggest that genomic outliers of differentiation between native and introduced *M. galloprovincialis* populations were dominated by protein-coding transcripts. Small sample sizes, however, did not enable inferences regarding the source of the differentiated variation; that is, whether shared patterns of genetic differentiation across individual genes were due to new coding mutations arising in introduced populations or a result of shifts in alleles frequencies from standing genetic variation present in the native range. Replicated species introductions provide ideal frameworks for deciphering between selection on standing genetic variation and novel alleles arising in independent populations, as native populations can be used to characterise ancestral variation and haplotypes (Sherman et al. 2016). Recent advances in statistical approaches using coalescent theory and genome-wide data provide promising avenues towards distinguishing between these mechanisms of parallel evolution across multiple populations or taxa (Lee and Coop 2017). A particular advantage of the approach proposed by Lee and Coop (2017) is that molecular signatures of selection on linked sites can also be leveraged to detect when beneficial alleles arise in separate populations through gene flow. Such insight would help clarify the role of allele sharing through stepping stone introductions in invasive species evolution, and can be applied in the absence of *a priori* knowledge about phenotypic trait variation or candidate loci.

The third result relevant for interpreting the actions of selection is that lncRNAs evolving under purifying selection or other evolutionary constraints are unlikely to underlie adaptive sequence variation associated with post-introduction evolution. For the case of non-coding transcripts, selection may act only on secondary of tertiary structures, rather than primary sequences (Washietl et al. 2014). lncRNA diversification may, therefore, occur through evolutionary mechanisms other than single nucleotide polymorphisms, which are the principal data type informing genomic outlier analyses and standard measures of differentiation. Indeed, the role of other structural polymorphisms (not easily detected by SNP-based methods) is gaining greater interest among invasion biologists, with recent

literature highlighting the utility of invasive species to study their contribution to rapid post-introduction evolution (Dlugosch et al. 2015; Sherman et al. 2016). Indeed, the first documented evidence for the genetic basis of adaptation in an invasive species involved chromosomal inversions in the salivary glands of *Drosophila* (Carson 1965; Dobzhansky 1965). It is now evident that chromosomal inversion polymorphisms can evolve repeatedly (Jones et al. 2012) and are associated with successful invasions (Kirkpatrick 2010; Pascual et al. 2007; Dlugosch et al. 2015). At the population level, intragenic copy number changes in tandem repetitive elements have also been associated with putatively adaptive variation (e.g., Popovic et al. 2014). All of these data suggest a wide range of mutation types other than point mutations occur frequently and evolve rapidly within species. Investigating higher-level dimensions of genetic variation may also hold important implications for studying large-effect phenotypic changes and adaptation in successful marine invasions.

Common genomic responses to introduction may also involve heritable mechanisms other than molecular sequence evolution most generally. Changes in gene expression can be adaptive, and there is growing evidence that post-introduction selection on heritable regulatory changes may have an important role in the initial establishment and adaptation in non-native species (Richards et al. 2006; Pérez et al. 2006; reviewed in Sherman et al. 2016; Wellband and Heath 2017). RNAseq approaches can also be used to detect shared differentiation in gene expression associated with non-native taxa that can illuminate common traits or selective pressures involved in introduced species adaptation (Rius et al. 2015b; Tepolt 2015) or reveal adaptive changes in plasticity in key gene functional groups (Hodgins et al. 2013). Indeed, one of the best examples of such an investigation comes from warm-adapted and cold-adapted *Mytilus* species (Lockwood et al. 2010), which formed the basis of the hypotheses tested in Chapter 2. The contributions of epigenetic or non-genetic heritable changes in invasive species success, however, are still largely unexplored (Chown et al. 2015; Verhoeven et al. 2016; Sherman et al. 2016; Bourne et al. 2018). Yet, there is evidence that epigenetic mechanisms through stress-induced genomic modifications, such as transposable elements, can directly affect traits involved in invasions by altering gene expression or function (e.g., Schrader et al. 2014; reviewed in Stapley et al. 2015; Bock et al. 2015). Because such changes occur through a number of evolutionary mechanisms that do not involve sequences changes (i.e. DNA methylation, histone alterations and modifications to non-coding RNAs), rapid evolution of the epigenome is not burdened by the time required for new mutations to arise. It is therefore posited that non-genetic evolution may have a central role in boosting genetic variation in the earliest stages of successful invasions (Prentis et al. 2008; Ardura et al. 2017). While Chapter 4 provides the first comparative insights into the evolution of non-coding regulatory RNAs in *Mytilus* species, there are currently no investigations

examining the role of epigenetic processes and potential links to adaptation in marine invasive species (Sherman et al. 2016).

Conclusions

The work of this thesis represents the most comprehensive genomic investigation of a marine invasive species in Australia to date and contributes to empirical research regarding the importance of pre- and post- introduction evolutionary processes in successful marine non-native species. I demonstrate that RNAseq is a powerful methodology for investigating multiple aspects of invasion biology, particularly in non-model species without a complete reference genome (Sherman et al. 2016). This thesis illustrates the application of RNAseq data in a range of genome-wide analyses within and between species: Comparative analyses of protein-coding sequence data to investigate the rate of molecular evolution between congeners; phylogenomic analyses of species relationships and coalescent demographic modelling using transcriptome derived haplotypes to contrast interspecific divergence and gene flow with contemporary introgression; and finally, population genomic analyses of SNP variants to partition the contributions of coding and non-coding expressed elements to intraspecific differentiation. The annotated transcriptome assemblies of five *Mytilus* congeners (including endemic Australian *M. planulatus*) are major contributions to existing genomic resources for *Mytilus* mussels and support the growing body of empirical studies that are extending high-throughput approaches to the study of non-model marine species.

Taken together, the empirical work of this thesis demonstrates that multiple modes of evolution are likely to operate at various stages of the introduction process, at least for the marine invasive mussel, *M. galloprovincialis*. For example, a species may be already genetically well-matched to an introduced environment thus allowing successful establishment in the introduced range; however, a founding population may subsequently undergo genetic adaptations to maximise fitness during later stages of the invasion. For the case of *M. galloprovincialis,* warm-temperature tolerance appears to have facilitated the initial invasions into California (Lockwood and Somero 2011a; Lockwood et al. 2015), and divergence from non-invasive congeners at thermal adaptation gene functions is evident at the molecular level (Chapter 2). This thesis also demonstrates that other post-introduction processes, specifically, introgression with native congeners (Chapter 3) and possibly local adaptation (Chapter 4) are involved in the evolution of independent *M. galloprovincialis* introductions. These results imply that even for a single species, model predictions of marine invasive spread should ideally incorporate all of these processes: pre-existing environmental tolerance (which is already the primary assumption of environmental niche models; reviewed in Tepolt 2015), shifts in

environmental niches through adaptive genetic changes following introduction and the potential for expansions of introduced hybrid populations introgressed with native genetic diversity. Discrepancies in such predictions may inform future study designs and sampling efforts where observed distributions of introduced taxa do not fit model expectations (Riginos et al. 2016).

As non-native species are increasingly identified in the most remote regions of our oceans, knowledge regarding potential impacts on native biodiversity, and whether species introductions are more likely to be successful under warming ocean conditions will be essential for predicting and minimising future introductions (Stachowicz et al. 2002; Molnar et al. 2008; Chown et al. 2015). Moving forward, interdisciplinary approaches could best fill significant gaps in our understanding regarding the predictability of species introductions and the frequency and nature of adaptation (including hybridisation) in successful marine non-native species. Genomic approaches can infer the action of selection on molecular targets, but without the context of the organism's environment or phenotypic traits involved in adaptation. Examining changes in gene expression variation may allow tests of specific hypotheses regarding the role of environmental pressures on introduced populations, without strong inferences about whether shared physiological responses are heritable or involve plastic responses to introduced conditions (Sherman et al. 2016). Traditional evolutionary approaches, including quantitative genetic studies can examine detailed heritable trait variation evolving in introduced and native ranges (Colautti and Lau 2015); however, such investigations do not reveal causative genetic variation and are contingent on multigenerational common garden or reciprocal transplant experiments not possible for many bentho-pelagic marine taxa. Leveraging the strengths of various genomic approaches, and harnessing existing ecological, physiological and molecular knowledge for marine invaders (e.g., *Ciona robusta* genome; Dehal et al. 2002) and their congeners will greatly inform comparative studies and yield new and exciting opportunities for understanding and predicting successful marine invasions.

Literature Cited

Ab Rahim, E. S., Nguyen, T. T., Ingram, B., Riginos, C., Weston, K. J., & Sherman, C. D. (2016). Species composition and hybridisation of mussel species (Bivalvia: Mytilidae) in Australia. *Marine and Freshwater Research*, 67(12), 1955–1963.

Abbott, R. J. (1992). Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution*, 7(12), 401–405.

Abele, D., & Puntarulo, S. (2004). Formation of reactive species and induction of antioxidant defence systems in polar and temperate marine invertebrates and fish. *Comparative Biochemistry and Physiology Part a: Molecular & Integrative Physiology*, 138(4), 405–415.

Alexa, A., & Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. R Package Version, 2(0).

Alexa, A., Rahnenfuhrer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13), 1600–1607.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

Allendorf, F. W., & Lundquist, L. L. (2003). Introduction: population biology, evolution, and control of invasive species. *Conservation Biology*, 17(1), 24–30.

Amaral, P. P., Dinger, M. E., & Mattick, J. S. (2013). Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Briefings in Functional Genomics*, 12(3), 254–278.

Anderson, A. S., Bilodeau, A. L., Gilg, M. R., & Hilbish, T. J. (2002). Routes of introduction of the Mediterranean mussel (*Mytilus galloprovincialis*) to Puget Sound and Hood Canal. *Journal of Shellfish Research*, 21(1), 75–79.

Anisimova, M., & Liberles, D. A. (2007). The quest for natural selection in the age of comparative genomics. *Heredity*, 99(6), 567.

Anisimova, M., & Liberles, D. A. (2012). Detecting and understanding natural selection. *Codon Evolution: Mechanisms and Models*, 73–96.

Anisimova, M., & Yang, Z. (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution*, 24(5), 1219–1228.

Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution*, 18(8), 1585–1592.

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., et al. (2012). The magnitude of global marine species diversity. *Current Biology*, 22(23), 2189–2202.

Apte, S., Holland, B. S., Godwin, L. S., & Gardner, J. P. (2000). Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions*, 2(1), 75–79.

Arcella, T. E., Perry, W. L., Lodge, D. M., & Feder, J. L. (2014). The role of hybridization in a species invasion and extirpation of resident fauna: hybrid vigor and breakdown in the rusty crayfish, *Orconectes rusticus. Journal of Crustacean Biology*, 34(2), 157–164.

Ardura, A., Zaiko, A., Morán, P., Planes, S., & Garcia-Vazquez, E. (2017). Epigenetic signatures of invasive status in populations of marine invertebrates. *Scientific Reports*, 7, 42193.

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, 23(1), 26–32.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25.

Asif, J. H., & Krug, P. J. (2012). Lineage distribution and barriers to gene flow among populations of the globally invasive marine mussel *Musculista senhousia. Biological Invasions*, 14(7), 1431–1444.

Ayres, D. R., Grotkopp, E., Zaremba, K., Sloop, C. M., Blum, M. J., Bailey, J. P., et al. (2008). Hybridization between invasive *Spartina densiflora* (Poaceae) and native *S. foliosa* in San Francisco Bay, California, USA. *American Journal of Botany*, 95(6), 713–719.

Bachtrog, D., & Andolfatto, P. (2006). Selection, recombination and demographic history in *Drosophila miranda. Genetics*, 174(4), 2045–2059.

Baker, H. G. (1955). Self-compatibility and establishment after “long-distance” dispersal. *Evolution*, 9(3), 347–349.

Baker, H. G. (1965). Characteristics and modes of origin of weeds. In: *The Genetics of Colonizing Species* (eds Baker H, Stebbins G), pp. 147–172. Academic Press, New York.

Baker, H. G. (1974). The evolution of weeds. *Annual Review of Ecology and Systematics*, 5(1), 1–24.

Barker, B. S., Andonian, K., Swope, S. M., Luster, D. G., & Dlugosch, K. M. (2017). Population genomic analyses reveal a history of range expansion and trait evolution across the native and invaded range of yellow starthistle (*Centaurea solstitialis*). *Molecular Ecology*, 26(4), 1131–1147.

Barreto, F. S., Moy, G. W., & Burton, R. S. (2010). Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. *Molecular Ecology*, 20(3), 560–572.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38–44.

Barrón, M. G., Fiston-Lavier, A.-S., Petrov, D. A., & González, J. (2014). Population genomics of transposable elements in Drosophila. *Annual Review of Genetics*, 48, 561–581.

Bay, R. A., & Palumbi, S. R. (2014). Multilocus adaptation associated with heat resistance in reef-building corals. *Current Biology*, 24(24), 2952–2956.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41, 379–406.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. Genetics, 162(4), 2025–2035.

