The Roles of Hybridisation, Contemporary Climate Change, and Recent Range Shifts on the Redistribution of Marine Biodiversity.

by

Jamie Hudson

ORCID ID 0000-0002-2294-1616

Thesis for the degree of Doctor of Philosophy

March 2020
Global change as a result of human activities has caused unprecedented alterations to biodiversity. One consequence of such alteration is the redistribution of species. Understanding the mechanisms that determine and maintain species’ ranges is at the forefront of ecological and evolutionary research. Current drivers of the redistribution of biodiversity include natural range expansions, anthropogenic transport of species, and contemporary climate change. These can alter both species ranges and evolutionary trajectories in multiple ways. A notorious example is when changes in species distributions bring divergent genotypes into secondary contact, facilitating hybridisation. Despite each of these topics being well studied individually, there remains a dearth of research studying these factors in combination to understand how they reshape biodiversity patterns in the marine environment. It is predicted that an integrative approach studying these factors in combination would enable further understanding how they interact to affect the redistribution of biodiversity.

This thesis used a multidisciplinary approach that combined population genomic data, controlled experimental crosses of divergent genotypes, and environmental datasets to investigate the role of different factors on past, current, and future changes in species distributions. My research has shown that whilst hybridisation can be beneficial for range expansions, this is not always the case. I found evidence that some species that have spread as a result of human mediated transport have undergone secondary contacts, facilitating colonisation of distant habitats. Conversely, other species with transoceanic distributions may be expected to undergo secondary contact due to anthropogenic transport and the presence of two native lineages, though I found no evidence that range expansion had been preceded by hybridisation. Finally, species that are able to currently hybridise may not perform as well as parental crosses under thermal stress, suggesting that under CCC, hybridisation may not be such a dominant driver of species redistribution.

Taken together, this thesis demonstrates the utility in using a multidisciplinary approach to study factors that can combine to determine species ranges. I have shown that different mechanisms can have dissimilar consequences on species ranges that can be difficult to predict, and that by studying multiple factors can one successfully disentangle the role each one plays in the redistribution of marine biodiversity.
# Table of Contents

Table of Contents.......................................................................................................................... i
Table of Tables................................................................................................................................. v
Table of Figures................................................................................................................................. vii
List of Accompanying Materials .................................................................................................... xi
Research Thesis: Declaration of Authorship .................................................................................. xiii
Acknowledgements......................................................................................................................... xv
Definitions and Abbreviations ......................................................................................................... xvii

## Chapter 1 Introduction ............................................................................................................... 1

1.1 Species redistribution .................................................................................................................. 1
  1.1.1 Natural range expansions ................................................................................................. 2
  1.1.2 Biological invasions ....................................................................................................... 3

1.2 Contemporary climate change ................................................................................................. 4

1.3 Hybridisation ............................................................................................................................ 5
  1.3.1 The costs associated with hybridisation ....................................................................... 5
  1.3.2 The benefits associated with hybridisation .................................................................. 6

1.4 Hybridisation and contemporary climate change leading to species redistributions.......... 9

1.5 Ascidians: a unique group to study changes in marine ecosystems .................................... 10
  1.5.1 Ecology of ascidians .................................................................................................... 12
  1.5.2 Ascidians as invaders .................................................................................................. 14

1.6 Thesis structure and author contributions ............................................................................. 15

## Chapter 2 Secondary contacts and genetic admixture shape colonization by an amphiatlantic epibenthic invertebrate ................................................................. 21

2.1 Abstract .................................................................................................................................... 21

2.2 Introduction ............................................................................................................................... 22

2.3 Materials and Methods ............................................................................................................ 25
  2.3.1 Field sampling .............................................................................................................. 25
  2.3.2 DNA extraction and genotyping .................................................................................. 28
  2.3.3 Analysis of genotyping-by-sequencing data ................................................................. 28
  2.3.4 Population structure ...................................................................................................... 29
  2.3.5 Reconstructing invasion routes ..................................................................................... 30
Chapter 3  Using oceanographic and genomic data to study regionally constrained distributions of non-indigenous species

3.1 Abstract ................................................................................................................. 41
3.2 Introduction ............................................................................................................ 42
3.3 Methods .................................................................................................................. 44
  3.3.1 Study species and distribution ........................................................................... 44
  3.3.2 Sampling ............................................................................................................ 45
  3.3.3 DNA extraction and genotyping ....................................................................... 46
  3.3.4 Analysis of GBS data ....................................................................................... 47
  3.3.5 Population structure ......................................................................................... 48
  3.3.6 Reconstructing invasion routes ........................................................................ 49
  3.3.7 Environmental matching ............................................................................... 50
3.4 Results .................................................................................................................... 51
  3.4.1 Processing of raw GBS data ............................................................................. 51
  3.4.2 Global genomic diversity indices ..................................................................... 51
  3.4.3 Evolutionary history of the native range ........................................................ 52
  3.4.4 Population structure across all sample sites .................................................... 53
  3.4.5 Reconstructing invasion routes ....................................................................... 54
  3.4.6 Environmental matching ............................................................................... 55
  3.4.7 Fine-scale population genomic analysis within the introduced range .......... 57
3.5 Discussion .............................................................................................................. 58
  3.5.1 Evolutionary history within Australia ............................................................. 59
  3.5.2 Invasion history .............................................................................................. 59
  3.5.3 Patterns within the introduced range ............................................................... 61
  3.5.4 Conclusions and future directions ................................................................. 62
3.6 Acknowledgements .............................................................................................. 62
# Table of Contents

## Chapter 4  Contemporary climate change hinders hybrid performance of ecologically dominant marine invertebrates

4.1 Abstract .................................................................................................................. 65
4.2 Introduction ........................................................................................................... 66
4.3 Methods .................................................................................................................. 69
  4.3.1 Study species ................................................................................................. 69
  4.3.2 Field sampling ............................................................................................... 70
  4.3.3 Sea surface temperature data ......................................................................... 70
  4.3.4 Laboratory housing of animals ...................................................................... 71
  4.3.5 Fertilization methods .................................................................................... 72
  4.3.6 Data analysis and statistics ........................................................................... 74
4.4 Results .................................................................................................................... 74
  4.4.1 Temperature results ...................................................................................... 74
  4.4.2 Development at in situ temperature .............................................................. 75
  4.4.3 Development at experimental treatment temperatures .............................. 76
  4.4.4 Effects of temperature and cross on pre-metamorphic development .......... 77
  4.4.5 Effects of temperature and cross on post-metamorphic development .......... 78
4.5 Discussion ............................................................................................................. 80
4.6 Acknowledgements ............................................................................................... 84

## Chapter 5  Genomic signatures of local adaptation to different depths in a dominant marine invertebrate

5.1 Abstract .................................................................................................................. 85
5.2 Introduction ........................................................................................................... 85
5.3 Methods .................................................................................................................. 88
  5.3.1 Sampling ....................................................................................................... 88
  5.3.2 DNA extraction and genotyping .................................................................... 90
  5.3.3 Identification of outlier loci .......................................................................... 90
  5.3.4 Identification of outlier loci associated to population structure ................... 90
  5.3.5 Identification of outlier loci associated with environmental variables .......... 91
  5.3.6 Population genomic analyses ...................................................................... 92
  5.3.7 Annotation of outlier loci ............................................................................. 93
  5.3.8 Laboratory housing of animals, fertilisation methods and data analysis of experimental crosses .................................................................................. 93
Table of Contents

5.4  Results .........................................................................................................................95
5.4.1 Loci assembly ............................................................................................................95
5.4.2 Identification of outlier loci ......................................................................................96
5.4.3 Genetic structuring inferred by outlier loci ...............................................................97
5.4.4 Functional annotation of outlier loci ........................................................................102
5.4.5 Laboratory crosses ....................................................................................................105

5.5  Discussion .....................................................................................................................107
5.5.1 Population genomic data and differences between depths .......................................107
5.5.2 Laboratory crosses ....................................................................................................109
5.5.3 Gene annotation and conclusions ............................................................................110

5.6  Acknowledgements ......................................................................................................111

Chapter 6  Conclusions and synthesis ..............................................................................113
Chapter 7  Appendices .......................................................................................................119
List of References ...........................................................................................................161
Table of Tables

Table 2.1. Sampling information for *Ciona intestinalis* ................................................................. 26
Table 3.1. Sampling sites where *Pyura praeputialis* were collected .............................................. 46
Table 3.2. Descriptive population genomic statistics of all sampling sites ........................................ 52
Table 4.1. Results of generalised linear models with a binomial error distribution and a logit link function testing the effects of cross at 20°C................................................................. 79
Table 4.2. Results of generalised linear models with binomial error distributions and a logit link function testing the effects of temperature and cross ......................................................... 79
Table 5.1. Sampling information for *Ciona intestinalis* ................................................................. 88
Table 5.2. Number of loci of *Ciona intestinalis* identified using Bayenv2.0 at different Bayes Factors ........................................................................................................................................ 96
Table 5.3. Significantly enriched (P<0.05 after FDR correction) gene ontologies for *Ciona intestinalis* based on 231 genes ........................................................................................................ 104
Table 7.1. Pairwise population comparisons of F<sub>ST</sub> values and associated P values ............... 119
Table 7.2. Analysis of Molecular Variance under the assumption of four clusters as identified by ADMIXTURE ........................................................................................................................................ 120
Table 7.3. Analysis of Molecular Variance under the assumption of three clusters as identified by DAPC ........................................................................................................................................... 120
Table 7.4. Analysis of Molecular Variance under the assumption of two clusters (shallow artificial v shallow natural sites) ................................................................................................... 121
Table 7.5. Prior distribution parameters describing the set of scenarios investigated using DIYABC .............................................................................................................................................. 122
Table 7.6. Posterior probabilities and 95% confidence intervals of the sets of DIYABC scenarios 123
Table 7.7. Model checking in DIYABC ............................................................................................. 124
Table 7.8. F<sub>IS</sub> values (values in italics are statistically significant [P<0.05]) and population mean expected heterozygosity (HE) for all sites using the dataset with all putative F<sub>ST</sub> outlier loci excluded ............................................................................................................. 130
Table of Tables

Table 7.9. Number of loci recovered for the different datasets under different ipyrad clustering parameters. .................................................................144

Table 7.10. Locations where sea surface temperature (SST) data were collected along the coast of Chile. ..............................................................................145

Table 7.11. Analysis of Molecular Variance under the assumption of three geographic clusters. 146

Table 7.12. Posterior probabilities and 95% confidence intervals of the first set of DIYABC scenarios. ...........................................................................................................147

Table 7.13. Posterior probabilities and 95% confidence intervals of the second set of DIYABC scenarios. ...........................................................................................................147

Table 7.14. Number of replicates produced for each reciprocal cross ........................................149

Table 7.15. Matrix of Pearson’s R (below horizontal) and associated P values (above the horizontal) between environmental variables...............................................153
Table of Figures

Figure 1.1. Visual evidence of heterosis for yield in maize ........................................... 6

Figure 1.2. Hypothetical example of transgressive segregation ........................................ 7

Figure 1.3. Relative fitness of the hybrid sunflower Helianthus paradoxus, and parental species H. annuus and H. petiolaris, at different salt concentrations in soil ...................... 8

Figure 1.4. An aggregation of the ascidian Ciona intestinalis ........................................... 12

Figure 1.5. Generalised life cycle of a solitary ascidian ...................................................... 13

Figure 1.6. Aggregates of invasive ascidians ....................................................................... 15

Figure 1.7. Generalised schematic showing the different themes and questions associated with the four different data chapters of this thesis .................................................. 17

Figure 2.1. North Atlantic coastlines where the samples of Ciona intestinalis were collected .. 25

Figure 2.2. ADMIXTURE plots representing all sampled populations of Ciona intestinalis ...... 32

Figure 2.3. Discriminant Analysis of Principal Components using unlinked loci .................. 33

Figure 2.4. Matrix of F_{ST} values. .................................................................................... 34

Figure 2.5. The most likely evolutionary scenario involving the sample sites as calculated using DIYABC. ............................................................................................................... 35

Figure 3.1. Genomic differences between south-eastern Australian and Chilean sites inferred using putatively neutral loci ................................................................. 53

Figure 3.2. The most likely colonisation scenarios depicting the putative introduction route of P. praeputialis from Australia to Antofagasta Bay .................................................. 55

Figure 3.3. Average sea surface temperature (SST; back dots), maximum and minimum SST (blue crosses), and range between the maximum and minimum SST obtained from the satellite derived GHRSSST dataset .............................................................. 56

Figure 3.4. Discriminant Analysis of Principal Components using unlinked loci with population information provided as a prior and sites from Australia excluded ............... 58

Figure 4.1. The distribution of two Pyura herdmani lineages and Pyura stolonifera along the coasts of southern Africa .......................................................................................... 69
Table of Figures

Figure 4.2. Experimental design used to test the effects of temperature on inter/intraspecific crosses. .......................................................... 72

Figure 4.3. Violin plots depicting the development of control and hybrid crosses at 20°C. ....... 75

Figure 4.4. Stacked bar plots showing the percentage of post-metamorphs, pre-metamorphs, and other/dead stages of larvae at different times across temperature treatments. .......................................................... 76

Figure 4.5. Effects of temperature on the pre-metamorph development after 120 hours of control and hybrid crosses. .......................................................... 77

Figure 4.6. Effect of temperature on the development of different control and hybrid crosses on post-metamorphic development after 120 hours. .......................................................... 78

Figure 5.1. Eastern North Sea coastline where samples of Ciona intestinalis was collected ...... 89

Figure 5.2. Experimental set up for laboratory crosses .......................................................... 94

Figure 5.3. Discriminant analysis of principal components showing clear genetic differentiation by depth based on loci under selection .......................................................... 98

Figure 5.4. ADMIXTURE plots of multiple Ks .......................................................... 100

Figure 5.5. FST heatmaps for outlier loci .......................................................... 101

Figure 5.6. Mean posterior probability of assignment to hybrid classes for each empirical and simulated individual in parallelnewhybrid .......................................................... 101

Figure 5.7. Venn diagram representing the number of outlier loci exhibiting significant correlation to one or more environmental variable .......................................................... 102

Figure 5.8. Location along each of the 14 Ciona robusta chromosomes of the 106 outlier loci that could be mapped to a genetic location .......................................................... 103

Figure 5.9. The percentage of larvae that have undergone development to (A) settler stage, (B) pre-metamorph stage, and (C) post-metamorph stage over 24, 72, and 120 hours .......................................................... 106

Figure 7.1. Cross validation error estimate from the ADMIXTURE analysis showing K = 4 as an appropriate modelling choice .......................................................... 125

Figure 7.2. Prior checking for each DIYABC run .......................................................... 126
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3</td>
<td>DIYABC scenarios used to assess colonisation history within the northeast Atlantic (A and B) and across the north Atlantic (C)</td>
</tr>
<tr>
<td>7.4</td>
<td>The first two axes of the PCA as part of the model checking function of DIYABC</td>
</tr>
<tr>
<td>7.5</td>
<td>ADMIXTURE plots representing all sampled populations of <em>Ciona intestinalis</em>, using the dataset with all F$_{ST}$ outlier recovered from either BayeScan and pcadapt removed</td>
</tr>
<tr>
<td>7.6</td>
<td>Discriminant Analysis of Principal Components with all putative F$_{ST}$ outlier loci (77) excluded</td>
</tr>
<tr>
<td>7.7</td>
<td>DIYABC scenarios used to assess the introduction of <em>Pyura praeputialis</em> from Australia to Chile</td>
</tr>
<tr>
<td>7.8</td>
<td>DIYABC scenarios used to assess the introduction of <em>Pyura praeputialis</em> from Eastern Australia to Chile</td>
</tr>
<tr>
<td>7.9</td>
<td>Heatmap of F$_{ST}$ values</td>
</tr>
<tr>
<td>7.10</td>
<td>Discriminant Analysis of Principal Components with population information included as priors</td>
</tr>
<tr>
<td>7.11</td>
<td>Cross-validation plots produced from ADMIXTURE for the (A) neutral, (B) bayenv, and (C) pcadapt dataset</td>
</tr>
<tr>
<td>7.12</td>
<td>Admixture plots depicting the inferred ancestry of each <em>Pyura praeputialis</em> individual to pre-defined genetic clusters</td>
</tr>
<tr>
<td>7.13</td>
<td>Discriminant Analysis of Principal Components with population information included as priors</td>
</tr>
<tr>
<td>7.14</td>
<td>Prior checking for each DIYABC scenario</td>
</tr>
<tr>
<td>7.15</td>
<td>Prior checking for each DIYABC scenario</td>
</tr>
<tr>
<td>7.16</td>
<td>Comparison of the posterior probabilities of five scenarios simulated in DIYABC pertaining to the number of individuals originating from eastern Australia that were introduced to Antofagasta, Chile</td>
</tr>
<tr>
<td>7.17</td>
<td>Average yearly sea surface temperature (SST) over a fifteen-year period (2003-2017) of different sites within Australia and Chile</td>
</tr>
</tbody>
</table>
Table of Figures

Figure 7.18. Daily differences in sea surface temperature (SST) obtained from the satellite derived GHR SST dataset (A Group for High Resolution Sea Surface Temperature; JPL MUR MEaSUREs Project, 2015) between the most northerly (C1) and southerly (C6) sites within Antofagasta Bay ......................................................... 144

Figure 7.19. Microscope images of early life history stages that were counted .................. 150

Figure 7.20. Sea surface temperature at Port Elizabeth between 2003 – 2017 ............... 151

Figure 7.21. Cross-Validation values as inferred by ADMIXTURE at different \( K \)s using the pcadapt dataset................................................................. 154

Figure 7.22. Cross-Validation values as inferred by ADMIXTURE at different \( K \)s using the bayenv dataset........................................................................ 155

Figure 7.23. Fertilisation curve for different sperm concentrations at shallow water salinity and 16°C ..................................................................................................... 156

Figure 7.24. Discriminant analysis of principal components on the neutral dataset .......... 157

Figure 7.25. Cross-Validation values as inferred by ADMIXTURE at different \( K \)s using the neutral dataset......................................................................................... 158

Figure 7.26. ADMIXTURE plots of \( K = 2-4 \) using the neutral dataset showing depth-defined divergence........................................................................................................... 158

Figure 7.27. \( F_{ST} \) heatmap for the neutral dataset..................................................... 159

Figure 7.28. Replicate of Figure 5.9 with raw data points rather than bar plots to demonstrate the unbalanced nature of our experimental crosses.................................................. 160
List of Accompanying Materials

All data supporting this study are openly available from the University of Southampton repository at https://doi.org/10.5258/SOTON/D1372 and from the Dryad Digital Repository at https://doi:10.5061/dryad.wh70rxwhw
Research Thesis: Declaration of Authorship

Print name: JAMIE HUDSON

Title of thesis: The Roles of Hybridisation, Contemporary Climate Change, and Recent Range Shifts on the Redistribution of Marine Biodiversity

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:-

Signature: ............................................................................................................. Date: ..............................
Acknowledgements

First and foremost I would like to thank my supervisor and mentor, Dr Marc Rius, for his guidance, support, and importantly patience throughout the past six years. I have been very lucky to have a supervisor whose door was always open for discussion and advice. I also need to extend gratitude to Prof Christopher McQuaid for being an extremely knowledgeable, inspiring, and calming influence during my research experience (and continue to be to this day). Despite the countless snippets of scientific gold-dust you have sprinkled my way, for some unknown reason, teaching me that “I am well” rather than “I am good” is something that will always stay with me. I would also like to thank Dr Mark Chapman, for always being available to assist with specific needs and issues- even if seemingly out of the blue! This thesis was very international in nature, and whilst I have tried to include specifics in each chapter, I would like to extend a broad thanks to everyone whose path I have crossed during this PhD, you have all contributed to this.

For all of the lab members of the Ecology and Evolution lab group, thank you for your advice and contributions to intellectual discussion during our lab meetings. I would especially like to thank Dr Steve Bourne and Luke Holman. Steve, you were always a brilliant guide for me from when I was a wee sprog. Whenever I had an issue, I could rely on you having already experienced it and offering more time and patience to help than I could ever have asked for. Luke, we have been on this PhD journey together, and I consider myself very lucky to have been able to work closely with you (although not enough for my liking). Your constant enthusiasm was infectious, and your drive for better research and continuous conveyor-belt of ideas was inspiring. I hold our discussions (both trivial and scientific) and friendship very close to my heart.

There are too many of the postgrad community at NOCS that deserve individual mention for what they have provided during the past three and a half years, however I cannot do it justice in few words, so just know that I am thankful for everyone. To all in office 344/35 (and honorary member David Riley), I thank you all for providing many laughs and smiles throughout the PhD. Millie and Sarah, I hope you both are able to step back and look at how much you have grown as people and researchers from the shy first years entering our office in 2016.

I would also like to extend gratitude to the Southampton Marine and Maritime Institute for providing the funding to support me during this research. Furthermore, I acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.
Acknowledgements

To my family, words cannot describe how grateful I am for everything you have sacrificed to enable me to be where I am right now. Dad, without you I would not have the keen interest in the natural world that I do. In hindsight I am very thankful for your attempts to get me interested in rock pools during my youth- even if I was more interested in being David Beckham than David Attenborough then. Mum, I can’t begin to describe everything you have done for me, and how much I owe you. You have always been my biggest supporter, and I pray one day I can repay your altruism. I hope you are proud. Nix, thank you for all the brotherly support over the years. Even though I have the high ground now, remember you were the chosen one, clever girl.

Finally, Bryony, you have been my rock throughout this whole journey. Whenever times got tough, you were always there for me- even if I was the other side of the world. Sometimes I feel you have lived this PhD as much as I have, especially for the past year as I have been constantly glued to my computer. I will forever be appreciative of your love, kindness, and patience as I explain for the thousandth time to a friend that I study sea squirts and not sea cucumbers.
Definitions and Abbreviations

ABC: Approximate Bayesian computation
AMOVA: Analysis of molecular variation
CCC: Contemporary climate change
DA: Discriminant analysis
DAPC: Discriminant analysis of principal components
ELHS: Early life-history stages
FSW: Filtered sea water
GBS: Genotyping-by-sequencing
GHRSSST: A Group for High Resolution Sea Surface Temperature
GLM: Generalised linear model
GO: Gene ontology
HTS: High-throughput sequencing
MCMC: Markov chain Monte Carlo
NIS: Non-indigenous species
PCA: Principal component analysis
PSU: Practical salinity units
RAD-seq: Restriction site-associated DNA sequencing
SNP: Single nucleotide polymorphism
SST: Sea surface temperature
Chapter 1  Introduction

A central theme in ecology and evolutionary biology is understanding the factors that determine and maintain where species live. For some species, there are obvious discontinuities in habitat that preclude survival or migration, such as an air-water interface or geographic barriers such as mountains (Kirkpatrick & Barton, 1997). For others however, limits in their range are the result of nuanced physiological or biotic factors such as temperature tolerance (Payne et al., 2016), drought resilience (Esquivel-Muelbert et al., 2017), or availability of appropriate food sources or competition (deRivera et al., 2005). These factors are not mutually exclusive, for example the upper limit of the intertidal red turf algae, *Mazzaella parksii*, is set by physiological stress, whereas the lower limit is constrained by herbivory (Harley, 2003). When the conditions influencing these factors change (in magnitude or intensity), the distribution of species may shift (Chen et al., 2011). This introduction to the thesis briefly outlines different factors that can alter species ranges, including *range expansions* (section 1.1.1) and *biological invasions* (section 1.1.2). These changes in species distributions often differ in speed and drivers behind evolutionary processes promoting range shifts. In addition, there is an appreciation that changes in species distribution are also being affected by contemporary climate change (CCC; section 1.2). Such a redistribution of species’ ranges in many cases promote hybridisation between divergence lineages (section 1.3), which can have consequences for individual and population fitness, and have knock-on effects for future movement of species.

1.1  Species redistribution

Human activities are responsible for major shifts in species distributions around the world, often called the redistribution of biodiversity. There are several ways that humans alter the distribution of species. One is the human-mediated alteration of the climate (*contemporary climate change* [CCC]; section 1.2), which can reshape species ranges by influencing habitat-suitability (Sultana et al., 2017), modifying dispersal mechanisms (Cetina-Heredia et al., 2015), affecting species interactions (Hughes, 2012), and directly impacting on species physiology (Binet & Doyle, 2013). From an evolutionary perspective, the movement of individuals through space can be conceptualised as a reshuffling of genotypes. This reshuffling can bring divergent populations into contact, and lead to an exchange of genes forming hybrid genotypes (*hybridisation*; section 1.3). Evolutionary processes associated with hybridisation have the potential to increase the likelihood of further species redistribution and alter the ecological and evolutionary consequences of such redistribution. Furthermore, evidence shows that CCC can both increase the chance, and effect of,
hybridisation, and the two factors can act concurrently and exacerbate the consequences of species redistribution (section 1.4). To effectively study these different topics covering species distributions, one needs to choose appropriate study systems that enable thorough examination of the subject matter. Coastal marine ecosystems provide suitable systems to further our understanding of species redistribution. Shipping is a major vector of species redistribution, with ports, marinas, and harbours representing a central hub for biological invasions (Zabin et al., 2014). Coastal ecosystems are some of the most susceptible to CCC, and future predictions suggest nuances in the intensity of CCC to such regions. Organisms with well-known dispersal methods, high affinities to artificial structures (i.e. marinas), complex intraspecific phylogenies, and that are easily sampled such as ascidians (section 1.5) provide unique windows into researching the roles of hybridisation, local adaptation, and environmental change on the redistribution of biodiversity.

### 1.1.1 Natural range expansions

The areas that species inhabit (species ranges) are not stationary entities but rather are in flux through time (Sexton et al., 2009). Natural changes in these ranges occur across vastly different time scales, from short-term rapid changes to long-term gradual adjustments. Ranges can undergo transient shifts throughout the year, as is the case in migratory birds which inhabit vastly different geographical locations and habitats during different seasons of the year (Dingle, 2008), or they can be longer lasting. Range shifts materialise when organisms alter their geographical spread and colonise a new area (Vermeij, 2005), and may occur naturally when two previously isolated habitats come into contact. A prominent example of this is the large redistribution of biodiversity that occurred after the last glacial maximum (LGM) ca. 16,000 YA. Numerous species including invertebrates (Cooper et al., 1995), mammals (Seddon et al., 2001), reptiles (Ursenbacher et al., 2006), amphibians (Vences et al., 2013), and plants (Magri, 2008) expanded their distribution from southern European refugia to newly available habitat throughout the rest of Europe (reviewed in Hewitt, 1999). This post-glacial colonisation also occurred in the marine environment, and contemporary patterns of genetic variation have enabled us to identify marine glacial refugia and reconstruct recolonization pathways (García-Marín et al., 1999; Hoarau et al., 2007; Campo et al., 2010; Silva et al., 2014). These natural range expansions occur gradually, often over geological time scales, and thus evolutionary processes acting on these species operate over a large period of time. Critically, range expansion due to natural mechanisms depends on the adaptive potential of fringe populations at a range edge (Kirkpatrick & Barton, 1997; Sexton et al., 2009). When species' ranges change, the species are often faced with novel evolutionary pressures at their range edges which they must adapt to, or else face the possibility of extirpation.
from this new habitat (Bridle & Vines, 2007). This is in contrast to other methods of species redistribution (such as biological invasions [section 1.2]), which are characterised by different dispersal dynamics and often occur over much shorter time-scales.