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

Bernardi, G., Azzurro, E., Golani, D., & Miller, M. R. (2016). Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. *Molecular Ecology*, 25(14), 3384–3396.

Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, 19(13), 2609–2625.

Bierne, N., & Eyre-Walker, A. (2004). The genomic rate of adaptive amino acid substitution in *Drosophila*. *Molecular Biology and Evolution*, 21(7), 1350–1360.

Bierne, N., Bonhomme, F., & Arnaud Haond, S. (2016). Editorial: Dedicated population genomics for the silent world: the specific questions of marine population genetics. *Current Zoology*, 62(6), 545–550.

Bierne, N., Bonhomme, F., & David, P. (2003a). Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1522), 1399–1406.

Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., & David, P. (2003b). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology*, 12(2), 447–461.

Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1591), 1253–1260.

Bierne, N., David, P., Boudry, P., & Bonhomme, F. (2002). Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution*, 56(2), 292–298.

Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? why are FST outliers sometimes so frequent? *Molecular Ecology*, 22(8), 2061–2064.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20(10), 2044–2072.

Blackburn, T. M., Lockwood, J. L., & Cassey, P. (2015). The influence of numbers on invasion success. *Molecular Ecology*, 24(9), 1942–1953.

Blum, M. G., & François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, 20(1), 63–73.

Blum, M. J., Walters, D. M., Burkhead, N. M., Freeman, B. J., & Porter, B. A. (2010). Reproductive isolation and the expansion of an invasive hybrid swarm. *Biological Invasions*, 12(8), 2825–2836.

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., et al. (2015). What we still don't know about invasion genetics. *Molecular Ecology*, 24(9), 2277–2297.

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

Boon, E., Faure, M. F., & Bierne, N. (2009). The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. *Journal of Molecular Evolution*, 68(5), 461–474.

Borsa, P., Daguin, C., & Bierne, N. (2007). Genomic reticulation indicates mixed ancestry in Southern-Hemisphere *Mytilus spp*. mussels. *Biological Journal of the Linnean Society*, 92(4), 747–754.

Bouchemousse, S., Liautard Haag, C., Bierne, N., & Viard, F. (2016). Distinguishing contemporary hybridization from past introgression with postgenomic ancestry-informative SNP s in strongly differentiated *Ciona* species. *Molecular Ecology*, 25(21), 5527–5542.

Bourne, S. D., Hudson, J., Holman, L. E., & Rius, M. (2018). Marine invasion genomics: Revealing ecological and evolutionary consequences of biological invasions. In: *Population Genetics: Marine Organisms, Population Genomics*. pp 1-36.

Bownes, S. J., & McQuaid, C. D. (2006). Will the invasive mussel *Mytilus galloprovincialis* Lamarck replace the indigenous *Perna perna* L. on the south coast of South Africa? *Journal of Experimental Marine Biology and Ecology*, 338(1), 140–151.

Braby, C. E., & Somero, G. N. (2006a). Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus *Mytilus*). *Journal of Experimental Biology*, 209(13), 2554–2566.

Braby, C. E., & Somero, G. N. (2006b). Ecological gradients and relative abundance of native (*Mytilus trossulus*) and invasive (*Mytilus galloprovincialis*) blue mussels in the California hybrid zone. *Marine Biology*, 148(6), 1249–1262.

Branch, G. M., & Steffani, C. N. (2004). Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology*, 300(1-2), 189–215.

Brannock, P. M., Wethey, D. S., & Hilbish, T. J. (2009). Extensive hybridization with minimal introgression in *Mytilus galloprovincialis* and *M. trossulus* in Hokkaido, Japan. *Marine Ecology Progress Series*, 383, 161–171.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, 81(5), 1084–1097.

Brunetti, R., Gissi, C., Pennati, R., Caicci, F., Gasparini, F., & Manni, L. (2015). Morphological evidence that the molecularly determined *Ciona intestinalis* type A and type B are different species: *Ciona robusta* and *Ciona intestinalis*. *Journal of Zoological Systematics and Evolutionary Research*, 53(3), 186–193.

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., et al. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, 68(4), 935–949.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo Ferreira, J., Ballenghien, M., Weinert, L., et al. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, 12(5), 834–845.

Campagna, L., Gronau, I., Silveira, L. F., Siepel, A., & Lovette, I. J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, 24(16), 4238–4251.

Carlton, J. T., & Geller, J. B. (1993). Ecological roulette: the global transport of nonindigenous marine organisms. *Science*, 261(5117), 78–82.

Carlton, J. T. (1996). Pattern, process, and prediction in marine invasion ecology. *Biological Conservation*, 78(1), 97–106.

Carson, H. L. (1965). Chromosomal morphism in geographically widespread species of *Drosophila*. In: *The Genetics of Colonizing Species* (eds Baker H, Stebbins G), pp. 503–531. Academic Press, New York.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, 2(4), e64.

Charlesworth, J., & Eyre-Walker, A. (2008). The McDonald-Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution*, 25(6), 1007–1015.

Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *Evolutionary Applications*, 8(1), 23–46.

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, 26(1), 59–76.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.

Clark, A. G., Glanowski, S., Nielsen, R., Thomas, P., Kejariwal, A., Todd, M. J., et al. (2003). Positive selection in the human genome inferred from human–chimp–mouse orthologous gene alignments. In: *Cold Spring Harbor symposia on quantitative biology*, (Vol. 68, pp. 479–486). Cold Spring Harbor Laboratory Press.

Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24(9), 1999–2017.

Colgan, D. J., & Middelfart, P. (2011). *Mytilus* mitochondrial DNA haplotypes in southeastern Australia. *Aquatic Biology*, 12(1), 47–53.

Colhoun, E. A., Turner, E., & Van de Geer, G. (1982). Late Pleistocene marine molluscan faunas from four sites in Tasmania. In: *Papers and Proceedings of the Royal Society of Tasmania*. (Vol. 116, pp. 91–97).

Connor, K. M., & Gracey, A. Y. (2011). Circadian cycles are the dominant transcriptional rhythm in the intertidal mussel *Mytilus californianus*. *Proceedings of the National Academy of Sciences*, USA, 108, 16110-16115.

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1749), 5039–5047.

Corander, J., Majander, K. K., Cheng, L., & Merilä, J. (2013). High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology*, 22(11), 2931–2940.

Craft, J. A., Gilbert, J. A., Temperton, B., Dempsey, K. E., Ashelford, K., Tiwari, B., et al. (2010). Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns. *PLoS ONE*, 5(1), e8875.

Crandall, E. D., Toonen, R. J., Laboratory, T., & Selkoe, K. A. (2018). A coalescent sampler successfully detects biologically meaningful population structure overlooked by F-statistics. *Evolutionary Applications*. doi: 10.1111/eva.12712

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13), 3133–3157.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7), 410–418.

Csilléry, K., François, O., & Blum, M. G. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3(3), 475–479.

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. *Evolution*, 62(8), 1908–1920.

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

Daguin, C., & Borsa, P. (2000). Genetic relationships of *Mytilus galloprovincialis* Lamarck populations worldwide: evidence from nuclear-DNA markers. *Geological Society, London, Special Publications*, 177(1), 389–397.

Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D., & Milzani, A. (2009). Protein S-glutathionylation: a regulatory device from bacteria to humans. *Trends in Biochemical Sciences*, 34(2), 85–96.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.

Darling, J. A., Bagley, M. J., Roman, J., Tepolt, C. K., & Geller, J. B. (2008). Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, 17(23), 4992–5007.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879–886.

De Wit, P., Pespeni, M. H., & Palumbi, S. R. (2015). SNP genotyping and population genomics from expressed sequences–current advances and future possibilities. *Molecular Ecology*, 24(10), 2310–2323.

Dean, A. M., & Thornton, J. W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. *Nature Reviews. Genetics*, 8(9), 675.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., et al. (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*, 298(5601), 2157–2167.

Dias, P. J., Fotedar, S., & Snow, M. (2014). Characterisation of mussel (*Mytilus sp.*) populations in Western Australia and evaluation of potential genetic impacts of mussel spat translocation from interstate. *Marine and Freshwater Research*, 65(6), 486–496.

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.

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, 24(9), 2095–2111.

Dobzhansky, T. (1965). 'Wild' and "domestic" species of Drosophila. *In: The Genetics of Colonizing Species* (eds Baker H, Stebbins G), pp. 533–551. Academic Press, New York.

Dong, Y., & Somero, G. N. (2008). Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus *Lottia*): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. *Journal of Experimental Biology*, 212(2), 169–177.

Donner, J., & Jungner, H. (1981). Radiocarbon dating of marine shells from southeastern Australia as a means of dating relative sea-level changes. *Ann. Acad. Sci. Fenn., Geol. Geogr.*, 131, 1-44

Dowd, W. W., Felton, C. A., Heymann, H. M., Kost, L. E., & Somero, G. N. (2013). Food availability, more than body temperature, drives correlated shifts in ATP-generating and antioxidant enzyme capacities in a population of intertidal mussels (*Mytilus californianus*). *Journal of Experimental Marine Biology and Ecology*, 449, 171–185.

Duforet-Frebourg, N., Bazin, E., & Blum, M. G. (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, 31(9), 2483–2495.

Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E., & Blum, M. G. (2015). Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 Genomes data. *Molecular Biology and Evolution*, 33(4), 1082–1093.

Durand, C., Manuel, M., Boudouresque, C. F., Meinesz, A., Verlaque, M., & Le Parco, Y. (2002). Molecular data suggest a hybrid origin for the invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta) in the Mediterranean Sea. *Journal of Evolutionary Biology*, 15(1), 122–133.

Ellegren, H. (2008). Comparative genomics and the study of evolution by natural selection. *Molecular Ecology*, 17(21), 4586–4596.

Ellstrand, N. C., & Schierenbeck, K. A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences*, 97(13), 7043–7050.

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, 26(6), 298–306.

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

Estoup, A., Beaumont, M., Sennedot, F., Moritz, C., & Cornuet, J. M. (2004). Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, 58(9), 2021–2036.

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion? *Annual Review of Ecology, Evolution, and Systematics*, 47, 51–72.

Estoup, A., Wilson, I. J., Sullivan, C., Cornuet, J.-M., & Moritz, C. (2001). Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, 159(4), 1671–1687.

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501.

Eyre-Walker, A., & Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, 26(9), 2097–2108.

Fagundes, N. J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F. M., Bonatto, S. L., & Excoffier, L. (2007). Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, 104(45), 17614–17619.

Fay, J. C., Wyckoff, G. J., & Wu, C.-I. (2001). Positive and negative selection on the human genome. *Genetics*, 158(3), 1227–1234.

Feis, M. E., John, U., Lokmer, A., Luttikhuizen, P. C., & Wegner, K. M. (2018). Dual transcriptomics reveals co-evolutionary mechanisms of intestinal parasite infections in blue mussels *Mytilus edulis*. *Molecular Ecology*, 27(6), 1505–1519.

Felsenstein, J., & Churchill, G. A. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, 13(1), 93–104.

Ferrero, V., Barrett, S. C., Castro, S., Caldeirinha, P., Navarro, L., Loureiro, J., & Rodríguez Echeverría, S. (2015). Invasion genetics of the Bermuda buttercup (*Oxalis pescaprae*): complex intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native and introduced populations. *Molecular Ecology*, 24(9), 2143–2155.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4(4), 423–425.

Fields, P. A., Rudomin, E. L., & Somero, G. N. (2006). Temperature sensitivities of cytosolic malate dehydrogenases from native and invasive species of marine mussels (genus *Mytilus*): sequence-

function linkages and correlations with biogeographic distribution. *Journal of Experimental Biology*, 209(4), 656–667.

Fields, P. A., Dong, Y., Meng, X., & Somero, G. N. (2015). Adaptations of protein structure and function to temperature: there is more than one way to "skin a cat." *Journal of Experimental Biology*, 218(12), 1801–1811.

Fields, P. A., Zuzow, M. J., & Tomanek, L. (2012). Proteomic responses of blue mussel (*Mytilus*) congeners to temperature acclimation. *Journal of Experimental Biology*, 215(7), 1106–1116.

Fitzpatrick, B. M., & Shaffer, H. B. (2007). Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proceedings of the National Academy of Sciences*, 104(40), 15793–15798.

Fitzpatrick, B. M., Johnson, J. R., Kump, D. K., Smith, J. J., Voss, S. R., & Shaffer, H. B. (2010). Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences*, 107, 3606–3610.

Fly, E. K., & Hilbish, T. J. (2012). Physiological energetics and biogeographic range limits of three congeneric mussel species. *Oecologia*, 172(1), 35–46.

Fly, E. K., Hilbish, T. J., Wethey, D. S., & Rognstad, R. L. (2015). Physiology and biogeography: The response of European mussels (*Mytilus spp.*) to climate change. *American Malacological Bulletin*, 33(1), 136–149.

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* 180, 977–993

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217), 1258524.

Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2015). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology*, 25(1), 269–286.

Fraïsse, C., Roux, C., Gagnaire, P.-A., Romiguier, J., Faivre, N., Welch, J. J., & Bierne, N. (2018a). The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: the effects of sequencing techniques and sampling strategies. *Peer J*, 6, e5198.

Fraïsse, C., Haguenauer, A., Gerard, K., Weber, A. A.-T., Bierne, N., & Chenuil, A. (2018b). Fine-grained habitat-associated genetic connectivity in an admixed population of mussels in the small isolated Kerguelen island. *bioRxiv*, 239244.

Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene flow in a mosaic hybrid zone: is local introgression adaptive? *Genetics*, 197, 939-951.

Frankham, R. (2005). Resolving the genetic paradox in invasive species. *Heredity*, 94(4), 385–385.

Frantz, L. A., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature Genetics*, 47(10), 1141.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152.

Fu, Yun-Xin, & Li, W.-H. (1997). Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution*, 14(2), 195–199.

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

Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., et al. (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biology and Evolution*, 10(9), 2518–2534.

Gagnaire, P.-A., Normandeau, E., & Bernatchez, L. (2012). Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Molecular Biology and Evolution*, 29(10), 2909–2919.

Gaiti, F., Calcino, A. D., Tanurdžić, M., & Degnan, B. M. (2017). Origin and evolution of the metazoan non-coding regulatory genome. *Developmental Biology*, 427(2), 193–202.

Gaiti, F., Fernandez-Valverde, S. L., Nakanishi, N., Calcino, A. D., Yanai, I., Tanurdzic, M., & Degnan, B. M. (2015). Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. *Molecular Biology and Evolution*, 32(9), 2367–2382.

Gaiti, F., Degnan, B. M., & Tanurdžić, M. (2018a). Long non-coding regulatory RNAs in sponges and insights into the origin of animal multicellularity. *RNA Biology*, 01–22.

Galindo, J., Grahame, J. W., & Butlin, R. K. (2010). An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology*, 23(9), 2004–2016.

Galtier, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genetics*, 12(1), e1005774–23.

García-Ramos, G., & Rodríguez, D. (2002). Evolutionary speed of species invasions. *Evolution*, 56(4), 661–668.

Gardner, J. P., Zbawicka, M., Westfall, K. M., & Wenne, R. (2016). Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: the need for baseline data and enhanced protection in the Southern Ocean. *Global Change Biology*, 22(9), 3182–3195.

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., et al. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLoS Genetics*, 9(4), e1003457.

Geller, J. B. (1999). Decline of a native mussel masked by sibling species invasion. *Conservation Biology*, 13(3), 661–664.

Geller, J. B., Carlton, J. T., & Powers, D. A. (1994). PCR-based detection of mtDNA haplotypes of native and invading mussels on the northeastern Pacific coast: latitudinal pattern of invasion. *Marine Biology*, 119(2), 243–249.

Geller, J. B., Darling, J. A., & Carlton, J. T. (2010). Genetic perspectives on marine biological invasions. *Annual Review of Marine Science*, 2, 367–393.

Geller, J., Sotka, E. E., Kado, R., Palumbi, S. R., & Schwindt, E. (2008). Sources of invasions of a northeastern Pacific acorn barnacle, *Balanus glandula*, in Japan and Argentina. *Marine Ecology Progress Series*, 358, 211–218.

Gerdol, M., De Moro, G., Manfrin, C., Milandri, A., Riccardi, E., Beran, A., et al. (2014). RNA sequencing and de novo assembly of the digestive gland transcriptome in *Mytilus galloprovincialis* fed with toxinogenic and non-toxic strains of *Alexandrium minutum. BMC Research Notes*, 7(1), 722.

Gérard, K., Bierne, N., Borsa, P., Chenuil, A., & Féral, J.-P. (2008). Pleistocene separation of mitochondrial lineages of *Mytilus spp*. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. *Molecular Phylogenetics and Evolution*, 49(1), 84–91.

Gilchrist, G. W., Jeffers, L. M., West, B., Folk, D. G., Suess, J., & Huey, R. B. (2008). Clinal patterns of desiccation and starvation resistance in ancestral and invading populations of *Drosophila subobscura. Evolutionary Applications*, 1(3), 513–523.

Gilg, M. R., Kirby, S. E., Sullivan, R., Knapp, L. W., & Hilbish, T. J. (2007). Dispersal vs. retention: correspondence of species-specific reproductive cycles and settlement periods in a blue mussel hybrid zone. *Marine Ecology Progress Series*, 351, 151–161.

Gleason, L. U., Miller, L. P., Winnikoff, J. R., Somero, G. N., Yancey, P. H., Bratz, D., & Dowd, W. W. (2017). Thermal history and gape of individual *Mytilus californianus* correlate with oxidative damage and thermoprotective osmolytes. *Journal of Experimental Biology*, 220(22), 4292–4304.

Glotzbecker, G. J., Walters, D. M., & Blum, M. J. (2016). Rapid movement and instability of an invasive hybrid swarm. *Evolutionary Applications*, 9(6), 741–755.

Gosling, E. (1992). The mussel *Mytilus*: ecology, physiology, genetics and culture. Elsevier Amsterdam.

Gosling, E. M., & McGrath, D. (1990). Genetic variability in exposed-shore mussels, *Mytilus spp*., along an environmental gradient. *Marine Biology*, 104(3), 413–418.

Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high FST outlier loci within a mussel species. *Journal of Evolutionary Biology*, 26(1), 14–26.

Goubert, C., Henri, H., Minard, G., Valiente Moro, C., Mavingui, P., Vieira, C., & Boulesteix, M. (2017). High-throughput sequencing of transposable element insertions suggests adaptive evolution of the invasive Asian tiger mosquito towards temperate environments. *Molecular Ecology*, 26(15), 3968–3981.

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

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., & Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644.

Gracey, A. Y., Chaney, M. L., Boomhower, J. P., Tyburczy, W. R., Connor, K., & Somero, G. N. (2008). Rhythms of gene expression in a fluctuating intertidal environment. *Current Biology*, 18(19), 1501–1507.

Grant, W. S., & Cherry, M. I. (1985). *Mytilus galloprovincialis* Lmk. in southern Africa. *Journal of Experimental Marine Biology and Ecology*, 90(2), 179–191.

Griffiths, C. L., Hockey, P., Van Erkom Schurink, C., & Le Roux, P. J. (1992). Marine invasive aliens on South African shores: implications for community structure and tropillc functioning. *South African Journal of Marine Science*, 12(1), 713–722.

Grime, J. P. (1977). Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist*, 111(982), 1169–1194.

Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26(20), 5362–5368.

Guillemaud, T., Beaumont, M. A., Ciosi, M., Cornuet, J.-M., & Estoup, A. (2010). Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, 104(1), 88.

Guzinski, J., Ballenghien, M., Daguin Thiébaut, C., Lévêque, L., & Viard, F. (2018). Population genomics of the introduced and cultivated Pacific kelp *Undaria pinnatifida*: Marinas—not farms—drive regional connectivity and establishment in natural rocky reefs. *Evolutionary Applications*, 11(9), 1582–1597.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-

seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494.

Haddrill, P. R., Bachtrog, D., & Andolfatto, P. (2008). Positive and negative selection on noncoding DNA in *Drosophila simulans. Molecular Biology and Evolution*, 25(9), 1825–1834.

Haerty, W., & Ponting, C. P. (2013). Mutations within lncRNAs are effectively selected against in fruitfly but not in human. *Genome Biology*, 14(5), R49.

Hamilton, G., Stoneking, M., & Excoffier, L. (2005). Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proceedings of the National Academy of Sciences*, 102(21), 7476–7480.

Hamilton, J. A., Okada, M., Korves, T., & Schmitt, J. (2015). The role of climate adaptation in colonization success in *Arabidopsis thaliana. Molecular Ecology*, 24(9), 2253–2263.

Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspective. In: *Hybrid Zones and the Evolutionary Process,* 3–12. Oxford University Press, Oxford.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25(11), 2454–2466.

Hart, M. W. (2012). Next-generation studies of mating system evolution. *Evolution*, 66(6), 1675–1680.

Hart, M. W. (2014). Models of selection, isolation, and gene flow in speciation. *The Biological Bulletin*, 227(2), 133–145.

Hart, M. W., Sunday, J. M., Popovic, I., Learning, K. J., & Konrad, C. M. (2014). Incipient speciation of sea star populations by adaptive gamete recognition coevolution. *Evolution*, 68(5), 1294–1305.

Haslbeck, M., Franzmann, T., Weinfurtner, D., & Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nature Structural & Molecular Biology*, 12(10), 842–846.

Hayes, K., Sliwa, C., Migus, S., McEnnulty, F., & Dunstan, P. (2005). National priority pests. *Part II. Ranking of Australian Marine Pests*. Independent Report for Department of Environment and Heritage, Australian Government Department of Environment and Heritage, Canberra, Australia.

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22(11), 2898–2916.

Hey, J., & Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics*, 167(2), 747–760.

Hey, J., & Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, 104(8), 2785–2790.

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports*, 11(7), 1110–1122.

Hilbish, T. J., & Koehn, R. K. (1985). The physiological basis of natural selection at the LAP locus. *Evolution*, 39(6), 1302–1317.

Hilbish, T. J., Brannock, P. M., Jones, K. R., Smith, A. B., Bullock, B. N., & Wethey, D. S. (2010). Historical changes in the distributions of invasive and endemic marine invertebrates are contrary to global warming predictions: the effects of decadal climate oscillations. *Journal of Biogeography*, 37(3), 423–431.

Hilbish, T. J., Lima, F. P., Brannock, P. M., Fly, E. K., Rognstad, R. L., & Wethey, D. S. (2012). Change and stasis in marine hybrid zones in response to climate warming. *Journal of Biogeography*, 39(4), 676–687.

Hilbish, T. J., Mullinax, A., Dolven, S. I., Meyer, A., Koehn, R. K., & Rawson, P. D. (2000). Origin of the antitropical distribution pattern in marine mussels (*Mytilus spp*.): routes and timing of transequatorial migration. *Marine Biology*, 136(1), 69–77.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., et al. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist,* 188(4), 379–397.

Hodgins, K. A., & Rieseberg, L. (2011). Genetic differentiation in life-history traits of introduced and native common ragweed (*Ambrosia artemisiifolia*) populations. *Journal of Evolutionary Biology*, 24(12), 2731–2749.

Hodgins, K. A., Bock, D. G., Hahn, M. A., Heredia, S. M., Turner, K. G., & Rieseberg, L. H. (2014). Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Molecular Ecology*, 24(9), 2226–2240.

Hodgins, K. A., Lai, Z., Nurkowski, K., Huang, J., & Rieseberg, L. H. (2013). The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, 22(9), 2496–2510.

Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Rieseberg, L. H., & Aitken, S. N. (2016). Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Molecular Biology and Evolution*, 33(6), 1502–1516.

Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, 97(3), 222.

Hofmann, G. E., & Somero, G. N. (1996). Interspecific variation in thermal denaturation of proteins in the congeneric mussels *Mytilus trossulus* and *M. galloprovincialis*: evidence from the heat-shock response and protein ubiquitination. *Marine Biology*, 126(1), 65–75.

Holland, B. S. (2000). Genetics of marine bioinvasions. *Hydrobiologia*, 420(1), 63–71.

Holloway, A. K., Lawniczak, M. K. N., Mezey, J. G., Begun, D. J., & Jones, C. D. (2007). Adaptive gene expression divergence inferred from population genomics. *PLoS Genetics*, 3(10), e187–7.

Hope, J. H., Lampert, R. J., Edmondson, E., Smith, M. J., & Van Tets, G. F. (1977). Late Pleistocene faunal remains from Seton rock shelter, Kangaroo Island, South Australia. *Journal of Biogeography*, 363–385.

Hovick, S. M., & Whitney, K. D. (2014). Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. *Ecology Letters*, 17(11), 1464–1477.

Hovick, S. M., Campbell, L. G., Snow, A. A., & Whitney, K. D. (2011). Hybridization alters early life-history traits and increases plant colonization success in a novel region. *The American Naturalist*, 179(2), 192–203.

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9), 868–877.

Huey, R. B., Gilchrist, G. W., Carlson, M. L., Berrigan, D., & Serra, L. (2000). Rapid evolution of a geographic cline in size in an introduced fly. *Science*, 287(5451), 308–309.

Hunt, B. G., Ometto, L., Keller, L., & Goodisman, M. A. (2012). Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Molecular Biology and Evolution*, 30(2), 263–271.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267.

Ibáñez, I., Diez, J. M., Miller, L. P., Olden, J. D., Sorte, C. J., Blumenthal, D. M., et al. (2014). Integrated assessment of biological invasions. *Ecological Applications*, 24(1), 25–37.

Jeffery, N. W., Bradbury, I. R., Stanley, R. R., Wringe, B. F., Van Wyngaarden, M., Lowen, J. B., et al. (2018). Genome-wide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. *Evolutionary Applications*, 11, 869–882.