1.1.2 Biological invasions

The examples of range expansions above have an abiotic driver (e.g. formation of new land masses, retreat of glacial regions), although drivers of species redistribution may also be artificial in essence (biological invasions or artificial range expansions). Anthropogenic activity can directly alter the distribution of organisms through intentional [e.g. introduction of cane toads as biocontrol in Australia (Easteal, 1981)] or unintentional [e.g. hull fouling by marine organisms (Clarke Murray et al., 2011)] release. Humans can also alter the distribution of species through additional factors including habitat disturbance (Hodgson et al., 2012) and in extreme cases extinction (Pimm et al., 2014), however these are beyond the scope of this thesis. Whilst outlined in detail in Chapter 3, biological introductions are the movement of species via anthropogenic transport to a novel location. This process can be considered as a series of stages, each containing barriers which species must overcome to progress to the following stage (Richardson et al., 2000; Blackburn et al., 2011). Generally, the invasion process can be conceptualised in four main stages: i) Transportation of individuals or propagules from the native range via a vector; ii) Introduction of the species to the novel environment; iii) Establishment of the species to the novel environment where it can survive and form a self-sustaining population through reproduction; iv) Further significant spread from the initial point of introduction. Successful progression to this final stage generally deems a species to be an “invasive species”. Invasive species are often high priority in research and management, due to economic and ecological impacts associated with them. These impacts are often exacerbated, as NIS are transported via anthropogenic means they often have a high affinity to man-made structures. For example, in the marine environment non-indigenous tunicates outcompete and overgrow commercial shellfish, foul aquaculture gear, restrict the cycling of water and nutrients, and require control mitigation costs (Carman et al., 2010). When quantitative estimates of the impact of non-indigenous species have been made, the values are sizeable – 88 species of invasive molluscs, for example, have an estimated annual cost of > $2.2 billion in the United States (Pimentel et al., 2005). Invasive species have been reported as the second most common reason behind species extinctions (Bellard et al., 2016), and cause an immense threat to ecosystem services (Walsh et al., 2016). In a pan-European review of the impacts of marine invasive species, Katsanevakis et al., (2014) found that 87 species have been documented to have a high impact on ecosystem services or biodiversity. Forty-nine of these species were ecosystem engineers modifying existing habitats or creating novel habitats by
changing their abiotic properties. Understanding the mechanisms that promote successful invasive species, and the consequences of such a shuffling of biodiversity is therefore paramount.

There is current debate regarding the scientific value of studying range expansions and biological invasions as different entities, or whether such separation is gratuitous (Hoffmann & Courchamp, 2016; Wilson et al., 2016a). Whilst there is merit to both sides of the argument, Wilson et al., (2016a) succinctly summarise the applicability for separation: both Stonehenge and a rocky shore were both formed by rolling stones, however the processes shaping these were quite different and would not be expected to be studied in the same way. As mentioned previously, whereas natural range expansions rely on the adaptive potential of fringe populations, the source of an introduction can be anywhere within the native range. This can be significant as species can have differing levels of genetic diversity throughout their native range, from edge populations to those occupying the centre of a species’ distribution (Vucetich & Waite, 2003; Diekmann & Serrão, 2012), which can affect invasion success. Despite the different drivers promoting, and evolutionary factors facilitating, range expansions and biological invasions, studying their role in the redistribution of marine biodiversity remains fundamental.

1.2 Contemporary climate change

Perhaps the most familiar aspect of current global change is CCC. This has led to a multitude of environmental effects such as altered precipitation regimes, acidification of the world’s oceans, and an increase in mean global temperature (IPCC, 2013). CCC is causing unprecedented consequences throughout all major ecosystems on Earth. For example, some tropical rainforest regions are expected to be replaced by savannah and grassland, with a net reduction in primary productivity associated with increased temperatures and reduced precipitation (Lyra et al., 2017); Arctic ecosystems are facing reductions in reproductive success of Arctic species and changes in community structure (Wassmann et al., 2011); and even deep sea ecosystems, despite absorbing large quantities of heat and carbon dioxide (Levin & Bris, 2015), have been shown to be extremely vulnerable to small temperature changes with limited recovery potential of biodiversity (Danovaro et al., 2004; Yasuhara et al., 2008). The ocean has absorbed 93% of the extra energy arising from greenhouse emissions and led to an increase in average global sea surface temperature (SST) of almost 1°C (IPCC, 2013), and has taken up ca. 30% of the anthropogenic carbon dioxide released into the atmosphere. In turn, research on CCC effects in the ocean are mainly associated with rising temperature (Bindoff et al., 2007) and the reduction in ocean pH leading to ocean acidification (Doney et al., 2012). These two consequences are linked with alterations in circulation, increased ocean stratification, changes in ocean chemistry, rising sea level, and reduced oxygen content (Doney et al., 2012). Changing SST is, however, perhaps the
effect of CCC that is most ubiquitous and pertinent in coastal marine environments. SST fluctuations have numerous effects on biological processes ranging from the metabolic level (e.g. enzyme reactions and diffusion) to key processes such as primary production and lead to alterations in food web dynamics (reviewed in Hoegh-Guldberg & Bruno, 2010).

1.3 Hybridisation

“We used to make fun of Edgar Anderson by saying that he was finding hybrids under every bush. Then we realized that even the bushes were hybrids” – Late American botanist Warren H. Wagner (Abbott et al., 2013).

Hybridisation occurs when genetically distinct populations (either within or between species) interbreed, producing offspring with genotypes constituting a combination of both parental gene pools (Roman & Darling, 2007). Hybridisation has excited, intrigued, and perplexed biologists for many years. Charles Darwin even dedicated a chapter of “On the Origin of Species” to hybridisation and noted that interspecific mating can be difficult (Darwin, 1859). In the 1900s the quest to understand hybridisation gained momentum, with research direction split between botanists and zoologists. Botanists focused on understanding how hybridisation produces genetic diversity (Stebbins, 1950), whilst zoologists concentrated on resolving the role of hybridisation in speciation, with the opinion that hybrids were rare entities (Dobzhansky, 1937; Mayr, 1942).

Currently, there is a recognition that hybridisation is more common than initially thought, with estimates of 25% of plant species and 10% of animal species involved in hybridisation (Mallet, 2005).

1.3.1 The costs associated with hybridisation

Despite increasing appreciation of the prevalence of hybridisation within a variety of flora and fauna, progeny from hybridisation of closely related species are still often infertile, sterile, or inviable. Hybrid incompatibility is commonly attributed to deleterious epistatic interactions between alleles at different loci of parental genomes (Coyne & Orr, 2004). These interactions, known as “Dobzhansky-Muller incompatibilities” (Dobzhansky, 1937; Muller, 1942) arise due to genetic incompatibilities between the divergent genomes of two populations which once lived in sympatry. Hybridisation can also lead to outbreeding depression, where hybrid progeny are less fit than parental crosses. For example, hybrids of Chilean neotropical species of Drosophila pavani and D. gaucha exhibit reduced fitness related to larval foraging behaviours, notably feeding behaviour and locomotion. This is suggested to be due to the breakdown of co-adapted gene
complexes during recombination (Godoy-Herrera et al., 2005). Reduced hybrid fitness is commonly observed. In studies where hybrid fitness has been directly measured, reduced fitness was reported in 56% of F1 hybrids compared to either parent (Arnold & Hodges, 1995). Furthermore, even if hybrids are genetically fertile, reproduction may not occur, or be reduced, due to prezygotic reproductive isolation. An example is the case of hybrid male fur seals, which have a lower reproductive success than pure-species males, possibly due to phenotypic traits that affect mate choice (Lancaster et al., 2007).

1.3.2 The benefits associated with hybridisation

Hybridisation may also provide many evolutionary “benefits” for species. Inter- or intraspecific hybridisation can promote adaptive variation (Rius & Darling, 2014; Stelkens et al., 2014). Hybridisation brings together divergent genotypes, potentially increasing genetic variation and creating a larger pool of genotypes on which selection can act (Hegarty, 2012).

Figure 1.1. Visual evidence of heterosis for yield in maize. Reciprocal first-generation hybrids (F1) can be seen to possess a larger yield (and hence seeds and reproductive fitness) than either parents (P1 and P2). Image from Xing et al., (2016).

Heterosis, also known as hybrid vigour, is the phenotypic superiority of hybrid offspring in comparison to its parents. The genetic basis of heterosis is hotly debated (Kaeppler, 2012), with dominance (the masking of deleterious recessive alleles from one parent by dominant alleles from the other parent), overdominance (the beneficial interactions of alleles from different lineages), or epistasis (the complementary interaction between multiple loci from different lineages) being
proposed as mechanisms behind observed hybrid vigour (Bock et al., 2015). Notwithstanding these mechanisms, heterosis can remove inbreeding depression (Keller & Waller, 2002), and provide first-generation hybrids (F1) with phenotypic superiority over their parents (Lippman & Zamir, 2007). Indeed, F1 progeny of ring-necked pheasant (*Phasianus colchicus*) subspecies exhibit raised mean individual fitness relative to non-hybrid subspecies (Drake, 2006). There has been debate regarding the efficacy of heterosis, as despite being important in asexually reproducing plants, it is seen as transient in sexually reproducing species due to the effects of genetic segregation (Lee, 2002). Nonetheless, even a short-term increase in fitness can allow a population to increase in size rapidly [due to a so-called ‘catapult effect’ (Drake, 2006)] and overcome the disadvantages associated with small population sizes (Drake, 2006). Hybridisation can reduce the detrimental effects of genetic bottlenecks by masking or purging deleterious mutations (Ellstrand & Schierenbeck, 2000; Rius & Darling, 2014), or lead to the formation of novel phenotypes, which could provide a selective advantage due to transgressive segregation (Figure 1.2).

Transgressive segregation has been identified in a range of organisms (Stelkens et al., 2009; Pritchard et al., 2013) and may be an expected consequence of hybridisation (Rieseberg et al.,...
1999). A review of 171 studies that report phenotypic variation in hybrids found that 91% of studies report transgressive traits (Rieseberg et al., 1999). One widely cited example of transgressive segregation is evident in the hybrid sunflower species *Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus*. These three species are stabilised hybrid derivatives of the same parental sunflower species *H. annuus* and *H. petiolaris* (Rieseberg, 1991). *Helianthus annuus* is found in clay-based soils that exhibit seasonal fluctuations in rainfall, whilst *H. petiolaris* inhabits dryer, sandier soils. Due to transgressive segregation, the hybrid species are found in habitats too extreme for the parental species: *H. anomalus* is found on sand dunes, *H. deserticola* in dry, sandy desert soil, and *H. paradoxus* inhabits brackish salt marshes (Rieseberg et al., 2003). As the parents are unable to tolerate these extreme habitats, gene flow between hybrids and parents halted, allopatric speciation occurred, and the hybrids became genetically distinct (Figure 3). It is now widely acknowledged that hybrid species exhibit ecological differentiation from their parental species (Gross & Rieseberg, 2005).

![Relative fitness of the hybrid sunflower Helianthus paradoxus, and parental species H. annuus and H. petiolaris, at different salt concentrations in soil. Whilst all species are able to survive low salt concentration, only H. paradoxus is able to tolerate high salt concentrations. Adapted from Welch & Rieseberg (2002).](image)
1.4 Hybridisation and contemporary climate change leading to species redistributions

In order to understand recent changes in species distribution, one needs to understand the evolutionary history of the species. If populations or species have diverged whilst experiencing similar environmental pressures, they can be expected to build up genomic differentiation owing to the stochastic effects of drift. If, however, populations or species diverge and experience differing environmental conditions, selective pressures of these environments may promote local adaptation, where each population is fitter than the other in their own habitat (Kawecki & Ebert, 2004). Whilst studies of local adaptation have invariably assessed environmental gradients associated with latitude and longitude, there is a dearth of studies studying the effects of the ‘vertical plane’ (altitude or depth) on local adaptation. Understanding the mechanisms driving the evolutionary history of species is important, as they can have knock-on effects in relation to species distribution. For example, knowledge of local adaptation to environments can identify species that are potentially “pre-adapted” to recently introduced areas (Guo et al., 2014). In areas where diverging populations come into contact and reproduce, local adaptation associated with divergent selection due to steep selective pressures can temper the scale of hybridisation (Johannesson, 2003).