Johannesson, K. (1988). The paradox of Rockall: why is a brooding gastropod (*Littorina saxatilis*) more widespread than one having a planktonic larval dispersal stage (*L. littorea*)? *Marine Biology*, 99(4), 507–513.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55.

Jones, J. B., & Creeper, J. (2006). Diseases of pearl oysters and other molluscs: A Western Australian perspective. *Journal of Shellfish Research*, 25(1), 233–238.

Jong, C. J., Azuma, J., & Schaffer, S. (2012). Mechanism underlying the antioxidant activity of taurine: prevention of mitochondrial oxidant production. *Amino Acids*, 42(6), 2223–2232.

Jordan, G., & Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution*, 29(4), 1125–1139.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, 17(18), 4015–4026.

Kane, N. C., King, M. G., Barker, M. S., Raduski, A., Karrenberg, S., Yatabe, Y., et al. (2009). Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*, 63(8), 2061–2075.

Kane, N. C., Barker, M. S., Zhan, S. H., & Rieseberg, L. H. (2011). Molecular evolution across the Asteraceae: micro-and macroevolutionary processes. *Molecular Biology and Evolution*, 28(12), 3225–3235.

Keightley, P. D., & Eyre-Walker, A. (2012). Estimating the rate of adaptive molecular evolution when the evolutionary divergence between species is small. *Journal of Molecular Evolution*, 74(1-2), 61–68.

Keinan, A., Mullikin, J. C., Patterson, N., & Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, 39(10), 1251.

Keller, S. R., & Taylor, D. R. (2010). Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology*, 23(8), 1720–1731.

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, 9(1), e86175.

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501.

Kirkpatrick, M., & Barrett, B. (2015). Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Molecular Ecology*, 24(9), 2046–2055.

Klopfstein, S., Currat, M., & Excoffier, L. (2005). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3), 482–490.

Kocot, K. M., Citarella, M. R., Moroz, L. L., & Halanych, K. M. (2013). PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics*, 9, EBO–S12813.

Koester, J. A., Swanson, W. J., & Armbrust, E. V. (2013). Positive Selection within a diatom species acts on putative protein interactions and transcriptional regulation. *Molecular Biology and Evolution*, 30(2), 422–434.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 35(suppl_2), W345–W349.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics*, 22(24), 3096–3098.

Kosiol, C., Vinař, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., & Siepel, A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS Genetics*, 4(8), e1000144–17.

Kuhner, M. K. (2009). Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution*, 24(2), 86–93.

la Fuente, de, M., Valera, S., & Martínez-Guitarte, J. L. (2012). ncRNAs and thermoregulation: a view in prokaryotes and eukaryotes. *FEBS Letters*, 586(23), 4061–4069.

la Rosa, de, J., & Stipanuk, M. H. (1985). Evidence for a rate-limiting role of cysteine-sulfinate decarboxylase activity in taurine biosynthesis in vivo. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 81(3), 565–571.

Ladner, J. T., Barshis, D. J., & Palumbi, S. R. (2012). Protein evolution in two co-occurring types of Symbiodinium: an exploration into the genetic basis of thermal tolerance in Symbiodinium clade D. *BMC Evolutionary Biology*, 12(1), 217.

Lai, Z., Kane, N. C., Kozik, A., Hodgins, K. A., Dlugosch, K. M., Barker, M. S., et al. (2012). Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. *American Journal of Botany*, 99(2), 209–218.

Lai, Z., Kane, N. C., Zou, Y., & Rieseberg, L. H. (2008). Natural variation in gene expression between wild & weedy populations of *Helianthus annuus*. *Genetics*, 179, 1881–1890.

Lamichhaney, S., Barrio, A. M., Rafati, N., Sundström, G., Rubin, C.-J., Gilbert, E. R., et al. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, 109(47), 19345–19350.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359.

Larraín, M. A., Zbawicka, M., Araneda, C., Gardner, J. P., & Wenne, R. (2018). Native and invasive taxa on the Pacific coast of South America: Impacts on aquaculture, traceability and biodiversity of blue mussels (*Mytilus spp.*). *Evolutionary Applications*, 11(3), 298–311.

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, 21(7), 1548–1566.

Le Roux, J., & Wieczorek, A. M. (2009). Molecular systematics and population genetics of biological invasions: towards a better understanding of invasive species management. *Annals of Applied Biology*, 154(1), 1–17.

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics*, 12(1), 124.

Lee, C. E. (1999). Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, 53(5), 1423–1434.

Lee, C. E. (2016). Evolutionary mechanisms of habitat invasions, using the copepod *Eurytemora affinis* as a model system. *Evolutionary Applications*, 9(1), 248–270.

Lee, C. E., Kiergaard, M., Gelembiuk, G. W., Eads, B. D., & Posavi, M. (2011). Pumping ions: rapid parallel evolution of ionic regulation following habitat invasions. *Evolution*, 65(8), 2229–2244.

Lee, C. E., Remfert, J. L., & Gelembiuk, G. W. (2003). Evolution of physiological tolerance and performance during freshwater invasions. *Integrative and Comparative Biology*, 43(3), 439–449.

Lee, K. M., & Coop, G. (2017). Distinguishing among modes of convergent adaptation using population genomic data. *Genetics,* 117, 300-417.

Lee, S. Y., & Morton, B. (1985). The introduction of the Mediterranean mussel *Mytilus galloprovincialis* into Hong Kong. *Malacological Review*, 18(1-2), 107–109.

Lemos, B., Bettencourt, B. R., Meiklejohn, C. D., & Hartl, D. L. (2005). Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Molecular Biology and Evolution*, 22(5), 1345–1354.

Lessios, H. A. (2007). Reproductive isolation between species of sea urchins. *Bulletin of Marine Science*, 81(2), 191–208.

Lessios, H. A., Kessing, B. D., & Pearse, J. S. (2001). Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema. Evolution*, 55(5), 955–975.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659.

Li, Y. F., Costello, J. C., Holloway, A. K., & Hahn, M. W. (2008). "Reverse ecology" and the power of population genomics. *Evolution*, 62(12), 2984–2994.

Liao, M.-L., Zhang, S., Zhang, G.-Y., Chu, Y.-M., Somero, G. N., & Dong, Y.-W. (2017). Heat-resistant cytosolic malate dehydrogenases (cMDHs) of thermophilic intertidal snails (genus *Echinolittorina*): Protein underpinnings of tolerance to body temperatures reaching 55 C. *Journal of Experimental Biology*, 220(11), 2066–2075.

Limborg, M. T., Helyar, S. J., De Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., et al. (2012). Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, 21(15), 3686–3703.

Lindberg, D. R. (1991). Marine biotic interchange between the northern and southern hemispheres. *Paleobiology*, 17(3), 308–324.

Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, *325*(5944), 1095-1098.

Lockwood, B. L., Sanders, J. G., & Somero, G. N. (2010). Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. *Journal of Experimental Biology*, 213(20), 3548–3558.

Lockwood, B. L., & Somero, G. N. (2011a). Invasive and native blue mussels (genus *Mytilus*) on the California coast: the role of physiology in a biological invasion. *Journal of Experimental Marine Biology and Ecology*, 400(1-2), 167–174.

Lockwood, B. L., & Somero, G. N. (2011b). Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Molecular Ecology*, 20(3), 517–529.

Lockwood, B. L., & Somero, G. N. (2012). Functional determinants of temperature adaptation in enzymes of cold- versus warm-adapted mussels (Genus *Mytilus*). *Molecular Biology and Evolution*, 29(10), 3061–3070.

Lockwood, B. L., Connor, K. M., & Gracey, A. Y. (2015). The environmentally tuned transcriptomes of *Mytilus* mussels. *Journal of Experimental Biology*, 218(12), 1822–1833.

Lockwood, J. L., Cassey, P., & Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, 20(5), 223–228.

Lockwood, J. L., Cassey, P., & Blackburn, T. M. (2009). The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions*, 15(5), 904–910.

Lombaert, E., Guillemaud, T., Thomas, C. E., Lawson Handley, L. J., Li, J., Wang, S., et al. (2011). Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird *Harmonia axyridis*. *Molecular Ecology*, 20(22), 4654–4670.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031–1046.

Lowe, S., Browne, M., Boudjelas, S., & De Poorter, M. (2000). 100 of the world's worst invasive alien species: a selection from the global invasive species database (Vol. 12). Invasive Species Specialist Group Auckland.

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, 320(5883), 1632–1635.

Lu, A., & Guindon, S. (2013). Performance of standard and stochastic branch-site models for detecting positive selection among coding sequences. *Molecular Biology and Evolution*, 31(2), 484–495.

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.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences*, 104(suppl 1), 8597–8604.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963.

Mallick, S., Gnerre, S., Muller, P., & Reich, D. (2009). The difficulty of avoiding false positives in genome scans for natural selection. *Genome Research*, 19(5), 922–933.

Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, 26(9), 448–456.

Marko, P. B., & Hart, M. W. (2012). Retrospective coalescent methods and the reconstruction of metapopulation histories in the sea. *Evolutionary Ecology*, 26(2), 291–315.

Markova-Raina, P., & Petrov, D. (2011). High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Research*, 21(6), 863–874.

Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. Genetics, 206, 429–438.

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, 15(suppl_1), R17–R29.

Mayr, E. (1954). Geographic speciation in tropical echinoids. *Evolution*, 8(1), 1–18.

McDonald, J. H., & Koehn, R. K. (1988). The mussels *Mytilus galloprovincialis* and *M. trossulus* on the Pacific coast of North America. Marine Biology, 99(1), 111–118.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. Nature, 351(6328), 652.

McDonald, J. H., Seed, R., & Koehn, R. K. (1991). Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. *Marine Biology*, 111(3), 323–333.

McGuire, M. J., Lipsky, P. E., & Thiele, D. L. (1992). Purification and characterization of dipeptidyl peptidase I from human spleen. *Archives of Biochemistry and Biophysics*, 295(2), 280–288.

McQuaid, C. D., & Phillips, T. E. (2000). Limited wind-driven dispersal of intertidal mussel larvae: in situ evidence from the plankton and the spread of the invasive species *Mytilus galloprovincialis* in South Africa. *Marine Ecology Progress Series*, 201, 211–220.

Mendes, F. K., & Hahn, M. W. (2016). Gene Tree discordance causes apparent substitution rate variation. *Systematic Biology*, 65(4), 711–721.

Mesgaran, M. B., Lewis, M. A., Ades, P. K., Donohue, K., Ohadi, S., Li, C., & Cousens, R. D. (2016). Hybridization can facilitate species invasions, even without enhancing local adaptation. *Proceedings of the National Academy of Sciences*, 113(36), 10210–10214.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 28(11), 659–669.

Michalek, K., Ventura, A., & Sanders, T. (2016). *Mytilus* hybridisation and impact on aquaculture: a mini review. *Marine Genomics*, 27, 3–7.

Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, 6(9), 485–492.

Moyers, B. T., & Rieseberg, L. H. (2013). Divergence in gene expression is uncoupled from divergence in coding sequence in a secondarily woody sunflower. *International Journal of Plant Sciences*, 174(7), 1079–1089.

Mugal, C. F., Wolf, J. B. W., & Kaj, I. (2013). Why time matters: Codon evolution and the temporal dynamics of dN/dS. *Molecular Biology and Evolution*, 31(1), 212–231.

Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., & Canchaya, C. (2016). A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis. PLoS ONE*, 11(3), e0151561.

Narum, S. R., Gallardo, P., Correa, C., Matala, A., Hasselman, D., Sutherland, B. J., & Bernatchez, L. (2017). Genomic patterns of diversity and divergence of two introduced salmonid species in Patagonia, South America. *Evolutionary Applications*, 10(4), 402–416.

Natarajan, C., Hoffmann, F. G., Lanier, H. C., Wolf, C. J., Cheviron, Z. A., Spangler, M. L., et al. (2015). Intraspecific polymorphism, interspecific divergence, and the origins of function-altering mutations in deer mouse hemoglobin. *Molecular Biology and Evolution*, 32(4), 978–997.

Nei, M., & Kumar, S. (2000). Molecular evolution and phylogenetics. Oxford university press.

Nei, M., & Maruyama, T. (1975). Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics*, 80, 395.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29(1), 1–10.

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., & Bekkevold, D. (2009a). Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, 18(15), 3128–3150.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., et al. (2009b). Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, 9(1), 276.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., & White, T. J. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology*, 3(6), e170.

Nuzhdin, S. V., Wayne, M. L., Harmon, K. L., & McIntyre, L. M. (2004). Common pattern of evolution of gene expression level and protein sequence in *Drosophila. Molecular Biology and Evolution*, 21(7), 1308–1317.

Oduor, A. M. (2013). Evolutionary responses of native plant species to invasive plants: a review. *New Phytologist*, 200(4), 986–992.

Oliver, T. A., Garfield, D. A., Manier, M. K., Haygood, R., Wray, G. A., & Palumbi, S. R. (2010). Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome Biology and Evolution*, 2, 800–814.