Both natural and human-driven species redistributions increase the likelihood of secondary contact between divergent populations or species, which in turn promotes the opportunity for hybridisation. Cyclically, hybridisation has also been shown to facilitate range expansions and invasions (Ellstrand & Schierenbeck, 2000), both in the form of intraspecific admixture between different source populations within a native range (Rius & Darling, 2014; Wagner et al., 2017), or through interspecific hybridisation (c.f. Helianthus spp. example from section 1.3; Rieseberg et al., 2003). Hybridisation is not limited to pre-introduction, it can also occur as a consequence post-introduction or expansion (Egger et al., 2012). Hybridisation offers a mechanism for increased genetic diversity, and enable the overcoming of Allee effects (density dependent fitness; Mesgaran et al., 2016), both of which can counteract demographic stochasticity associated with population bottlenecks often accompanying shifting distributions. Hybrids can be detrimental to recipient ecosystems. Post-introduction hybridisation can lead to the displacement of native species (Huxel, 1999), either through introgression (Rhymer & Simberloff, 1996) or competition (Parepa et al., 2014). A fitting example demonstrating the relationship between hybridisation and biological invasions is observed cordgrass. In salt marshes across eastern North America, Spartina alterniflora form dense aggregates that dominate the lower intertidal zone. Accidental release of this species to Southampton, UK, via seeds transported in ballast water introduced the species to
Chapter 1

a novel environment. Here, *S. alterniflora* came into contact with the native congener, *Spartina maritima*, and hybridised forming the F1 hybrid *S. x townsendii*, which itself underwent chromosome doubling to form the new species *S. anglica* (Thompson, 1991). The hybrid *S. anglica* has since colonised large portions of the British Isles through a combination of natural dispersal and anthropogenic transport. This example draws attention to the situation when the introduction of species (or range expansion) is a driver of hybridisation, but also when hybridisation is then a driver of further range expansion. Climate change similarly can promote range shifts and alter species distributions. Shifts in abiotic factors are associated with the range expansion of the butterfly *Atalopedes campestris* across western North America (Crozier, 2004), and by altering the biotic conditions of ecosystems (van der Knaap *et al.*, 2005; Huang *et al.*, 2012) can promote range shifts (Svenning *et al.*, 2014). In the marine environment, extreme heat waves associated with CCC have led to the range contraction of kelp forests, altering community-wide species composition (Wernberg *et al.*, 2013, 2016). Additionally, CCC is predicted to alter dispersal pathways and population connectivity through the alteration of oceanic currents (Wilson *et al.*, 2016b). The relationship between hybridisation, CCC and the movement of species is becoming increasingly apparent (Scriber, 2013; Chunco, 2014; Chown *et al.*, 2015; Vallejo-Marín & Hiscock, 2016; Canestrelli *et al.*, 2017), with the consequences of hybridisation and range expansions often exacerbated by CCC. For example, CCC has been shown to alter pre- and post-mating reproductive barriers (Chunco, 2014) and change life-history stages of hybrids (Canestrelli *et al.*, 2017), altering spatio-temporal patterns of hybridisation (Muhlfeld *et al.*, 2014). The barred tiger salamander, *Ambystoma tigrinum*, provides an appropriate system encompassing this. *Ambystoma tigrinum* was intentionally released within western regions of North America (Riley *et al.*, 2003). It has since become invasive, and its range overlaps with that of the native *Ambystoma californiense*. Hybridisation between the two species has occurred, with increased temperature due to CCC being shown to positively influence juvenile dispersal, and hence the further rapid range expansion of a hybrid swarm (Johnson *et al.*, 2010).

1.5 *Ascidiants*: a unique group to study changes in marine ecosystems

There are few study systems that are suitable for the study of the above topics. For example, experiments involving plants enable multi-generational assessments of hybrid fitness through common-garden or field manipulation (Mitchell *et al.*, 2019). Similarly, range expansions have been assessed using terrestrial invertebrates in laboratory settings due to adult life spans in the scale of days (Wagner *et al.*, 2017). The marine environment provides its own set of challenges. In addition to the obvious difficulty sampling many marine ecosystems, often marine organisms produce extremely motile larvae (Cowen & Sponaugle, 2009), rendering controlled multi-
generational experiments nearly impossible. Finally, the use of genomic tools available for marine biologists has lagged behind our terrestrial counterparts. For example, the number of marine species with sequenced genomes only reached double figures in 2015, almost ten years after terrestrial species reached the same value (Kelley et al., 2016). Ascidians (Chordata, Ascidiacea; Figure 1.4) are sessile marine invertebrates that inhabit every ocean in the world, living at depths from the intertidal zone to > 8,000 m (Sanamyan & Sanamyan, 2006). The complex taxonomic history of ascidians has been summarised nicely by Shenkar & Swalla (2011), whereby they have previously been grouped with bryozoans (Milne-Edwards, 1843), identified as a mollusc (Hancock, 1850), before now being accepted as part of the sub-phylum Tunicata. Phylogenomic studies have placed the Tunicata as a sister group to the vertebrates (Bourlat et al., 2006; Delsuc et al., 2006). This position as basal chordates and phylogenetic relatedness to vertebrates has promoted the use of ascidians to study evolutionary developmental biology (evo-devo) as well as studies of immunology where research has discovered natural products from ascidians that help develop anti-cancer and malarial drugs (Mendiola et al., 2006; Rajesh & Annapan, 2015; Watters, 2018). Indeed, the applicability of evolutionary developmental studies of ascidians to humans is seen by the vase tunicate, *Ciona robusta*, being only the seventh animal, first marine animal, and first non-laboratory invertebrate to have its genome sequenced (one week after the mouse genome was published; Dehal et al., 2002).
Figure 1.4. An aggregation of the solitary ascidian *Ciona intestinalis* from Fiskebäckskil, Sweden. Photo taken by Jamie Hudson.

Despite the pertinence of ascidians to the fields of “evo-devo” and medicinal progression, it is their ecology and appropriateness to population genomics study that makes them suitable study species for this thesis.

### 1.5.1 Ecology of ascidians

Ascidians readily settle on hard substrata including rocks, boulders, and anthropogenic surfaces; however some species can attach to softer sediment (Monniot *et al.*, 1991). From the post-metamorph stage (i.e. once their two siphons have developed), ascidians use siphonal cilia and tentacles to bring water in through an oral siphon before filtering food through a mesh-like branchial sac and pumping filtered water out via their atrial siphon. It is the structure of this branchial sac that defines the three orders within the class Ascidiacea. The three orders, Aplousobranchia, Phlebobranchia, and Stolidonbranchia have a simple, vascular, and folded branchial sac respectively (Shenkar & Swalla, 2011), with these groupings being supported by phylogenetic analysis based on 18S rDNA (Zeng & Swalla, 2005). Ascidians consist of two main morphological types; colonial and solitary (although social ascidians have been identified as an intermediate between the two types). Colonial ascidians consists individual zooids which all contain their own oral siphon, but share a communal atrial siphon, and reproduce both by sexual and asexual reproduction. Solitary ascidians are independent organisms that only reproduce
sexually through external fertilisation. Both colonial and solitary ascidians are simultaneous hermaphrodites. Whilst different species exhibit slightly different life-history strategies (such as unique gamete retention and egg brooding), the general reproduction strategy of solitary ascidians is as follows (Figure 1.5): Eggs and sperm (rarely both by the same individual to reduce self-fertilisation) are broadcast released into the water column where they remain viable for only a few hours. Fertilisation occurs in the water column and the newly formed embryo develops into a tadpole-shaped lecithotrophic larvae which hatches within a couple of hours. The tadpole larvae remains pelagic for ca. 12 hours before using sensory cues to settle on a substratum head first, where they rapidly undergo metamorphosis by absorbing their tail, digesting their notochord, and developing the internal and external structures found in the adult forms. This whole process occurs extremely quickly (between a few hours and a few days), supposedly as an evolutionary response to get to the safety of the adult stage as quickly as possible (Pineda et al., 2012).

Figure 1.5. Generalised life cycle of a solitary ascidian. 1) Reproductively mature adults spawn, releasing gametes into the water column, where external fertilisation occurs, and eggs hatch into larvae after ca. 12 hours. 2) The free swimming lecithotrophic larvae remain pelagic for ca. 12 hours before attaching to the substrate and becoming a settler. 3) This settler begins tail reabsorption, utilising the last bits of the yolk. 4) Metamorphosis occurs and the newly developed siphons enable the post-
metamorph to start filter feeding. 5) The final stages of development occur as the post-metamorph develops into first a juvenile, and then an adult that will eventually be reproductively active. Figure adapted from Rius et al. (2017).

Temperature and salinity are the main abiotic drivers of ascidian distribution (Auker & Oviatt, 2008; Epelbaum et al., 2009; Nagar & Shenkar, 2016). As expected by the circumglobal distribution of ascidians, different species experience markedly different conditions. For example, despite a prolonged larval and development stage, the Antarctic ascidian *Cnemidocarpa verrucosa* can develop at temperatures as low as 0°C (Strathmann et al., 2006), and species have been recorded in the Arabian Gulf where temperatures exceed 35°C (Monniot & Monniot, 1997). All species are marine, with the general lower limit of salinity tolerance being approximately 20 PSU (Dybern, 1967). Some generalists are able to survive a broad range of conditions however, such as *Ciona intestinalis*, which can develop under salinities ranging from 11 – 40 PSU (Dybern, 1967). As with most marine invertebrates, different life history stages exhibit differing tolerances to environmental pressures with adults being more susceptible to a range of conditions than larvae and early life history stages (Pineda et al., 2012). Thus, despite an individual being capable of surviving a range of environmental conditions, it is the ability for the full life cycle (i.e. reproduction, fertilisation, development etc) to complete for a population to be sustained under certain environmental conditions.

### 1.5.2 Ascidians as invaders

That ascidians offer excellent study systems for further understanding biological invasions has been extensively reviewed elsewhere (Zhan et al., 2015). For this reason, I will only briefly overview the characteristics that support their use in this thesis. Firstly, as mentioned previously, ascidian larvae have an extremely shortened pelagic duration. It is therefore impossible for them to naturally disperse over great distances [although ascidians can raft great distances on natural (Thiel & Gutow, 2005) and artificial debris (Carlton et al., 2017; Simkanin et al., 2019)]. As such, disjunct distributions, or distributions across large oceanic basins can generally be considered to be human-mediated in origin. Secondly, ascidians have a high propensity to settle and establish on artificial structures and are readily found in marinas and harbours. In other words, they are generally located within hubs of transoceanic transportation. Additionally, numerous human-mediated vectors are suitable to spread ascidians, including: ballast tanks (Briski et al., 2011), hull fouling (Aldred & Clare, 2014), and recreational boating (Clarke Murray et al., 2011). Ascidians therefore face many transport opportunities, but also, if associated with aquaculture infrastructure, often cause economic and ecological impacts (Figure 1.6). Thirdly, ascidians generally have a broad tolerance of environmental conditions (see section 1.5.1). Not only does
Chapter 1

This showcase their applicability to study the invasion process (Zhan et al., 2015), as they can be expected to experience a wide range of environments during the invasion process, but it also renders them suitable to study the effects of CCC on their biology. Finally, the sessile nature of ascidians means that once identified, they are generally easy to sample, especially from artificial structures. Furthermore, they can be maintained in aquaria for further experimentation (Joly et al., 2007), with well-developed protocols for gamete extraction (Marshall et al., 2000; Christiaen et al., 2009) enabling study on early life history stages (Rius et al., 2010a; b; Pineda et al., 2012).


1.6 Thesis structure and author contributions

This thesis uses three study systems to advance knowledge of the evolutionary and ecological mechanisms that both shape current distributions of species, but also predict the potential of future range shifts. I use a multidisciplinary approach that includes genomic techniques, experimental crosses, and environmental datasets to examine the role of local adaptation across extremely steep gradients (metres), to illuminate factors facilitating transoceanic movement of species (>10,000 km) and determine the potential of hybridisation and CCC to promote range shifts. Each chapter offers a unique insight to mechanisms acting at different spatial and temporal
Chapter 1

scales that determine the distribution of marine biodiversity (Figure 1.7). To investigate the above, this thesis has five objectives:

1. To elucidate the genomic patterns underlying the distribution of widespread ascidian species throughout their native and introduced ranges.

2. To identify evolutionary mechanisms that may have led to the genomic patterns identified in objective 1.

3. To examine the role that environmental-matching can play in determining successful and failed introduction events.

4. To determine whether hybridisation may promote range expansion under conditions relevant to climate change.

5. To further our understanding of the evolutionary forces in play during periods of secondary contacts.
To address these objectives, the thesis presented is made up of four data chapters outlined as follows:

**Chapter Two** contributes towards objectives 1 and 2. I use a population genomic approach to understand how the genomic makeup of the native range affects the introduced range. Additionally, this chapter contributes towards furthering our understanding of how historic range shifts and secondary contacts may bring about patterns of biodiversity observed currently. This Chapter is published in Evolutionary Applications as:
Chapter 1


Author contributions: JH, KJ, and MR designed the study; JH and KJ collected individuals. JH collected samples, performed DNA extraction and measured quality and concentration. JH filtered the genomic data and wrote the code required for the population genomic analyses. JH led the writing of the paper and wrote the first draft; KJ, CDM, and MR commented on the manuscript and assisted with subsequent drafts.

Chapter Three contributes towards objectives 1, 2, and 3. I use a combination of genomic and environmental data to clarify the invasion route, evolutionary history, and reasons behind the failed and successful species introductions. This chapter provides evidence that using a multifaceted approach offers advantages to studying marine invasions. This Chapter is prepared for submission for publication as:

Hudson, J., Castilla, J.C., Teske, P.R., Beheregaray, L.B., Haigh, I.D., McQuaid, C.D., & Rius, M. Using oceanographic and genomic data to study regionally constrained distributions of non-indigenous species.

Author contributions: MR, PT, and JCC conceived the study; MR and JCC collected samples; collected samples, performed DNA extraction and measured quality and concentration. JH filtered the genomic data, wrote the code required for the population genomic analyses, and analysed the genomic data. IDH assisted in collection of temperature data; JH analysed temperature data. JH led the writing of the manuscript; MR, PT, JCC, CDM, IDH have all commented on the chapter.

Chapter Four contributes towards objective 4. By performing controlled crosses this chapter enables further understanding of the potential combination hybridisation and CCC can play in determining species ranges. This Chapter is published in the Journal of Evolutionary Biology as:


Author contributions: MR conceived the study; JH, CDM, MR designed the study; JH collected individuals, housed them in aquaria, and performed laboratory crosses; JH analysed data, with MR contributing towards later development of analysis; JH led the writing of the manuscript; CDM and MR commented on the manuscript and assisted with subsequent drafts.
Chapter Five contributes towards objective 5. I conducted an innovative analysis to understand the drivers of strong population structure over short distances. This chapter facilitates our understanding of genomic divergence during periods of secondary contact and enables the further clarification of the different evolutionary forces at play under these unique conditions. This Chapter is prepared for submission for publication as:

Hudson, J., Johannesson, K., McQuaid, C.D., Chapman, M.A., & Rius, M. Genomic signatures of local adaptation to different depths in a dominant marine invertebrate.

Author contributions: JH and MR designed the study; JH collected samples, performed DNA extraction and measured quality and concentration; JH filtered the genomic data and performed subsequent analyses; JH and led the writing of the chapter; KJ, CDM, MC, and MR commented on the chapter.
Chapter 2  Secondary contacts and genetic admixture shape colonization by an amphiatlantic epibenthic invertebrate

This chapter is a reproduction of text published with *Evolutionary Applications* in the special issue titled “An evolutionary perspective on marine bioinvasions: evolutionary history, adaptation, and species interactions”. As such, this chapter is written in the style of the journal.

2.1  Abstract

Research on the genetics of invasive species often focuses on patterns of genetic diversity and population structure within the introduced range. However, a growing body of literature is demonstrating the need to study how native genotypes affect both ecological and evolutionary mechanisms within the introduced range. Here, we used genotyping-by-sequencing to study both native and introduced ranges of the amphiatlantic marine invertebrate *Ciona intestinalis*. A previous study using microsatellites analysed samples collected along the Swedish west coast and showed the presence of genetically distinct lineages in deep and shallow waters. Using 1,653 single nucleotide polymorphisms (SNPs) from newly collected samples (285 individuals), we first confirmed the presence of this depth-defined genomic divergence along the Swedish coast. We then used approximate Bayesian computation to infer the historical relationship among sites from the North Sea, the English Channel and the northwest Atlantic and found evidence of ancestral divergence between individuals from deep waters off Sweden and individuals from the English Channel. This divergence was followed by a secondary contact that led to a genetic admixture between the ancestral populations (i.e., deep Sweden and English Channel), which originated the genotypes found in shallow Sweden. We then revealed that the colonization of *C. intestinalis* in the northwest Atlantic was as a result of an admixture between shallow Sweden and the English Channel genotypes across the introduced range. Our results showed the presence of both past and recent genetic admixture events that together may have promoted the successful colonizations of *C. intestinalis*. Our study suggests that secondary contacts potentially reshape the evolutionary trajectories of invasive species through the promotion of intraspecific hybridization and by altering both colonization patterns and their ecological effects in the introduced range.
2.2 Introduction

It is well established that attributes of nonindigenous species (NIS) such as genetic diversity (Dupont et al., 2003), founder group size (Lockwood et al., 2005), inbreeding depression (Roman & Darling, 2007) and genetic admixture (Verhoeven et al., 2011) influence their colonization success. These attributes are not mutually exclusive and often combine to allow or deter species introductions (Rius et al., 2015b). In addition, genetic data are critical for (a) reconstructing invasion routes, (b) identifying the source population(s) and (c) understanding how anthropogenic factors affect colonization success (Estoup & Guillemaud, 2010; Cristescu, 2015). Despite a great deal of recent research on invasion genetics (Bock et al., 2015; Rius et al., 2015b; Bourne et al., 2018), there remains a dearth of studies investigating how genetic patterns in the native range influence the introduced range.

Biological invasions act as unique experiments in evolution (Yoshida et al., 2007), allowing observations of how NIS spread and adapt to novel environments on a human timescale. The genetic study of NIS furthers our understanding on how contemporary gene flow and local adaptation contribute to colonization success (Verhoeven et al., 2011). In addition, studies of NIS have shown that genetic admixture of divergent lineages can affect fitness of colonizing populations through transgressive segregation (Johansen-Morris & Latta, 2006; Wagner et al., 2017), by masking deleterious mutations (Keller & Waller, 2002), and/or by increasing standing genetic variation on which selection can act (Rius & Darling, 2014). Genetic admixture can also disrupt locally adapted gene pools, which may negatively affect colonization success (Gilk et al., 2004). Therefore, understanding how ecological and evolutionary mechanisms influence colonization success is key for unravelling how genetic patterns found in native and introduced ranges relate. Research progress on the evolutionary effects of NIS has largely been dominated by studies conducted in terrestrial ecosystems (Abbott, 1992; Rius & Darling, 2014), with considerably less effort being devoted to study aquatic organisms.

Ascidians (Chordata, Tunicata, Asciidiacea) are marine sessile invertebrates that are notoriously invasive (Lambert & Lambert, 1998) and frequently foul aquaculture facilities (Rius et al., 2011; Fitridge et al., 2012) and marine infrastructures (Johnston et al., 2017). The early life-history stages of ascidians are ephemeral and represent the only dispersive stages of their life cycle (Millar, 1971), offering only highly restricted natural dispersal. Thus, long-distance dispersal of ascidians is attributed to artificial transport (Hudson et al., 2016) or rare rafting events (Carlton et al., 2017). As such, they are relevant and unique models for studying colonization success in
marine ecosystems (Zhan et al., 2015). *Ciona intestinalis* is a solitary ascidian with a disjunct amphiatlantic (i.e., inhabiting both sides of the Atlantic) distribution throughout the North Atlantic Ocean (Bouchemousse et al., 2016a). It is generally accepted that the northeast Atlantic coastline is its native range (Gulliksen & Skjæveland, 1973; Bellas et al., 2003; Bouchemousse et al., 2016a; Nydam et al., 2017), while the introduced range includes the north-west Atlantic coastline (Nydam & Harrison, 2007; Bouchemousse et al., 2016a). As with all solitary ascidians, *C. intestinalis* is hermaphroditic and reproduces through broadcast spawning, with external fertilization. The short-lived pelagic larval stage normally lasts <24 hr, though this stage can be extended to five days (Petersen & Svane, 1995). Larvae of *C. intestinalis* are often retained close to the adults and the production of adhesive mucus strings together with the eggs (Svane & Havenhand, 1993) may result in lower dispersal potential. Consequently, transcontinental dispersal of *C. intestinalis* is attributed to anthropogenic transport or rafting events of individuals only. *Ciona intestinalis* shows a high affinity for marine infrastructures (e.g., pontoons and ropes in harbours and marinas), which are known to concentrate NIS (Aldred & Clare, 2014). This propensity to foul can lead to negative economic and ecological impacts when this species is found in aquaculture facilities (Lutz-Collins, 2009; Rius et al., 2011; Fitridge et al., 2012). Consequently, most research studying the extensive distribution of *C. intestinalis* has been performed considering individuals found on artificial structures (e.g., Zhan et al., 2010; Bouchemousse et al., 2016a; c; Hudson et al., 2016). This has led to a good understanding of the distribution of *C. intestinalis* on artificial structures, but there is still limited knowledge of the relative importance of natural and artificial habitats for the spread and establishment of this species in new areas.

The west coast of Sweden is a coastline where *C. intestinalis* is present on natural substrata from the surface to depths of more than 100m (Dybern, 1965, 1967; Svane & Havenhand, 1993). There, the opening of the brackish waters of the Baltic Sea to the Atlantic means that individuals inhabiting shallow water experience a wide range of salinities (10–30 PSU) and variable temperatures (~0–20°C, Dybern, 1965; Renborg et al., 2014), whereas individuals at depth live in both more constant temperatures and stable, high salinities (~34 PSU). The difference in density between surface and deeper waters leads to a strong pycnocline separating the less saline surface water of the Baltic Sea from the high salinity bottom water (often more than ~10-15 m in depth) from the Atlantic (Johannesson et al., 2018). There are observable differences in the biology and life history of individual *C. intestinalis* found in different depths. For example, individuals inhabiting shallow waters (<15 m) have two generations a year (each spawning period lasting a couple of months) during boreal spring and late summer, whereas deeper individuals (>15 m) have one generation per year, with spawning lasting approximately one month during boreal...
summer (Dybern, 1965). Additionally, there appears to be slight morphological variation across depths, with shallow individuals being smaller and more heavily pigmented than deeper individuals (Dybern, 1965; Svane, 1983). This may be due to genetically driven phenotypic variation. A recent study using microsatellites showed that the deep and shallow water populations of *C. intestinalis* along the Swedish west coast are genetically differentiated (Johannesson *et al.*, 2018). Strong pycnoclines can act as distinct barriers to vertical movement of larvae within the water column (e.g., Gallager *et al.*, 1996), and the existence of genetically distinct populations has tentatively been ascribed to the pycnocline present at ~10-15 m acting as a barrier to reproductive exchange. In addition, local adaptation may contribute to the genetic differences between shallow and deep populations as they are exposed to different conditions, including salinity, temperature, food availability and light. Thus, two distinct populations of *C. intestinalis* separated by an abiotic barrier have evolved along the Swedish west coast.

Here, we used *C. intestinalis* as a model organism to investigate how understanding genetic variability in the native range can help elucidate mechanisms shaping both colonization success and introduction pathways in new ranges. The objectives of the study were to (a) identify fine- and broad-scale population genomic patterns of *C. intestinalis*, (b) reveal evolutionary relationships among individuals collected along coastlines across the range of the species, (c) determine the presence or not of genetic admixture and (d) if admixture was present, infer if it could be associated with successful colonization of novel habitats. We hypothesized that the colonization success of *C. intestinalis* across its introduced range has been affected by the historic divergence of ancestral genotypes, the levels of genetic admixture between divergent lineages, and the intensity of gene flow between native and introduced ranges.
2.3 Materials and Methods

Figure 2.1. North Atlantic coastlines where the samples of *Ciona intestinalis* were collected.

Sampling sites are abbreviated as in Table 2.1, with bold quadrats around site codes representing deep sampling sites. The putative native range in the literature includes Scandinavia, the British Isles and the English Channel, whereas the introduced range includes the northwest Atlantic.

2.3.1 Field sampling

Tissue samples of 285 *C. intestinalis* were collected from 20 sites within the putative native and introduced range of the species (Table 2.1 and Figure 2.1). Samples from Sweden were collected from shallow natural, deep natural, and shallow artificial substrata (see details in Table 2.1), whereas sites outside Sweden were all from shallow artificial substrata. Individuals from natural substrata were sampled by either snorkelling, SCUBA diving or a remotely operated under-water vehicle. Artificial substrata were sampled in marinas by pulling up hanging ropes, submerged buoys and checking the undersides of pontoons. We attempted to leave a distance of at least one metre between each sampled individual to limit the chance of collecting closely related individuals. Once collected, tissue was immediately preserved in 95% ethanol which was periodically replaced until tissue pigment no longer leached into the ethanol. Finally, tissue samples were stored at −20°C until DNA extraction.
Table 2.1. Sampling information for *Ciona intestinalis*.

<table>
<thead>
<tr>
<th>Country</th>
<th>Site name</th>
<th>Code</th>
<th>Latitude (N)</th>
<th>Longitude (E or W)</th>
<th>Depth (category)</th>
<th>Substratum</th>
<th>No. of individuals sequenced</th>
<th>Fis</th>
<th>He</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>Vattenholmen</td>
<td>VAT</td>
<td>58.87°</td>
<td>11.09°</td>
<td>60m (Deep)</td>
<td>Natural</td>
<td>16</td>
<td>0.065</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>Gåseklåvan</td>
<td>GUL</td>
<td>58.31°</td>
<td>11.54°</td>
<td>20-25m (Deep)</td>
<td>Natural</td>
<td>15</td>
<td>0.121</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>Jämningarnas</td>
<td>JAM_D</td>
<td>58.26°</td>
<td>11.39°</td>
<td>17-20m (Deep)</td>
<td>Natural</td>
<td>8</td>
<td>0.086</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td>Kåvra</td>
<td>KAV</td>
<td>58.33°</td>
<td>11.36°</td>
<td>18-22m (Deep)</td>
<td>Natural</td>
<td>16</td>
<td>0.113</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>Burholmen</td>
<td>BUH</td>
<td>58.89°</td>
<td>11.13°</td>
<td>5m (Shallow)</td>
<td>Natural</td>
<td>16</td>
<td>0.139</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>South Koster</td>
<td>KOS</td>
<td>58.88°</td>
<td>11.05°</td>
<td>3-4m (Shallow)</td>
<td>Natural</td>
<td>15</td>
<td>0.078</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>Brattskär</td>
<td>BRA</td>
<td>58.86°</td>
<td>11.07°</td>
<td>1-4m (Shallow)</td>
<td>Artificial</td>
<td>15</td>
<td>0.078</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>Lindholmen</td>
<td>LIN</td>
<td>58.88°</td>
<td>11.15°</td>
<td>0-1m (Shallow)</td>
<td>Artificial</td>
<td>14</td>
<td>0.088</td>
<td>0.242</td>
</tr>
<tr>
<td></td>
<td>Porsholmen</td>
<td>POR</td>
<td>58.23°</td>
<td>11.40°</td>
<td>2-4m (Shallow)</td>
<td>Natural</td>
<td>15</td>
<td>0.056</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>Jämningarnas</td>
<td>JAM_S</td>
<td>58.26°</td>
<td>11.39°</td>
<td>5-7m (Shallow)</td>
<td>Natural</td>
<td>16</td>
<td>0.028</td>
<td>0.206</td>
</tr>
<tr>
<td>Denmark</td>
<td>Fiskebäckskil</td>
<td>FIS</td>
<td>58.24°</td>
<td>11.46°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>15</td>
<td>0.073</td>
<td>0.241</td>
</tr>
<tr>
<td>England</td>
<td>Limfjord</td>
<td>DEN</td>
<td>56.78°</td>
<td>9.18°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>20</td>
<td>0.095</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>Hartlepool</td>
<td>HPL</td>
<td>54.69°</td>
<td>-1.20°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>18</td>
<td>0.089</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>Town Quay</td>
<td>TNQ</td>
<td>50.89°</td>
<td>-1.41°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>12</td>
<td>0.121</td>
<td>0.231</td>
</tr>
<tr>
<td>Jersey</td>
<td>St. Helier</td>
<td>JER</td>
<td>49.18°</td>
<td>-2.12°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>15</td>
<td>0.082</td>
<td>0.217</td>
</tr>
<tr>
<td>France</td>
<td>St. Malo</td>
<td>STM</td>
<td>48.64°</td>
<td>-2.03</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>14</td>
<td>0.117</td>
<td>0.232</td>
</tr>
<tr>
<td>Canada</td>
<td>Yarmouth</td>
<td>YAM</td>
<td>43.83°</td>
<td>-66.13°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>12</td>
<td>0.069</td>
<td>0.229*</td>
</tr>
<tr>
<td></td>
<td>Shelburne</td>
<td>SB</td>
<td>43.76°</td>
<td>-65.32°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>2</td>
<td>0.017$</td>
<td>0.257$</td>
</tr>
<tr>
<td></td>
<td>Brudenell River</td>
<td>BR</td>
<td>46.20°</td>
<td>-62.59°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>9</td>
<td>0.017$</td>
<td>0.257$</td>
</tr>
<tr>
<td></td>
<td>Sydney</td>
<td>SD</td>
<td>46.14°</td>
<td>-60.19°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>2</td>
<td>0.017$</td>
<td>0.257$</td>
</tr>
</tbody>
</table>