Oyarzún, P. A., Toro, J. E., Cañete, J. I., & Gardner, J. P. (2016). Bioinvasion threatens the genetic integrity of native diversity and a natural hybrid zone: smooth-shelled blue mussels (*Mytilus spp*.) in the Strait of Magellan. *Biological Journal of the Linnean Society*, 117(3), 574–585.

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution*, 7(4), 114–118.

Palumbi, S. R., & Lessios, H. A. (2005). Evolutionary animation: How do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? *Proceedings of the National Academy of Sciences*, 102, 6566–6572.

Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics*, 22(1), 1–5.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M.-C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, 114(5), 450.

Paradis E., Claude J., Strimmer K., (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Pascual, M., Chapuis, M. P., Mestres, F., Balanya, J., Huey, R. B., Gilchrist, G. W., et al. (2007). Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Molecular Ecology*, 16(15), 3069–3083.

Patterson, N. J., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics*, 192, 1065-1093.

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Molecular Biology and *Evolution*, 29(10), 3237–3248.

Peischl, S., & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Molecular Ecology*, 24(9), 2084–2094.

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22(24), 5972–5982.

Perry, R. B.-T., & Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. *Development*, 143(21), 3882–3894.

Perry, W. L., Lodge, D. M., & Feder, J. L. (2002). Importance of hybridization between indigenous and nonindigenous freshwater species: an overlooked threat to North American biodiversity. *Systematic Biology*, 51(2), 255–275.

Pespeni, M. H., & Palumbi, S. R. (2013). Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology*, 22(13), 3580–3597.

Pérez, J. E., Nirchio, M., Alfonsi, C., & Muñoz, C. (2006). The biology of invasions: the genetic adaptation paradox. Biological Invasions, 8(5), 1115–1121.

Phillips, B. L., Brown, G. P., & Shine, R. (2010). Evolutionarily accelerated invasions: the rate of dispersal evolves upwards during the range advance of cane toads. *Journal of Evolutionary Biology*, 23(12), 2595–2601.

Phillips, B. L., Brown, G. P., Webb, J. K., & Shine, R. (2006). Invasion and the evolution of speed in toads. *Nature*, 439(7078), 803.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967.

Plough, L. V. (2016). Genetic load in marine animals: a review. *Current Zoology*, 62(6), 567–579.

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J. S., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 443(7108), 167.

Popovic, I., Marko, P. B., Wares, J. P., & Hart, M. W. (2014). Selection and demographic history shape the molecular evolution of the gamete compatibility protein bindin in *Pisaster* sea stars. *Ecology and Evolution*, 4(9), 1567–1588.

Preisser, E. L., Lodge, A. G., Orwig, D. A., & Elkinton, J. S. (2008). Range expansion and population dynamics of co-occurring invasive herbivores. *Biological Invasions*, 10(2), 201–213.

Prentis, P. J., Wilson, J. R., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13(6), 288–294.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5(3), e9490.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., & Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12), 1791–1798.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575.

Puzey, J., & Vallejo-Marín, M. (2014). Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Molecular Ecology*, 23(18), 4472–4485.

Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nature Reviews. *Genetics*, 17(1), 47.

Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A., & Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & Development*, 30(2), 191–207.

Rawson, P. D., & Hilbish, T. J. (1995). Distribution of male and female mtDNA lineages in populations of blue mussels, *Mytilus trossulus* and *M. galloprovincialis*, along the Pacific coast of North America. *Marine Biology*, 124(2), 245–250.

Rawson, P. D., & Hilbish, T. J. (1998). Asymmetric introgression of mitochondrial DNA among European populations of blue mussels (*Mytilus spp.*). *Evolution*, 52(1), 100–108.

Rawson, P. D., Agrawal, V., & Hilbish, T. J. (1999). Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Marine Biology*, 134(1), 201–211.

Rawson, P. D., Joyner, K. L., Meetze, K., & Hilbish, T. J. (1996). Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity*, 77(6), 599.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489.

Renaut, S., Owens, G. L., & Rieseberg, L. H. (2014). Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, 23(2), 311–324.

Richards, C. L., Bossdorf, O., Muth, N. Z., Gurevitch, J., & Pigliucci, M. (2006). Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters*, 9(8), 981–993.

Richardson, D. M. (2011). Fifty years of invasion ecology: the legacy of Charles Elton. John Wiley & Sons.

Rieseberg, L. H., Kim, S.-C., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C., & Clay, K. (2007). Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, 129(2), 149–165.

Riginos, C., Hickerson, M. J., Henzler, C. M., & Cunningham, C. W. (2004). Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution*, 58(11), 2438–2451.

Riginos, C., & Cunningham, C. W. (2005). Invited review: local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, 14(2), 381–400.

Riginos, C., Crandall, E. D., Liggins, L., Bongaerts, P., & Treml, E. A. (2016). Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Current Zoology*, 62(6), 581–601.

Riley, S. P., Bradley Shaffer, H., Randal Voss, S., & Fitzpatrick, B. M. (2003). Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. *Ecological Applications*, 13(5), 1263–1275.

Ripley, B., Venables, W., & Ripley, M. B. (2016). Package "nnet." R Package Version, 7–3.

Riquet, F., Daguin Thiébaut, C., Ballenghien, M., Bierne, N., & Viard, F. (2013). Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusc *Crepidula fornicata*. *Molecular Ecology*, 22(4), 1003–1018.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, 29(4), 233–242.

Rius, M., Turon, X., Bernardi, G., Volckaert, F. A., & Viard, F. (2015a). Marine invasion genetics: from spatio-temporal patterns to evolutionary outcomes. *Biological Invasions*, 17(3), 869–885.

Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015b). Applications of next-generation sequencing to the study of biological invasions. *Current Zoology*, 61(3), 488–504.

Robinson, T. B., Griffiths, C. L., Branch, G. M., & Govender, A. (2007). The invasion and subsequent die-off of *Mytilus galloprovincialis* in Langebaan Lagoon, South Africa: effects on natural communities. *Marine Biology*, 152(2), 225–232.

Rockman, M. V. (2008). Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature*, 456(7223), 738.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 66(1), 1–17.

Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F., & Lapegue, S. (2013). Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas. Evolutionary Applications*, 6(7), 1064–1078.

Rollins, L. A., Richardson, M. F., & Shine, R. (2015). A genetic perspective on rapid evolution in cane toads (*Rhinella marina*). *Molecular Ecology*, 24(9), 2264–2276.

Roman, J., & Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions. Trends in Ecology & Evolution, 22(9), 454–464.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261.

Rosenblum, E. B., Hickerson, M. J., & Moritz, C. (2007). A multilocus perspective on colonization accompanied by selection and gene flow. *Evolution*, 61(12), 2971–2985.

Ross-Ibarra, J., Wright, S. I., Foxe, J. P., Kawabe, A., DeRose-Wilson, L., Gos, G., et al. (2008). Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE*, 3(6), e2411.

Rougemont, Q., & Bernatchez, L. (2018). The demographic history of Atlantic Salmon (*Salmo salar*) across its distribution range reconstructed from Approximate Bayesian Computations. *Evolution*.

Roure, B., Rodriguez-Ezpeleta, N., & Philippe, H. (2007). SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evolutionary Biology*, 7(1), S2.

Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G. H., & Bierne, N. (2014a). Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27(8), 1662–1675.

Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., & Keller, L. (2014b). Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, 31(7), 1661–1685.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14(12), e2000234.

Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013). Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution,* 30(7), 1574–1587.

Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology*, 24(18), 4723–4738.

Saarman, N. P., Kober, K. M., Simison, W. B., & Pogson, G. H. (2017). Sequence-based analysis of thermal adaptation and protein energy landscapes in an invasive blue mussel (*Mytilus galloprovincialis*). *Genome Biology and Evolution*, 9(10), 2739–2751.

Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3, 509–535.

Sanjuan, A., Zapata, C., & Alvarez, G. (1997). Genetic differentiation in *Mytilus galloprovincialis* Lmk. throughout the world. *Ophelia*, 47(1), 13–31.

Sarver, S. K., & Foltz, D. W. (1993). Genetic population structure of a species' complex of blue mussels (*Mytilus spp.*). *Marine Biology*, 117(1), 105–112.

Schaffer, S. W., Jong, C. J., Ramila, K. C., & Azuma, J. (2010). Physiological roles of taurine in heart and muscle. *Journal of Biomedical Science*, 17(1), S2.

Schierenbeck, K. A., & Ellstrand, N. C. (2009). Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions*, 11(5), 1093.

Schneider, K. R., & Helmuth, B. (2007). Spatial variability in habitat temperature may drive patterns of selection between an invasive and native mussel species. *Marine Ecology Progress Series*, 339, 157–167.

Schrader, L., & Schmitz, J. (2018). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, 1-13

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, 5, 5495.

Secor, C. L., Day, A. J., & Hilbish, T. J. (2001). Factors influencing differential mortality within a marine mussel (Mytilus spp.) hybrid population in southwestern England: reproductive effort and parasitism. *Marine Biology*, 138(4), 731–739.

Seed, R. (1992). Systematics evolution and distribution of mussels belonging to the genus *Mytilus*- an overview. *American Malacological Bulletin,* 9(2), 123–137.

Sherman, C., Lotterhos, K. E., Richardson, M. F., Tepolt, C. K., Rollins, L. A., Palumbi, S. R., & Miller, A. D. (2016). What are we missing about marine invasions? Filling in the gaps with evolutionary genomics. *Marine Biology*, 163(10), 198.

Shields, J. L., Heath, J. W., & Heath, D. D. (2010). Marine landscape shapes hybrid zone in a broadcast spawning bivalve: introgression and genetic structure in Canadian west coast *Mytilus. Marine Ecology Progress Series*, 399, 211–223.

Shine, R. (2010). The ecological impact of invasive cane toads (*Bufo marinus*) in Australia. *The Quarterly Review of Biology*, 85(3), 253–291.

Silva, A. L., & Wright, S. H. (1992). Integumental taurine transport in *Mytilus* gill: short-term adaptation to reduced salinity. *Journal of Experimental Biology*, 162(1), 265–279.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.

Simberloff, D. (2009). The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 81–102.

Simberloff, D. (2013). Biological invasions: much progress plus several controversies. *Contributions to Science*, 7–16.

Simon-Bouhet, B., Garcia Meunier, P., & Viard, F. (2006). Multiple introductions promote range expansion of the mollusc *Cyclope neritea* (Nassariidae) in France: evidence from mitochondrial sequence data. *Molecular Ecology*, 15(6), 1699–1711.

Skibinski, D., Beardmore, J. A., & Cross, T. F. (1983). Aspects of the population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*, 19(2), 137–183.

Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, 393–430.

Sloop, C. M., Ayres, D. R., & Strong, D. R. (2009). The rapid evolution of self-fertility in *Spartina* hybrids (*Spartina alterniflora* × *foliosa*) invading San Francisco Bay, CA. *Biological Invasions*, 11(5), 1131–1144.

Smith, N. G., & Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila. Nature*, 415(6875), 1022–1024.

Somero, G. N. (2012). The physiology of global change: linking patterns to mechanisms. *Annual Review of Marine Science*, 4, 39–61.

Somero, G. N. (2018). RNA thermosensors: how might animals exploit their regulatory potential? *Journal of Experimental Biology*, 221(4), jeb162842.

Somero, G. N., Lockwood, B. L., & Tomanek, L. (2017). Biochemical adaptation: response to environmental challenges, from life's origins to the Anthropocene. Sinauer Associates, Incorporated Publishers.

Sorte, C. J., Ibáñez, I., Blumenthal, D. M., Molinari, N. A., Miller, L. P., Grosholz, E. D., et al. (2013). Poised to prosper? A cross-system comparison of climate change effects on native and non-native species performance. *Ecology Letters*, 16(2), 261–270.

Sorte, C. J., Williams, S. L., & Carlton, J. T. (2010a). Marine range shifts and species introductions: comparative spread rates and community impacts. *Global Ecology and Biogeography*, 19(3), 303–316.

Sorte, C. J., Williams, S. L., & Zerebecki, R. A. (2010b). Ocean warming increases threat of invasive species in a marine fouling community. *Ecology*, 91(8), 2198–2204.

Sotka, E. E., Baumgardner, A. W., Bippus, P. M., Destombe, C., Duermit, E. A., Endo, H., et al. (2018). Combining niche shift and population genetic analyses predicts rapid phenotypic evolution during invasion. *Evolutionary Applications*, 11(5), 781–793.

Sousa, V. M. C., Carneiro, M., Ferrand, N., & Hey, J. (2013). Identifying loci under selection against gene flow in isolation with migration models. *Genetics*, 194, 211-233.

Stachowicz, J. J., Terwin, J. R., Whitlatch, R. B., & Osman, R. W. (2002). Linking climate change and biological invasions: ocean warming facilitates nonindigenous species invasions. *Proceedings of the National Academy of Sciences*, 99(24), 15497–15500.

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, 24(9), 2241–2252.

Stebbins, G. L. (1985). Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden*, 824–832.

Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, 25(1), 79–88.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews. Genetics*, 14(11), 751.