265
Note: The table includes geographical region, site abbreviation code, coordinates of sampling sites, depth (shallow [<15m] or deep [>30m]), substratum collected from, and the number of individuals used in genomic analyses. Additionally included are F\textsubscript{IS} values (values in italics are statistically significant [P<0.05]) and population mean expected heterozygosity (H\textsubscript{e}). * refers to samples merged and known as CAN_1 (Yarmouth and Shelburne), § refers to merged samples known as CAN_2 (Brudenell River and Sydney).
2.3.2 DNA extraction and genotyping

DNA was extracted from preserved tissue using the Qiagen DNeasy® Tissue Kit (Qiagen) according to the manufacturer's protocol. Gel electrophoresis and the QuantiFluor® dsDNA System (Promega) were used to assess quality and quantity of extracted DNA, respectively. DNA was shipped to the University of Wisconsin Biotechnology Center where it was genotyped using the genotyping-by-sequencing methodology (GBS; Elshire et al., 2011). Briefly, GBS reduces the complexity of the sample genome by digesting the DNA using methylation-sensitive restriction enzymes and sequencing the ends of the digested fragments using barcoded adapter regions.

2.3.3 Analysis of genotyping-by-sequencing data

The GBS assembly was performed using ipyrad v. 0.7.28 (Eaton, 2014), a toolbox for assembly and analysis of restriction site-associated DNA sequencing (RAD-seq) type genomic data sets. We followed the seven sequential assembly steps of ipyrad using parameters based on those recommended for single-end GBS data in the ipyrad documentation (http://ipyrad.readthedocs.io/). As the C. intestinalis genome is not yet available, we used the de novo assembly method, which requires no prior genomic resources and used ipyrad to trim Illumina adapter reads. As we were working with only one species, we set the level of sequence similarity for clustering to be 90% (I. Overcast, pers. comm.). Following the iterative filtering framework outlined by O’Leary et al. (2018), we used vcftools v.0.1.13 (Danecek et al., 2011) to first filter for loci with a minimum single nucleotide polymorphism (SNP) call quality of 20, a minimum genotype depth of less than five, and a mean minimum depth (across individuals) of <15. Additionally, we chose to remove loci with a minor allele count of less than three, rather than the commonly used minor allele frequency threshold of 5%, because the latter will remove true rare alleles that are important in elucidating fine-scale structure and accurately drawing inference of past demographic events (O’Connor et al., 2015). We then iteratively increased our stringency for allowing missing data (on both loci and individuals separately), so that our final dataset contained loci with at least 50% call rate (i.e., a locus must be present in at least 50% of individuals), and up to 50% allowed missing data per individual. To remove the confounding effects of linkage disequilibrium, we used vcftools to thin markers so that only one SNP per locus was retained in our dataset.

We used BayeScan v.2.0 (Foll & Gaggiotti, 2008) to ensure our dataset contained only putatively neutral loci. BayeScan uses differences in allele frequency between populations to identify candidate loci under natural selection (Foll & Gaggiotti, 2008) and was run using a thinning
interval size of 20, with 25 pilot runs of length 10,000 and a burn-in length of 50,000. Prior odds for the neutral model were set to 100 rather than the default 10, to reduce the number of false positives in large datasets (>1,000 SNPs). While commonly used, BayeScan has often been shown to report false positives especially in species undergoing range expansions, while also assuming equal population exchange and evolutionary independence among all populations (Bierne et al., 2013; Whitlock & Lotterhos, 2015). We therefore also assessed for candidate loci using two newer methods, OutFlank v.0.2 (Whitlock & Lotterhos, 2015) and pcadapt v.4.1.0 (Luu et al., 2017).

Similarly to BayeScan, OutFlank groups individuals into predefined populations, before inferring candidate loci based on a trimmed distribution of $F_{ST}$ values for loci deemed to be neutral. As reported in similar studies (see results and Guzinski et al., 2018), OutFlank did not recover any $F_{ST}$ outlier loci, so we continued our analyses with other software. Regarding pcadapt, it ascertains population structure using principal component analysis (PCA) to find candidate loci excessively related to population structure. We classified loci that were recovered by both BayeScan and pcadapt as putatively under natural selection and removed them for the following analyses. Finally, we created a more conservative dataset that excluded all loci recovered by BayeScan and pcadapt and ran the whole set of analyses again (see Appendix A).

### 2.3.4 Population structure

We used the software ADMIXTURE v.1.3 (Alexander et al., 2009) to estimate the likelihood that an individual comes from one of a certain number of putative sample populations ($k$). Like STRUCTURE v.2.3.4 (Pritchard et al., 2000), ADMIXTURE uses a maximum-likelihood estimation from multilocus SNP genotype datasets, but calculates estimates using a faster numerical optimization algorithm. We performed a discriminant analysis of principal components (DAPC) to visualize between-population genomic variation (Jombart et al., 2010). DAPC transforms the data using PCA before using PCA factors as variables for a discriminant analysis (DA), ultimately maximizing the differences among groups while minimizing variation within groups (Jombart et al., 2010). We used the package adegenet v.2.1.1 (Jombart, 2008) for R (R Core Team, 2016) to perform the DAPC. We ran the DAPC with and without a priori knowledge of individual populations. We examined pairwise population genetic differentiation using $F_{ST}$ values and their p values by running 10,000 permutations with Arlequin v.3.5 (Excoffier & Lischer, 2010). We also used Arlequin to measure the inbreeding coefficient $F_{IS}$ and expected heterozygosity ($H_e$) per population. Finally, we ran an analysis of molecular variance (AMOVA) test using site clusters as inferred by ADMIXTURE and DAPC plots and also using only shallow Sweden sites to test whether there was an effect of substratum (natural vs. artificial). AMOVAs were performed in Arlequin v.3.5 (Excoffier & Lischer, 2010).
2.3.5 Reconstructing invasion routes

To obtain relevant and detailed information and infer the historical relationship among genotypes of *C. intestinalis* throughout its range, we analysed sets of evolutionary scenarios with the approximate Bayesian computation (ABC) method using DIYABC v.2.0.1 (Cornuet *et al.*, 2014). We grouped sites based on their geographical location and the results of above population structure analyses (shallow Sweden sites plus the Denmark site, deep Sweden sites, England, Jersey and France sites, and Canada sites). The high shipping traffic between the native and introduced ranges (Kaluza *et al.*, 2010), coupled with the similarity in genetic diversity across our sampled sites (see Results), meant we did not consider the presence of genetic bottlenecks while designing the evolutionary scenarios. Our first two sets of scenarios aimed to infer the evolutionary history within the northeast Atlantic (see details of scenario sets 1 and 2 in Figures 7.3A and B, respectively). Following the results of this initial analysis, we then added in data from Canada sites to infer the colonization history along the introduced range (scenario set 3, Figure 7.3c). As specific population sizes, divergence times and potential admixture rates were unknown, we used a uniform distribution with a large interval (population sizes and divergence times: 10–10⁷; admixture rates: 0.001–0.999; Table 7.5) when setting priors for each parameter (White *et al.*, 2018). We used the mean genic diversity, mean distribution of $F_{ST}$ distances, mean distribution of Nei distances, and whenever an admixture event was included in the scenario, mean admixture estimates for summary statistics. For all scenarios, we used the default 10⁶ simulated data per scenario to build reference tables. Upon creation of the reference table, we pre-evaluated scenarios and prior distributions by performing a PCA in the space of the summary statistics on 1,000 simulated datasets for each scenario and adding the observed dataset to each plane (Cornuet *et al.*, 2014). We used a logistic regression on the 1% simulated datasets that were closest to the observed dataset (using Euclidean distances between simulated and observed datasets) to calculate the posterior probability of each scenario. This approach produces 95% confidence intervals for each scenario’s posterior probability, with the most likely scenario defined as the highest estimate without overlapping confidence intervals (Cornuet *et al.*, 2008).

For the most probable scenario of scenario set 3 (Figure 7.3C), we calculated type I (the probability with which this scenario is rejected although it is the true scenario) and type II (the probability of choosing this scenario when simulating data according to other scenarios) error rates. Finally, we assessed the goodness of fit for the final chosen scenario by implementing the model checking feature of DIYABC. We simulated 1,000 datasets using posterior distribution values and compared these with the observed dataset by considering different summary statistics than were used during the generation of the reference table, and visualized this using a PCA (Cornuet *et al.*, 2014).
2.4 Results

2.4.1 Loci assembly and detection of outlier loci

GBS generated a total of 530,157,826 raw reads, with an average of 2,000,596 reads per sample. After filtering and clustering using ipyrad and vcftools, we retained a total of 1,667 putatively unlinked SNPs in the sequence assembly. Twenty individuals were removed from the dataset due to missing data (i.e., >50% missing data), which was likely caused by poor DNA quality or secondary contaminants within the samples (Federman et al., 2018). This led to a final dataset of 265 individuals from 20 separate sampling sites. However, the Canadian sites were merged as CAN_1 (sites Yarmouth and Shelburne) and CAN_2 (Brudenell River and Sydney) due to the limited number of individuals obtained from Shelburne and Sydney. Therefore, the total final number of sites was 18. BayeScan and pcdadapt recovered a total of 30 and 61 $F_{ST}$ outlier loci, respectively, of which 14 were found by both software, whereas OutFlank recovered no putative loci under selection. We subsequently removed the 14 loci found in both BayeScan and pcdadapt from our analyses, leaving a dataset of 1,653 SNPs. We also performed all analyses on a new dataset that excluded all $F_{ST}$ outlier loci recovered, irrespective of program used (77 loci in total, see Appendix A for details).

2.4.2 Heterozygosity and population structure

Values of $F_{IS}$ ranged from 0.017 to 0.139 (Table 2.1). Nine sites showed no signs of deviation from Hardy–Weinberg equilibrium, but nine of the sites exhibited significant positive $F_{IS}$ values indicating a deficiency of heterozygotes in these sites (Table 2.1). Expected heterozygosity ranged from 0.202 to 0.286 (Table 2.1), with no noticeable differences in genetic diversity between geographical regions (Table 2.1).
Figure 2.2. ADMIXTURE plots representing all sampled populations of *Ciona intestinalis*. The main regions are highlighted above. The different colours represent putative genetic clusters with $K$ ranging from 2 to 7, with $K=4$ being found to be the most optimal value (Figure 7.1).

The combination of ADMIXTURE, DAPC and pairwise site comparisons of $F_{ST}$ allowed us to identify fine- and broad-scale population genomic patterns. Cross-validation by ADMIXTURE inferred the most likely number of sampled populations was $K = 4$ (Figure 7.1) and broadly indicated the structuring of deep Sweden sites (green in Figure 2.2), shallow Sweden sites (orange) and those found in England, Jersey and France (blue). Individuals from Canada appeared to have a genetic background similar to both individuals found in England, France, Jersey and individuals from shallow Sweden. The Denmark samples clustered with a shallow Sweden site (shallow Jämningarna, purple cluster), and eight individuals from the shallow Sweden site Burholmen were grouped with samples from deep Sweden.
Figure 2.3. Discriminant Analysis of Principal Components using unlinked loci. (A) With no *a priori* population information. The first axis explain 58.2% of the variation, and the second axis explains 41.8%. (B) With *a priori* population information. The first axis explains 23.1% of total variation, and the second axis explains 18.1%. Site abbreviation as in
Table 2.1. Sites in (A) are assigned to clusters as follows; Cluster 1: FIS, KOS, BRA, LIN, POR, BUH, JAM_S, DEN, CAN_1, CAN_2, and eight individuals from BUH; Cluster 2: JER, TNQ, HPL, STM; Cluster 3: VAT, JAM_D, GUL, KAV, and eight individuals from BUH.

The ADMIXTURE patterns were supported by the DAPC analysis with and without prior sample assignment (Figures 2.3A and B), which recovered three genetic clusters, one of which (cluster 3) included the shallow Sweden sites (except the eight individuals from BUH) and sites from Denmark and Canada, and the other two clusters including deep Sweden sites and sites from England, Jersey and France, respectively.

Figure 2.4. Matrix of $F_{ST}$ values. Asterisks represent significant values, after Bonferroni correction. Sampling sites are abbreviated as in Table 2.1.

Pairwise comparisons of $F_{ST}$ suggested very strong genetic structuring among most sites (Figure 2.4), with 143 out of 153 comparisons (93%) being significant, including clear structuring between shallow and deep Sweden sites (Figure 2.4, Table 7.1). Notably, there was significant genetic differentiation among the deep Sweden sites with the exception of two deep sites (Kåvra vs. deep Jämningarna) that are very close to one another. Pairwise site comparisons among shallow
Sweden sites found 15 of 21 comparisons (71%) were significant, while comparison of the two Canadian sites provided a low, but significant, $F_{ST}$ value (Figure 2.4, Table 7.1).

The AMOVA test using site clusters as inferred by ADMIXTURE and DAPC plots showed that genetic differentiation was significant among groups, among sites within groups and within sites (Tables 7.2 and 7.3). The AMOVA test performed using only shallow Sweden sites to test whether there was an effect of substratum (natural vs. artificial) found no significant genetic differentiation between these two groups (Table 7.4).

### 2.4.3 Reconstructing invasion routes

![Figure 2.5. The most likely evolutionary scenario involving the sample sites as calculated using DIYABC. The Y axis represents time (not to scale). Included are shallow Sweden sites (SS), Canada sites (CAN), the Denmark site (DEN), England, Jersey, and France sites (EJF), and deep Sweden sites (DS).](image)

For all of our ABC analyses, our check of priors showed a good match between simulate datasets and the observed data (Figure 7.2). We firstly found that within our northeast Atlantic sampling sites, the ancestral population diverged and formed the deep Sweden and England, Jersey, France groups (Figure 7.3A), with the logistic estimate of posterior probability for this scenario being $p = .996$ (CI = 0.994, 0.997; Table 7.6). For our next set of scenarios, which assessed the origin of the shallow Sweden group (Figure 7.3B), we found the scenario with the highest support being an admixture event between deep Sweden and England, Jersey, and France groups ($p = .998$, CI = 0.998–0.999; Table 7.6) following secondary overlap of the two lineages. Our final set of scenarios, which assessed the scenario that best explains the colonization of the introduced range
Chapter 2

(Figure 7.3C), found that the most likely was a recent admixture between shallow Sweden sites and English, Jersey and France sites (p = .841. CI = 0.832, 0.851; Table 7.6; Figure 2.5). The type I error rate was 0.15, showing that 85% of our datasets simulated with the highest supported scenario (Figure 2.5) were correctly identified as being produced by the same scenario. Moreover, type II error rate was on average 0.04. Our model checking procedure for the most likely scenario found that for the 57 summary statistics used for model checking, 23 differed significantly from the simulated distribution (Table 7.7, Figure 7.4), suggesting that even though this is the most strongly supported scenario, there is some discordance between the scenario posterior combinations and the observed dataset.

2.5 Discussion

Our results showed high levels of genomic differentiation between the main regions of the northeast Atlantic (i.e., English Channel and North Sea) and identified the presence of historical genetic admixture among individuals from these regions. This seems to have resulted in genotypically and phenotypically distinctive individuals that are currently found in shallow sites in Sweden. In addition, we revealed genomic patterns suggesting secondary contacts and postulate that this may have promoted intraspecific hybridization. Our result supported the presence of genetic admixture during the spread to and colonization of the northwest Atlantic. More specifically, we found evidence of genetic admixture between genotypes from the English Chanel and genotypes from the shallow North Sea. While we found here no direct evidence that intraspecific hybridization influences colonization, our results indicate that this may be a possible mechanism promoting successful colonization of sites with new environmental conditions, such as trans-oceanic introductions.

This builds on a growing number of studies showing that the mixing of divergent genotypes as a result of human mediated transport of species has the potential to fundamentally alter colonization patterns and to unprecedentedly alter ecological and evolutionary patterns (Mooney & Cleland, 2001; Pineda et al., 2011; Bouchemousse et al., 2016c).

The presence of high genetic subdivision among genotypes found in deep sites off Sweden and in England, Jersey and France suggests that individuals found in these sites represent native populations (Figure 2.3). This is supported by the ABC analyses, which indicated an initial divergence between these two groups (Figure 2.5). This accords with the expectation that native ranges will show a highly defined population structure, often involving two main groups of ancestral genotypes (Reusch et al., 2010; Boubou et al., 2012; Rius et al., 2014a). Divergence of these C. intestinalis populations may reflect adaptation to differing local conditions and/or earlier
periods of allopatric isolation leading to the generation of genetic divergence through selection or genetic drift. Previous research has shown that high plasticity in *C. intestinalis* allows acclimatization of deep water individuals to shallow water salinities (Renborg *et al.*, 2014), which suggests that local adaptation and primary divergence are less likely. Rather, it seems more likely that during the last glacial maximum deep sites off Sweden and sites in England, Jersey and France were isolated as separate glacial refugia, leading to the divergence that we see today. ABC analyses suggest that secondary contacts leading to genetic admixture between the England, Jersey, France and the deep Sweden genotypes formed the genotypes found in shallow Sweden and Denmark. Thus, either historic artificial transport or postglacial expansion may have promoted such secondary contacts, as reported for other marine invertebrates (Pérez *et al.*, 2006). A similar situation has been identified with the ascidian *Pyura chilensis* in the southwest Pacific Ocean, where historical divergence of *P. chilensis* populations occurred due to isolation associated with glacial periods. This was followed by secondary contacts and genetic admixture between these previously isolated populations (Haye & Muñoz-Herrera, 2013). Our results suggest that genetic admixture may have had fitness effects that enabled *C. intestinalis* to expand to previously uninhabitable substrata and conditions within its native range. The ability of individuals from shallow Sweden populations to survive relatively high temporal variability in environmental conditions such as temperature and salinity compared to individuals found in England, Jersey, France and deep Sweden may be explained by the fitness benefits of genetic admixture (Wagner *et al.*, 2017), allowing survival in the face of strong selective pressures (Verhoeven *et al.*, 2011).

In contrast to the Sweden populations, individuals from England, Jersey and France formed a relatively homogeneous genetic cluster in both ADMIXTURE and DAPC analyses (Figures 2.2–2.4). Earlier work showed that samples from these locations were subdivided into two genetic groups (Hudson *et al.*, 2016), but this differentiation was weaker than what we found between samples from deep Sweden and those from England, Jersey and France. The native range of *C. intestinalis* has been previously described as the northeast Atlantic (e.g., Bouchemousse *et al.*, 2016a; Hudson *et al.*, 2016), and here, we show that this range comprises most of the genomic differentiation among populations, with more complex demographic histories among populations along the northeast coast of the Atlantic than the northwest coast. In line with previous studies that identified admixture within native ranges (Gillis *et al.*, 2009; Rius *et al.*, 2012), our findings suggest that historic artificial transport may have facilitated the admixture of the genotypes from deep Sweden and the English Channel.

It is well established that the recent spread of *C. intestinalis* has been promoted by the proliferation of man-made structures along coastlines that act both as stationary substrata and as
vectors (Clarke Murray et al., 2014). Throughout the study area, this species is widespread in harbours and marinas in much of its current distribution (Bouchemousse et al., 2016a; Hudson et al., 2016), but rare on natural substrata except along the Swedish coast (Johannesson et al., 2018). This raises the issue of the origin and evolutionary background of the many populations living on artificial substrata that may represent either extensions of large natural populations or completely new introduced populations. In the sampled shallow Sweden sites, comparison of genetic differentiation between natural and man-made sites showed no significant differences (Table 7.4), with individuals sampled on artificial substrata being generally more closely related to nearby shallow natural sites than to individuals from other artificial sites along the Swedish coast (Figure 2.4; Table 7.1). This suggests that the nature of natural and artificial substrata does not, in itself, create a barrier to local gene flow. In shallow waters off Sweden, the sampled natural substratum included seagrass beds where *C. intestinalis* lives at modest densities attached to blades of *Zostera marina*. Below the pycnocline, at depths of 20 m or more, dense populations occur on the natural vertical rock walls (see also Svane & Havenhand, 1993), whereas in the English Channel and south-western North Sea area, very few individuals have been documented on natural substrata. To our knowledge, there are no reports of *C. intestinalis* inhabiting natural substrata in the English Channel. However, small numbers have been recovered during dredging estuaries in the English Channel (authors pers. obs.) and the south-western North Sea (Rees et al., 2001). Such low densities may be due to the effects of predation on different life-history stages as seen for closely related species in other parts of the world (Dumont et al., 2011; Rius et al., 2014b).

Our study corroborated the findings of Johannesson et al., (2018) by identifying strong genetic differentiation between shallow and deep populations of *C. intestinalis* along the west coast of Sweden. While this genomic differentiation among populations appears surprising as some deep and shallow sites are geographically close to one another, this can be explained due to the effect of the aforementioned pycnocline promoting depth-defined divergence that has also been observed in corals (Prada & Hellberg, 2013). Admixture between deep Sweden and England, Jersey, France genotypes, as indicated by our ABC analysis, suggests that the pycnocline may not have always been the impenetrable barrier to gene flow as currently observed (Johannesson et al., 2018). Taken together, our results suggest that although historically the pycnocline may have allowed the mixing of divergent genotypes, it currently provides a strong barrier to gene flow, maintaining contemporary genomic differentiation between deep and shallow sites.

The northwest Atlantic range of *C. intestinalis* is restricted to the east coast of North America. It has been documented in eastern Canada since at least the mid-1800s (Carver et al., 2006), but its population size and range have only recently expanded (Ramsay et al., 2008). Our ABC analyses
suggest that the origin of the Canadian sites was due to secondary contact between populations from England, Jersey, France and shallow Sweden. This is supported by our ADMIXTURE analysis, which indicates high similarity between the individuals from Canada and the ones from both the English Channel and shallow sites within the North Sea (Figure 2.2). In addition, the DAPC analyses indicated that Canadian individuals were similar to individuals found in shallow Swedish waters (Figure 2.3). This interpretation accords with previous studies showing that multiple introductions facilitate marine biological invasions (Simon-Bouhet et al., 2006; Rius et al., 2015b) and that recurrent introductions of large numbers of individuals explain patterns of genetic diversity within introduced ranges (Uller & Leimu, 2011). Indeed, our results do not show a noticeable change in genetic diversity between Canadian and European sites (Table 2.1) and are compatible with contemporary introgression among divergent English Channel and North Sea genotypes followed by multiple introductions to Canadian sites.

Heterozygote deficiency at nine of our sites ($F_{IS}$ values significantly greater than zero; Table 2.1) could reflect either selection against heterozygotes or non-random mating. We can reject selection against heterozygotes as we excluded loci putatively under selection, and we can also exclude selfing as self-fertilization success is generally low in *C. intestinalis* (Byrd & Lambert, 2000; Bouchemousse et al., 2016b). A more likely explanation is a Wahlund effect (Wahlund, 1928), a reduction of expected heterozygosity due to mixing of two genetically differentiated populations, which has been reported in other studies (Dupont et al., 2009; Marescaux et al., 2015) including studies of *Ciona* spp. (Zhan et al., 2010; Hudson et al., 2016).

The history of the introduction of *C. intestinalis* to the western Atlantic coast is complex, starting with historical divergence in the native range involving two groups (England, Jersey, France and deep Sweden lineages, Figure 2.5), which was likely due to genetic drift during a period of isolation (allopatry) in different glacial refugia. More recent historic gene flow between these populations appears to have led to the formation of the admixed genotypes found in shallow Sweden and Denmark sites. Finally, the Canada specimens originated from secondary contacts between individuals from these sites and individuals from the western North Sea and English Channel. Our findings suggest that admixture between genetically diverse native genotypes preceded successful trans-oceanic colonization, in line with previous studies showing that genetic admixture facilitates the colonization of new habitats (Abbott et al., 2016). We suggest that artificial transport of species facilitates secondary contacts and intraspecific admixture among divergent native genotypes, strongly altering NIS evolutionary trajectories and influencing their ecological impacts within the introduced range.
Chapter 2

2.6 Acknowledgements

We would like to thank the guest editors who kindly invited us to contribute to this special issue and also the anonymous reviewers whose comments and suggestions greatly improved the manuscript. We would like to thank the divers at Tjärnö Marine Laboratory and Kristineberg Marine Station, and Dr Ellen Schagerström, for their help in collecting samples. We would also like to extend our gratitude to Dr Anna-Karin Ring for her assistance in the laboratory. Finally, we would like to thank Steven Bourne and Rachael Graham for guidance with the bioinformatic analyses. We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. We would also like to thank Prof Albin Zhan, Dr Camille Saurel and Prof Lene Møller Madsen for providing tissue samples. This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation. We are grateful to the KVA fund for internationalization and scientific renewal at the Sven Lovén Centre for Marine Sciences for awarding a grant to MR that supported JH stay at the Tjärnö Marine Laboratory.
Chapter 3  

Using oceanographic and genomic data to study regionally constrained distributions of non-indigenous species

3.1  Abstract

Invasive species are those that are transported to a location away from their native range, where they undergo further spread from their initial point of introduction. This secondary dispersal to neighbouring regions largely contributes to the significant impacts observed in most notorious invasive species. Although many invasive species show widespread distributions, some have regionally constrained introduced ranges. These dissimilar distributions are often explained by different stages of the invasion process, but they can also be as a result of disparities between realised and potential ranges. Restricted distributions suggest great potential for spread of the invasive species but few studies to date have assessed this. Here we studied the population genomics of the bioengineering marine invertebrate, *Pyura praeputialis*, that has only been reported as introduced in a single bay in Chile. Despite its restricted distribution, it causes important ecological impacts in the introduced range and displays the largest biomass of an intertidal metazoan in the world. Our genomic data showed strong population structure within the native range, with two genetically distinct lineages separated by a biogeographic break. We found no significant differences between native and introduced sites in terms of genomic diversity. We then used Bayesian methods to unravel the most likely invasion route and found a clear introduction from the eastern lineage of Australia to Chile. To assess environmental matching between native and introduced ranges, we compared sea surface temperature data and found high thermal suitability for *P. praeputialis* along most of the Chilean coastline. When we analysed the conditions across the restricted distribution of the bay in Chile, we found high temperatures in the north of the bay suggesting the possibility of increased suitability and larval retention around that area. Our study showed how oceanographic and genomic data can help understanding the potential distribution of invasive species, and how if conditions change, these highly invasive species with restricted distributions have a great spread potential. Studies focussing on environmental matching and population genomics are thus essential to enhance our predictions of distributions of invasive species.