Stern, D. L., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science*, 323(5915), 746–751.

Stewart, C. N., Tranel, P. J., Horvath, D. P., Anderson, J. V., Rieseberg, L. H., Westwood, J. H., et al. (2009). Evolution of weediness and invasiveness: charting the course for weed genomics. *Weed Science*, 57(5), 451–462.

Stoletzki, N., & Eyre-Walker, A. (2010). Estimation of the neutrality index. *Molecular Biology and Evolution*, 28(1), 63–70.

Storz, J. F., & Wheat, C. W. (2010). Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, 64(9), 2489–2509.

Storz, J. F., Bridgham, J. T., Kelly, S. A., & Garland, T., Jr. (2015). Genetic approaches in comparative and evolutionary physiology. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 309(3), R197–R214.

Storz, J. F., Runck, A. M., Sabatino, S. J., Kelly, J. K., Ferrand, N., Moriyama, H., et al. (2009). Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences*, 106(34), 14450–14455.

Strong, D. R., & Ayres, D. R. (2013). Ecological and evolutionary misadventures of *Spartina*. *Annual Review of Ecology, Evolution, and Systematics*, 44, 389–410.

Śmietanka, B., Burzyński, A., Hummel, H., & Wenne, R. (2014). Glacial history of the European marine mussels *Mytilus*, inferred from distribution of mitochondrial DNA lineages. *Heredity*, 113(3), 250.

Tepolt, C. K. (2015). Adaptation in marine invasion: a genetic perspective. *Biological Invasions*, 17(3), 887–903.

Tepolt, C. K., & Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology*, 24(16), 4145–4158.

Tepolt, C. K., & Somero, G. N. (2014). Master of all trades: thermal acclimation and adaptation of cardiac function in a broadly distributed marine invasive species, the European green crab, *Carcinus maenas*. *Journal of Experimental Biology*, 217(7), 1129–1138.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7(8), e41732.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S., et al. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5, 5770.

Tirosh, I., & Barkai, N. (2008). Evolution of gene sequence and gene expression are not correlated in yeast. *Trends in Genetics*, 24(3), 109–113.

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., et al. (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7), 892–908.

Tomanek, L. (2012). Environmental proteomics of the mussel *Mytilus*: Implications for tolerance to stress and change in limits of biogeographic ranges in response to climate change. *Integrative and Comparative Biology*, 52(5), 648–664.

Tomanek, L. (2015). Proteomic responses to environmentally induced oxidative stress. *Journal of Experimental Biology*, 218(12), 1867–1879.

Tomanek, L., & Zuzow, M. J. (2010). The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. *Journal of Experimental Biology*, 213(20), 3559–3574.

Tomanek, L., Zuzow, M. J., Hitt, L., Serafini, L., & Valenzuela, J. J. (2012). Proteomics of hyposaline stress in blue mussel congeners (genus *Mytilus*): implications for biogeographic range limits in response to climate change. *Journal of Experimental Biology*, 215(22), 3905–3916.

Turner, K. G., Hufbauer, R. A., & Rieseberg, L. H. (2014). Rapid evolution of an invasive weed. *New Phytologist*, 202(1), 309–321.

Turner, T. L., Bourne, E. C., Wettberg, Von, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, 42(3), 260.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews. Genetics*, 17(10), 601.

Uller, T., & Leimu, R. (2011). Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology*, 17(11), 3478–3485.

Vandepitte, K., De Meyer, T., Helsen, K., Van Acker, K., Roldán Ruiz, I., Mergeay, J., & Honnay, O. (2014). Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology*, 23(9), 2157–2164.

Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology*, 25(1), 89–103.

Verhoeven, K. J., Vonholdt, B. M., & Sork, V. L. (2016). Epigenetics in ecology and evolution: what we know and what we need to know. *Molecular Ecology*, 25(8), 1631–1638.

Vernot, B., & Akey, J. M. (2015). Complex history of admixture between modern humans and Neandertals. *The American Journal of Human Genetics*, 96(3), 448–453.

Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: three lessons from the past, and the way forward. *Current Zoology*, 62(6), 629–642.

Voolstra, C. R., Sunagawa, S., Matz, M. V., Bayer, T., Aranda, M., Buschiazzo, E., DeSalvo, M. K., Lindquist, E., Szmant, A. M., & Coffroth, M. A. (2011). Rapid evolution of coral proteins responsible for interaction with the environment. *PLoS ONE*, 6(5), e20392.

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89(5), 438–450.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15, 1419–1439.

Warton, D. I., Duursma, R. A., Falster, D. S., & Taskinen, S. (2012). smatr 3–an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution*, 3(2), 257–259.

Washietl, S., Kellis, M., & Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Research*, 24, 616-628.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370.

Wellband, K. W., & Heath, D. D. (2017). Plasticity in gene transcription explains the differential performance of two invasive fish species. *Evolutionary Applications*, 10(6), 563–576.

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2014). RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, 32(3), 820–832.

Westfall, K. M., & Gardner, J. P. (2010). Genetic diversity of Southern hemisphere blue mussels (Bivalvia: Mytilidae) and the identification of non-indigenous taxa. *Biological Journal of the Linnean Society*, 101(4), 898–909.

Westfall, K. M., & Gardner, J. P. (2013). Interlineage *Mytilus galloprovincialis* Lmk. 1819 hybridization yields inconsistent genetic outcomes in the Southern hemisphere. *Biological Invasions*, 15(7), 1493–1506.

Westram, A. M., Galindo, J., Alm Rosenblad, M., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, 23(18), 4603–4616.

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 FST. *The American Naturalist*, 186(S1), S24–S36.

Wilkins, N. P., Fujino, K., & Gosling, E. M. (1983). The Mediterranean mussel *Mytilus galloprovincialis* Lmk. in Japan. *Biological Journal of the Linnean Society*, 20(4), 365–374.

Winwood-Smith, H. S., Alton, L. A., Franklin, C. E., & White, C. R. (2015). Does greater thermal plasticity facilitate range expansion of an invasive terrestrial anuran into higher latitudes? *Conservation Physiology*, 3(1).

Wonham, M. J. (2004). Mini-review: distribution of the Mediterranean mussel, *Mytilus galloprovincialis* (Bivalvia: Mytilidae), and hybrids in the northeast Pacific. *Journal of Shellfish Research*, 23(2), 535–544.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15(1), 323–354.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851–865.

Yancey, P. H., & Siebenaller, J. F. (2015). Co-evolution of proteins and solutions: protein adaptation versus cytoprotective micromolecules and their roles in marine organisms. *Journal of Experimental Biology*, 218(12), 1880–1896.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591.

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4), 1107–1118.

Zayed, A., & Whitfield, C. W. (2008). A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences*, 105(9), 3421–3426.

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22(12), 2472–2479.

Supporting Information

Chapter 2

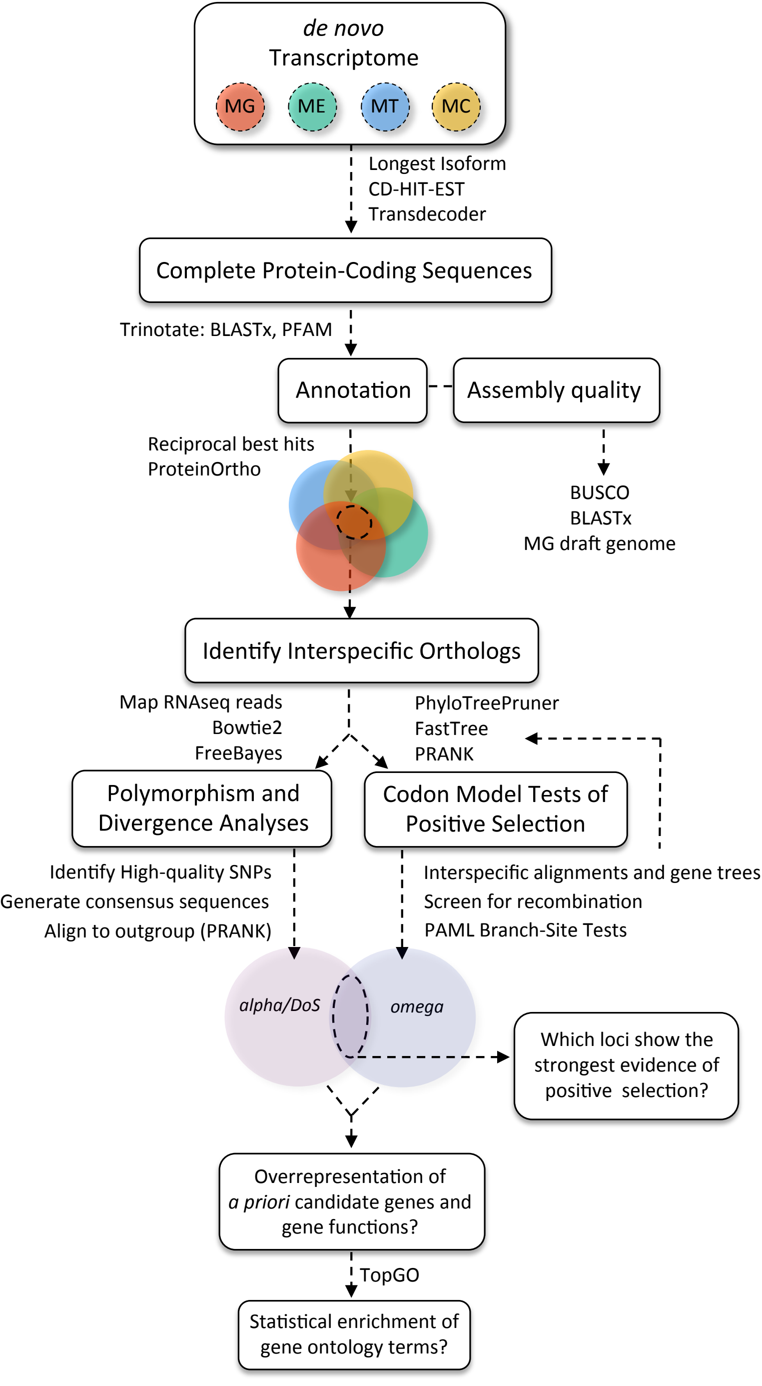


Figure S2 - 1. Conceptual summary of transcriptome sequence data preparation and statistical analyses. MC: *M. californianus*; MT: *M. trossulus*; ME: *M. edulis*; MG: *M. galloprovincialis.*

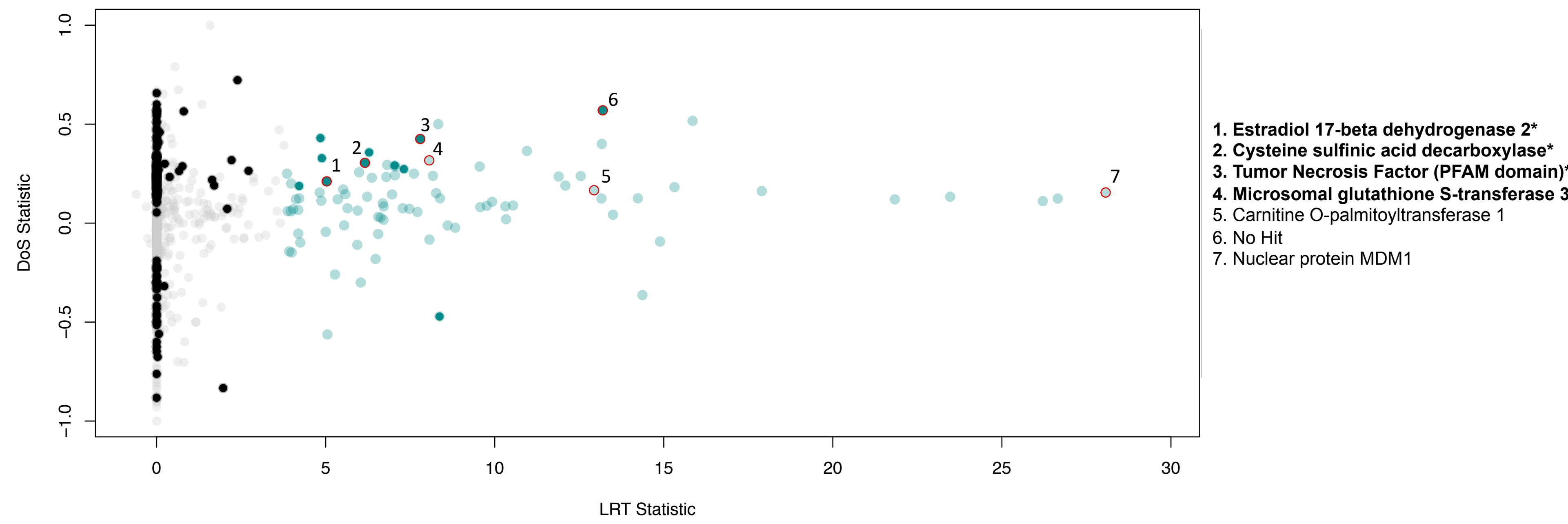


Figure S2 - 2. Strength of selection profiles for orthologous gene groups. Plot shows the relationship between per-gene likelihood ratio test (LRT) statistic and *DoS* statistic calculated for the Atlantic *M. galloprovincialis* population only. Significant LRT values are indicated in cyan. Significant *DoS* values are solid circles. Seven genes showing the strongest signatures of positive selection are outlined in red. Four positively selected genes corresponding to the 'highest-confidence alignments' are marked with an **asterisk.**

Table S2 - 1. Reference assembly (*de novo*) statistics for transcriptome sequencing data

Taxon	No. Processed Paired Read[a]	No. Normalized Paired Read	Trinity Assembly Statistics					
			Transcripts	Unique gene groups	N50 Length (bp)	Total Assembly Length (bp)	Mean Contig Length (bp)	Max Contig Length (bp)
M. galloprovincialis	20,462,657	7,568,056	164,075	126,895	1,313	131,613,630	802	16,625
M. edulis	34,854,901	7,229,075	142,112	113,219	1,135	106,889,258	752	17,358
M. trossulus	35,119,386	12,374,570	239,958	172,553	1,285	1,889,27,068	787	19,647
M. californianus	10,229,559	3,764,081	107,420	95,500	990	75,314,010	701	11,299

[a]Total number of quality-filtered paired reads from three sequenced individuals per taxon

Table S2 - 2. Filtering statistics for protein-coding sequence data

Taxon	Transcripts (Longest Isoform)	CD-Hit: Non-redundant Transcript No.	Transdecoder: No. Transcripts with ORFs			
			Complete	Partial (3' 5')	Internal	Total ORFs
M. galloprovincialis	126,895	102,716	11,370	12,230	14,934	38,534
M. edulis	113,219	90,170	9,660	12,139	14,693	36,492
M. trossulus	172,553	130,389	13,681	14,202	16,211	44,094
M. californianus	95,500	80,523	7,046	10,560	13,081	30,687

Table S2 - 3. BLAST statistics for complete ORF gene sets queried against Uniprot-Swissprot protein database ($e=10^{-3}$)

Taxon	Transcripts queried	Significant matches (non-contaminant)	Significant matches (contaminant)
M. galloprovincialis	11,370	8188	381
M. edulis	9,660	7244	815
M. trossulus	13,681	9583	478
M. californianus	7,046	5282	354

Table S2 - 4. BLAST statistics for gene sets queried against *M. galloprovincialis* (Mg) draft genome* ($e=10^{-3}$; Murgarella et al. 2016)

Taxon	Gene set	Transcripts queried	Count of unique contigs with significant hit to Mg genome	Count of unique genome reference sequences with significant hit to assembly
M. galloprovincialis	Trinity transcripts[a]	126,895	119,101	70,912
M. galloprovincialis	Complete ORFs	10,989	10,706	10,324

[a]Longest isoforms only

Table S2 - 5. Summary of assembly quality assessment of protein coding gene sets using proportion of 429 eukaryotic BUSCO orthologs

Taxon	% BUSCOs (complete)	% BUSCOs (complete+partial)	No. BUSCOs		
			Complete	Fragmented	Missing
M. galloprovincialis	82.7%	89.3%	355	28	46
M. edulis	84.8%	91.6%	364	29	36
M. trossulus	83.9%	89.7%	360	25	44
M. californianus	74.1%	87.8%	318	59	52

Table S2 - 6. Enrichment of Gene Ontology (GO) terms among positively selected gene sets at a significance level of $p \leq 0.001$. Analyses were conducted against the annotated set of 2719 orthogroups analysed in this study as the reference gene set.

Analysis	Selected Gene List	GO ID	Biological Process Term	P-value	Number of Significant Annotated Genes
1. Branch-site tests (PAML)					
1.a	LRT p≤0.05 (n=99)	GO:0006082	*organic acid metabolic process*	4.1e-05	2*
1.b	BEB p≥0.95 (n=38)	GO:0006082	*organic acid metabolic process*	1.8e-05	2*
2. Positive *alpha* and *DoS*					
		GO:0000375	*RNA splicing, via transesterification reaction*	4.5e-05	4
		GO:0022618	*ribonucleoprotein complex assembly*	8.3e-05	3
		GO:0002090	*regulation of receptor internalization*	1.7e-04	2*
		GO:0006082	*organic acid metabolic process*	3.5e-04	2
		GO:0006364	*rRNA processing*	7.0e-04	3

* Cysteine sulfinic acid decarboxylase (*CSAD*) and Carnitine O-palmitoyltransferase 1 (*CPT1A*) are annotated with significantly enriched GO terms.

Chapter 3

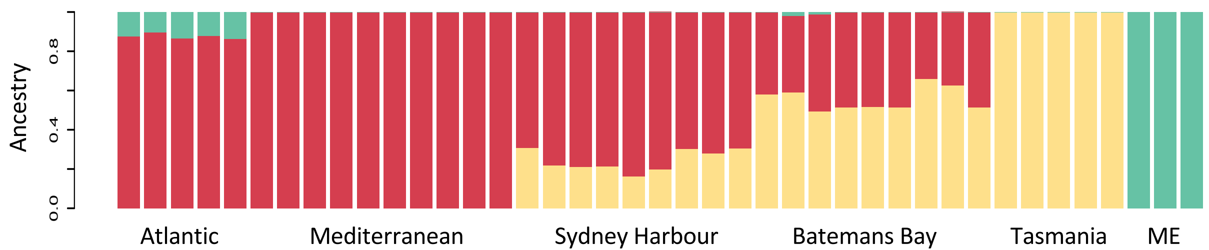


Figure S3 - 1. ADMIXTURE analyses for K=3 genetic clusters performed using only 1 SNP per contig to account for linkage effects. Each bar represents an individual belonging to one or more ancestral clusters, corresponding to different colours.

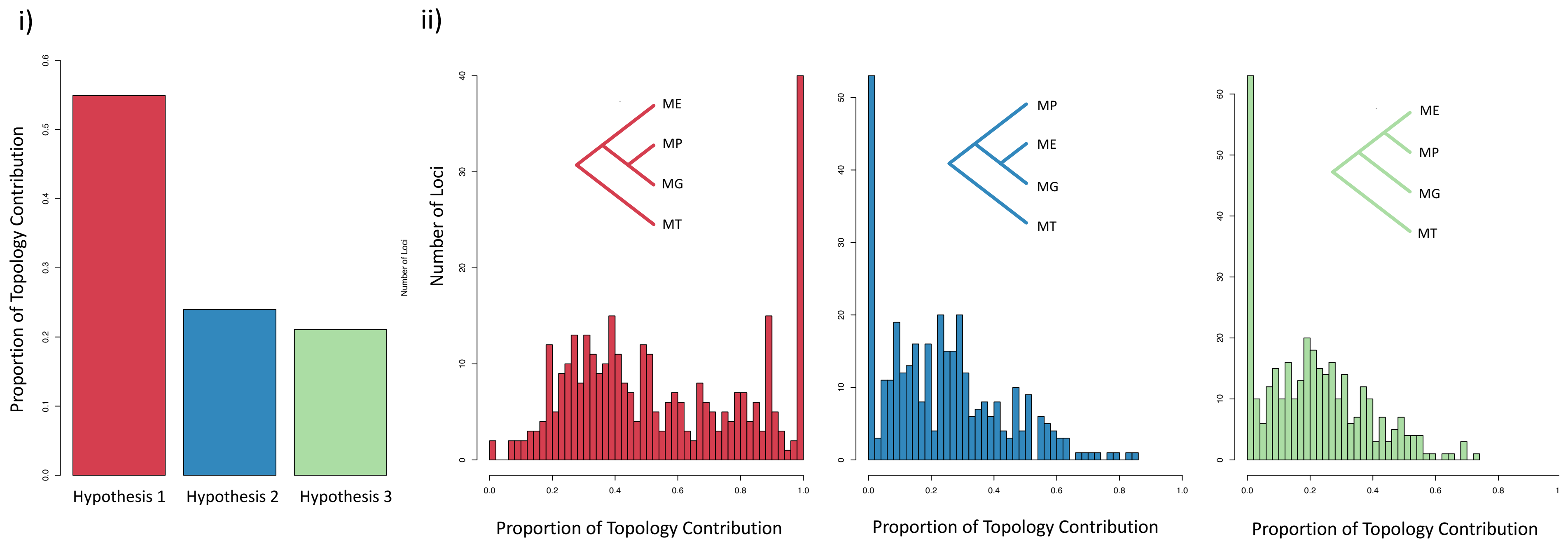


Figure S3 - 2. Summary of TWISST analyses of species relationships. Contributions of three possible unrooted topologies (hypotheses) to the nuclear species tree grouping *M. planulatus* (MP) with either *M. trossulus* (MT), *M. edulis* (ME) or *M. galloprovincialis* (MG). Plots indicate **i)** relative contributions of 343 topologies genealogies with a minimum tree length of 0.025; and **ii)** Distribution of the proportion of topology contributions for three tested topologies.

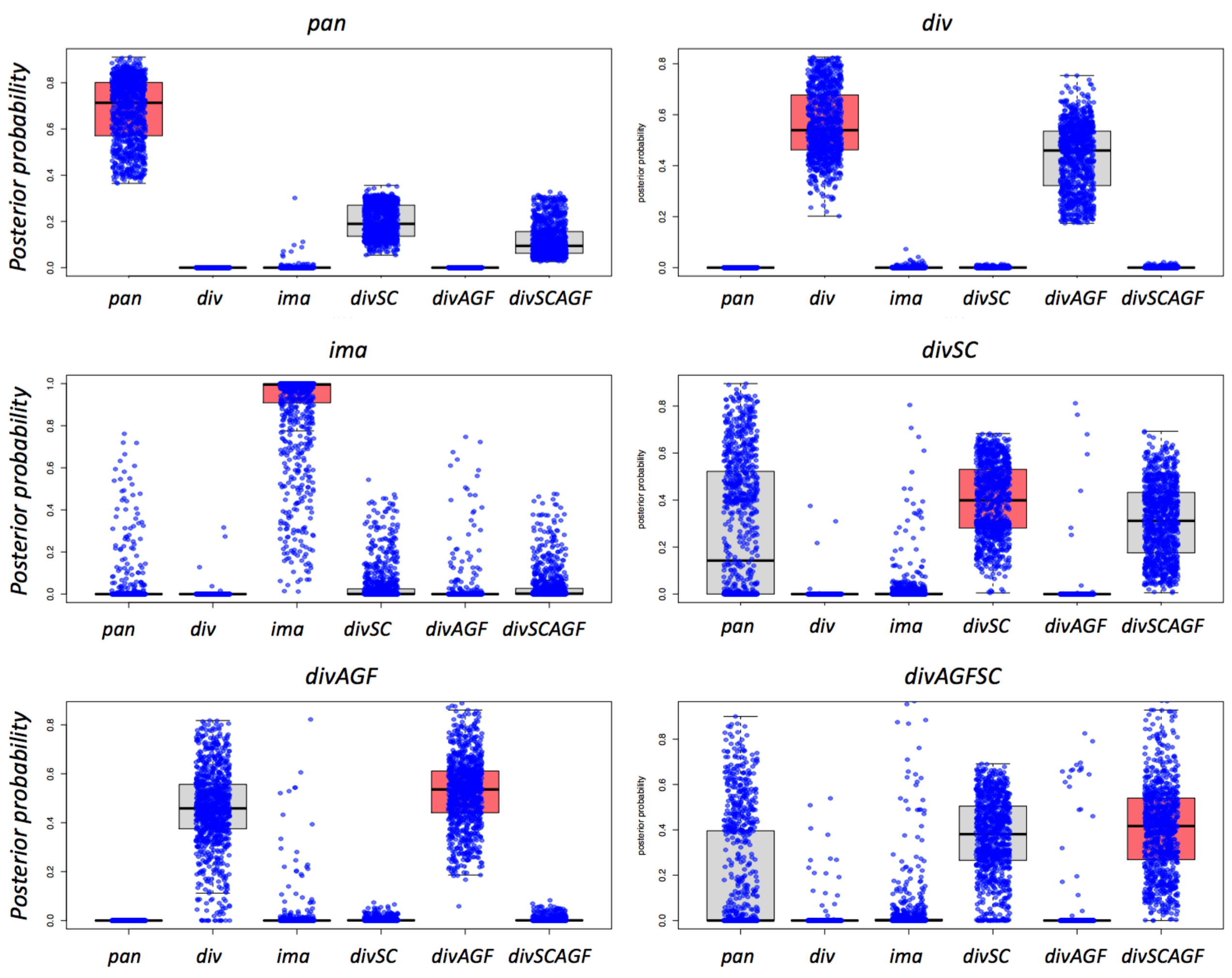


Figure S3 - 3. Summary of model choice validation using pseudo-observed datasets (PODS). Plot indicates distribution of posterior probabilities for each true model in model comparisons using 1000 PODS generated under the same model (shown in pink box plot).

Table S3 - 1. Summary of prior distribution lower and upper bounds for 13 parameters (for each demographic model) shown in generation units (generation time=2 years). Effective populations sizes of derived (N_i , N_j) and ancestral ($N_{ancestral}$) populations; migration rate (*m*), where m_{ij} is the proportion of migrants from population *j* into population *i*; ancient migration rate (*ma*); time of secondary contact (T_{sc}); time of the onset of ancient gene flow (T_{nc}; backwards in time); divergence time (T_{div}).

	Demographic Model Priors: Lower bound – Upper bound						
Parameter	***pan***	***div***	***im***	***divSC***	***divAGF***	***divSCAGF***	***divSCGP***
N_i N_j $N_{ancestral}$	1000-500000	1000-500000	1000-500000	1000-500000	1000-500000	1000-500000	1000-500000
m_{ij}	-	-	0-0.0001	0-0.5	-	0-0.5	0-0.5
m_{ji}	-	-	0-0.0001	0	-	0	0
m_{aij}	-	-	-	-	0-0.0001	0-0.0001	-
m_{aji}	-	-	-	-	0-0.0001	0-0.0001	-
T_{sc}	-	-	-	5-300	-	5-300	5-300
T_{nc}	-	-	-	-	10000-1750000	10000-1750000	-
T_{div}	-	100000-1750000	100000-1750000	100000-1750000	100000-1750000	100000-1750000	100000-1750000
ri	-	-	-	0.01-0.99	-	0.01-0.99	0.01-0.99
mu	0.00000002763	0.00000002763	0.00000002763	0.00000002763	0.00000002763	0.00000002763	0.00000002763
rho	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table S3 - 2. Summary of stepwise comparison Akaike Information Criterion (AIC) between sequential migration models in TreeMix. We did not consider additional migration events when the difference between nested models was less than two ($\Delta AIC < 2$).

Number of migration events	Ln(likelihood)	AIC value	Δ AIC
0	-1552.53	3109.06	-
1	-361.959	727.918	2381.142
2	174.862	-345.724	1073.642
3	215.341	-426.682	80.958
4	218.552	-433.104	6.422
5	219.126	-434.252	1.148
6	219.483	-434.966	0.714
7	219.483	-434.966	0
8	219.483	-434.966	0
9	219.483	-434.966	0
10	219.483	-434.966	0

Table S3 - 3. Summary of model validation using pseudo-observed datasets. Values are shown for the **A)** Mediterranean-Tasmania and **B)** Atlantic-Tasmania populations pairs. Values for the best inferred model is indicated in **bold.**

A)

Model	Precision	Misclassification rate (Type I error)	Mean Type II error
pan	1.000	0.000	0.145
div	0.618	0.382	0.0776
ima	0.933	0.067	0.0088
divSC	0.390	0.610	0.0678
divAGF	**0.607**	**0.393**	**0.0826**
divSCAGF	0.364	0.636	0.0354

B)

Model	Precision	Misclassification rate (Type I error)	Mean Type II error
pan	1.000	0.000	0.1408
div	0.708	0.292	0.812
ima	0.948	0.052	0.0076
divSC	0.422	0.578	0.0618
divAGF	**0.587**	**0.413**	**0.0614**
divSCAGF	0.397	0.603	0.0348

Table S3 - 4. Summary of demographic model selection under an approximate Bayesian computation framework. Model posterior probabilities comparing all homogeneous and heterogeneous models (accounting for variation in N_e | m parameters) under each demographic scenario independently; **Bold** indicates the highest probability model for each comparison at an acceptance threshold of 0.001.

Population			**Demographic Model Probability: Proportion of accepted simulations**					
Mediterranean-Tasmania	**Model**	**Method**	**hom|hom**	**hom|het1**	**het1|hom**	**het1|het1**	**het2|hom**	**het2|het1**
	div	Neural Net	0.0038	-	0.1170	-	**0.8792**	-
	im	Neural Net	0.0023	0.0471	0.0014	0.0825	0.0030	**0.8637**
	divSC	Neural Net	0.0191	0.1048	0.0196	0.1491	0.0199	**0.6875**
	divAGF	Neural Net	0.0013	0.0020	0.0259	0.0421	**0.6410**	0.2877
	divSCAGF	Neural Net	0.0217	0.0984	0.0201	0.2688	0.0246	**0.5664**
Atlantic-Tasmania								
	div	Neural Net	0.0001	-	0.0500	-	**0.9498**	**-**
	im	Neural Net	0.0009	0.0167	0.0016	0.0395	0.0042	**0.9369**
	divSC	Neural Net	0.0177	0.1129	0.0130	0.0674	0.0405	**0.7486**
	divAGF	Neural Net	0.0040	0.0398	0.0118	0.0265	0.4291	**0.4888**
	divSCAGF	Neural Net	0.0222	0.0974	0.0192	0.1168	0.0465	**0.6979**

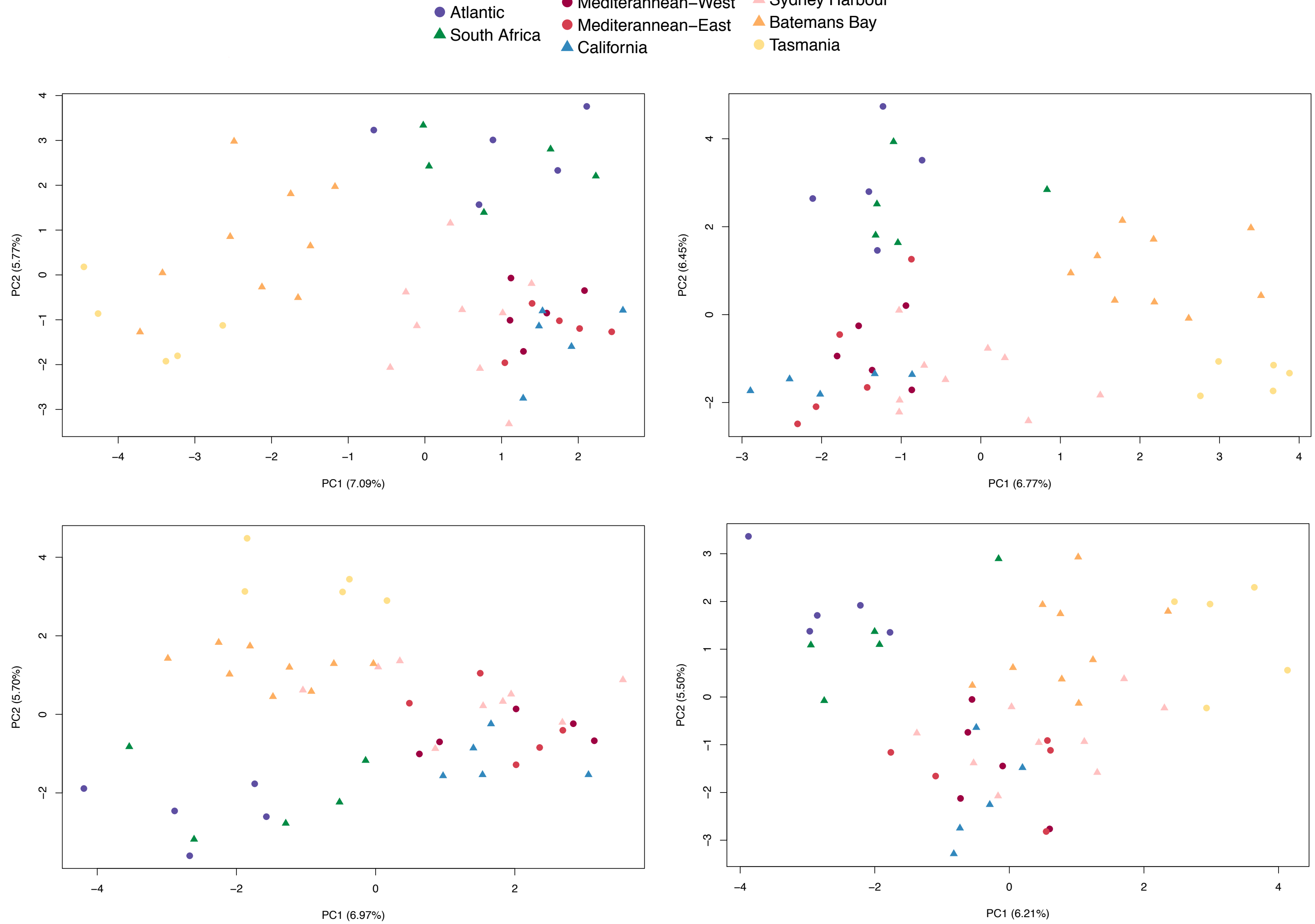


Figure S4 - 1. Results from four randomly chosen principal components analyses chosen out of 1000 randomly subsampled datasets of 471 SNPs (i.e. number of SNPs equivalent to the lncRNA dataset) from the full transcriptome assembly.

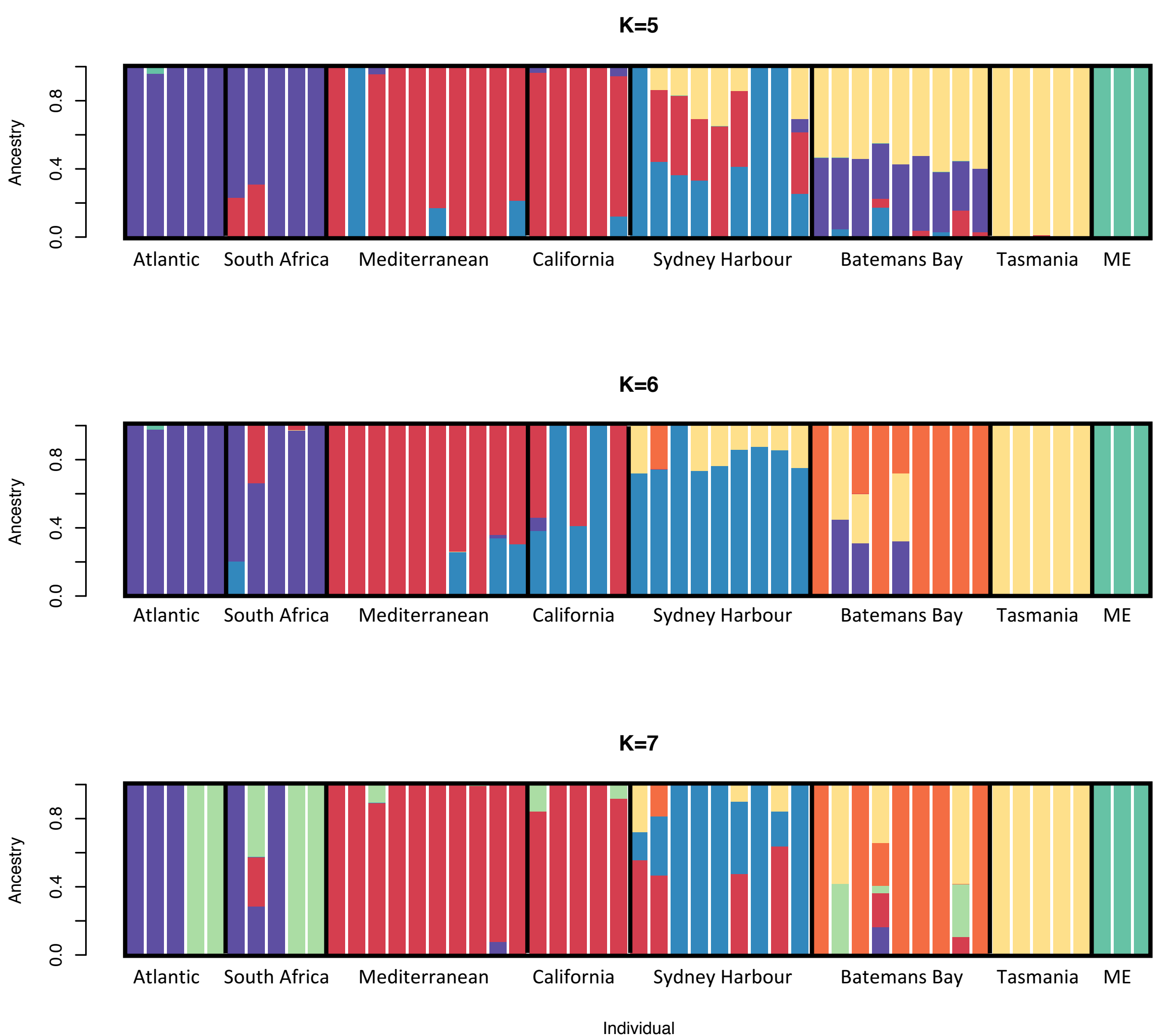


Figure S4 - 2. ADMIXTURE results for K=5-7 genetic clusters. Analyses did not recover differentiated genetic groupings for native and introduced populations from the Atlantic and South Africa, or for populations from the Mediterranean and California.