3.2 Introduction

Understanding the ecological and evolutionary mechanisms that shape species distributions, community assembly, and species coexistence is fundamental for understanding species ranges. These mechanisms can help explain why certain species have widespread distributions and others very restricted ones (Brown et al., 1996), and how species spread to new regions (Bridle & Vines, 2007). For example, factors including local adaptation (Atkins & Travis, 2010), genetic structure and diversity (Connallon & Sgrò, 2018), physiological tolerance (Calosi et al., 2010), and dispersal (Lester et al., 2007) have been shown to influence range size. Non-indigenous species (NIS) offer unique opportunities to witness how these mechanisms and factors influence the way species spread and establish new ranges. However, these species have rarely been used to understand how genomic attributes and environmental matching shape the size of their newly established ranges.

The process by which a species becomes invasive can be thought as a continuum from initial transportation to full biological invasion via conceptual barriers that must be overcome (Richardson et al., 2000; Blackburn et al., 2011). Generally, biological invasions can be broken down to four stages along this continuum: transportation from the native range, introduction to a novel area, establishment within this new location, and further spread to neighbouring areas (Bourne et al., 2018). As a result, there has been a drive to understand the mechanisms facilitating successful progression between each stage of the invasion process (Richardson & Pyšek, 2012). It is difficult to study early stages of this process though, as often species are not recognised as being invasive until they have reached the final stages of the continuum. This is supported in a literature review of 873 biological invasion articles by Puth & Post (2005), which found only 11% of studies concentrated on initial dispersal. It is therefore unsurprising that progress in invasion science is often led by studies focussing on highly invasive species [i.e. species that expand their range from this initial point of introduction (Richardson et al., 2000)] that are, in many cases, widely distributed (Moran & Alexander, 2014). These species are often the primary focus of research because of their potential ecological and economic effects (Bax et al., 2003; Molnar et al., 2008). In addition, some conceptual frameworks of invasion management suggest that impact is a unique invasion processes in of itself, occurring only once a species has spread from its initial point of introduction (Catford et al., 2009). Consequently, relatively little research is afforded to other aspects of the invasion continuum, such as naturalised species [species that establish self-sustaining populations beyond their native range, but do not expand from the point of introduction (Richardson et al., 2000; Blackburn et al., 2011)], or invasive species with restricted introduced ranges. Despite these species having a small area where they can be considered introduced, they can still have significant effects on ecosystems (Castilla et al., 2004;
Ricciardi & Cohen, 2007; Ricciardi et al., 2013). A unique and understudied area of research includes species that sit between “naturalised” and “widespread invasive”, such as successfully invasive species with a relatively small introduced range. Thus, these species have spread within the region of introduction, but do not exhibit further extensive spread, possibly due to mechanisms limiting the spread including biotic resistance (Harvey et al., 2004) or genetic bottlenecks (Kinziger et al., 2011).

The advent of high throughput sequencing (HTS) techniques has allowed the generation and analysis of genomic data at an unparalleled rate. HTS has increased our ability to study the evolutionary processes facilitating, and the evolutionary consequences of, biological invasions (Rius et al., 2015a). The increased marker coverage associated with genomic data provide enhanced power to address important eco-evolutionary processes associated with the introduction process (Viard et al., 2016). For example, genomic data have assisted the identification of fine-scale genetic structure not captured by previous genetic methods (Benestan et al., 2015; Gagnaire et al., 2015; Tepolt & Palumbi, 2015), have outlined the importance of rapid adaptation for invasion success by identifying genes associated with loci potentially under selection (Bernardi et al., 2016), and unravelled previously unknown roles of hybridisation and secondary contact in evolutionary histories (Bouchemousse et al., 2016c). High-resolution genomic data therefore offers an unprecedented opportunity to study the how current and past ecological and evolutionary events shape the range size of biological invasions.

Knowledge of ecological and environmental niches within the native habitat is often used to predict the potential ranges of introduced species (Peterson, 2003; Calosi et al., 2010). A high similarity between the thermal environments of species’ native and introduced ranges can promote invasion success (Facon et al., 2006), and such environmental matching has been used predict future spread of species (Holcombe et al., 2010). However, the opposite phenomenon, environmental-mismatch, has been posited as a reason behind failed introductions. This occurs when a species is unable to survive or colonise a new area despite being introduced (Bomford et al., 2009; Marshall et al., 2010). Studying failed introductions is inherently difficult as these events are transient and often do not leave an observable trace (e.g. individuals fail to establish). However, several reasons have been put forward to explain failed introductions including low propagule pressure, genetic constraints, and mutualist release (Zenni & Nuñez, 2013). Although studies combining population genomics and a detailed analysis of habitat suitability have the potential to transform our understanding of the distributions of a non-indigenous species (Massatti & Knowles, 2016; Malone et al., 2018), these studies remain rare.
Chapter 3

Here we compared genomic patterns of the native range with geographically constrained introduced populations of a sessile bioengineering marine invertebrate to: i) unravel the evolutionary history within the native range and identify the origin of introduced individuals, ii) reconstruct the invasion history and assess whether intraspecific hybridisation had any role, iii) understand the population genomic attributes of the highly restricted introduced range, and iv) ascertain environmental matching between native and introduced range and assess spread potential within the introduced range. We expected that differences in abiotic conditions (i.e. environmental mismatch) would be a good predictor of regions where failed and successful introductions can be found (extreme temperature conditions may preclude successful introductions). In addition, we predicted that introduced individuals would show a genomic composition that would provide insights into the potential distribution and dispersal within and beyond the current introduced range.

3.3 Methods

3.3.1 Study species and distribution

*Pyura praeputialis* (Heller, 1887), previously known as *P. stolonifera* or *P. stolonifera praeputialis* (e.g. Kott, 1985), is a solitary ascidian that forms packed aggregates and monopolises the intertidal and subtidal of extensive stretches of coastlines, achieving ecological dominance and one of the highest intertidal biomass reported (Castilla et al., 2000). This species was once believed to occur throughout temperate Australia and represented on the African continent (Kott, 1985). However, successive morphological (Monniot & Bitar, 1983; Rius & Teske, 2011) and genetic (Castilla et al., 2002a; Teske et al., 2011; Rius & Teske, 2013; Teske, 2014) studies have shown that the African populations comprise at least two distinct species, and that the Australasian distribution contains three species: *P. praeputialis*, and two other native Australian species, *P. dalbyi* (Rius & Teske, 2011) and *P. doppelganger* (Rius & Teske, 2013). The actual distribution of *P. praeputialis* along the Australian coastline is now well documented: it ranges from Cape Otway in Victoria (south-east coast) to southern Queensland (east coast), and a biogeographic barrier at Wilson’s Promontory separates two regional populations that differ primarily on the basis of allele frequencies (Teske et al., 2011; Rius & Teske, 2013). *Pyura praeputialis* is also present across the Pacific Ocean as a single isolated population exclusively in Antofagasta Bay, Chile (Castilla et al., 2002a; Teske et al., 2011). The idea that the population in Chile is a distinct species that is native to the Pacific coast of South America (Kott, 2006) has been ruled out by genetic studies which show a close relation between the Chilean and eastern Australian population (Castilla et al., 2002a; Teske et al., 2011). It has been proposed that *P.*
*Praeputialis* may have been transported to Antofagasta Bay via international shipping during the mid-late 19th century during the onset of the nitrate trade between Chile and Australia (Maino, 1985; Arce, 1997). Antofagasta Bay is situated in northern Chile within the Humboldt Current System. The bay is approximately 70 km in length with a 35 km wide southern-facing mouth. Within Antofagasta Bay, an ‘upwelling shadow’ (sensu Graham & Largier, 1997) exists, which leads to numerous unique oceanographic characteristics, including stratification due to a shallow thermocline, high concentrations of chlorophyll-a at the base of the thermocline, cyclonic circulation and high retention of water in the bay, plus the existence of a persistent warm-water patch (Escribano & Hidalgo, 2000; Piñones *et al.*, 2007). This warm-water patch is found within the bay immediately downwind of an upwelling centre, and temperatures are on average 2-3°C warmer than at sites outside the bay that are exposed to upwelling (Castilla *et al.*, 2002b; Piñones *et al.*, 2007). In addition, the geometry of the coastline aids in trapping surface water within the northern portion of the bay, creating an ‘upwelling trap’ (Castilla *et al.*, 2002b) where the surface waters are retained for several days. This has been suggested to represent a key retention mechanism for the planktonic larvae of *P. praeputialis*, which remain pelagic for less than three hours (Clarke *et al.*, 1999), and may be important in explaining the species’ limited distribution within Chile (Castilla *et al.*, 2002b, 2007).

### 3.3.2 Sampling

A total of 190 individuals of *P. praeputialis* were collected between 2009 and 2018 during low tides along the coastline of Antofagasta Bay and from several locations through the species’ Australian range (Table 3.1). Specimens were dissected *in situ*, and a piece of siphon and surrounding mantle tissue of each individual was placed in a 1.5 ml Eppendorf tube containing absolute ethanol. The ethanol was replaced several times until the tissues had turned white, indicating that any pigments had been removed. Samples were then shipped and stored in a freezer at -80°C at the National Oceanography Centre Southampton, United Kingdom, until further processing.
### Table 3.1. Sampling sites where *Pyura praeputialis* were collected, including information on geographical region, site abbreviation, geographic details, number of individuals sequenced (N) and used for genomic analyses (G).

<table>
<thead>
<tr>
<th>Region</th>
<th>Site name</th>
<th>Abbreviation</th>
<th>Latitude (S)</th>
<th>Longitude</th>
<th>N(G)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile</td>
<td>La Rinconada</td>
<td>C1</td>
<td>23.471</td>
<td>-70.513</td>
<td>8(7)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Las Conchitas</td>
<td>C2</td>
<td>23.527</td>
<td>-70.533</td>
<td>14(12)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Club de Yate</td>
<td>C3</td>
<td>23.643</td>
<td>-70.402</td>
<td>9(8)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Antofagasta</td>
<td>C4</td>
<td>23.707</td>
<td>-70.431</td>
<td>11(9)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>El Way</td>
<td>C5</td>
<td>23.743</td>
<td>-70.446</td>
<td>17(13)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Bolfin</td>
<td>C6</td>
<td>23.830</td>
<td>-70.499</td>
<td>11(7)</td>
<td>2013</td>
</tr>
<tr>
<td>Eastern Australia</td>
<td>Sydney</td>
<td>A1</td>
<td>33.992</td>
<td>151.231</td>
<td>19(16)</td>
<td>2018</td>
</tr>
<tr>
<td></td>
<td>Botany Bay</td>
<td>A2</td>
<td>33.992</td>
<td>151.231</td>
<td>19(17)</td>
<td>2018</td>
</tr>
<tr>
<td></td>
<td>Eden</td>
<td>A3</td>
<td>37.067</td>
<td>149.913</td>
<td>19(19)</td>
<td>2009</td>
</tr>
<tr>
<td>South-eastern Australia</td>
<td>Kilcunda</td>
<td>A4</td>
<td>38.556</td>
<td>145.481</td>
<td>15(14)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>Cowes</td>
<td>A5</td>
<td>38.447</td>
<td>145.240</td>
<td>15(15)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>Portsea</td>
<td>A6</td>
<td>38.319</td>
<td>144.712</td>
<td>18(17)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>Marengo Bay</td>
<td>A7</td>
<td>38.778</td>
<td>143.667</td>
<td>15(13)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td><strong>190(167)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.3.3 DNA extraction and genotyping

DNA was extracted using the Qiagen DNeasy® Tissue Kit (Qiagen) according to the manufacturer’s protocol, and the quality and quantity of extracted DNA was assessed by gel electrophoresis and the QuantiFluor® dsDNA System (Promega). DNA extractions were then genotyped on an Illumina NovaSeq6000 sequencer at the University of Wisconsin Biotechnology Center using genotyping-by-sequencing (GBS; Elshire *et al.*, 2011).
3.3.4 Analysis of GBS data

The GBS reads were assembled using ipyrad v. 0.7.30 (Eaton, 2014) with parameters recommended for paired-end GBS data in the ipyrad documentation (http://ipyrad.readthedocs.io/). As no published *P. praeputialis* genome is available, we used the *de novo* assembly method, which requires no prior genomic resources. The cutadapt algorithm v.1.12 (Martin, 2011) in ipyrad was used to trim Illumina adapter reads. We set the level of sequence similarity for clustering to be 90%, after comparing the number of recovered SNPs from datasets using values between 85 – 95% (see Table 7.9). We used vcftools v.0.1.13 (Danecek *et al.*, 2011) to first filter for loci with a minimum single nucleotide polymorphism (SNP) call quality of 20, a minimum genotype depth of less than five, and a mean minimum depth (across individuals) of less than 15. Additionally, we removed loci with a minor allele count of less than three to reduce the possibility of removing true rare alleles that are important in elucidating fine-scale structure and accurately drawing inference of past demographic events (O’Connor *et al.*, 2015).

We then followed the filtering framework outlined by O’Leary (2018), and iteratively increased our stringency for missing data allowance (on both loci and individuals separately), so that our final dataset contained loci with at least 50% call rate (i.e. a locus must be present in at least 50% of individuals), and up to 50% allowed missing data per individual. Finally, to remove the confounding effects of linkage disequilibrium, we used vcftools to also thin markers so that only one SNP per locus was retained in our dataset.

We used pcadapt v.4.1.0 (Luu *et al.*, 2017) to identify putative outlier loci using a false discovery rate of 5%. Pcadapt determines population structure using principal component analysis (PCA) to find candidate loci excessively related to population structure. However, as outliers identified by pcadapt are detected due to population structure, these outliers may have diverged through neutral process such as drift, and therefore may not exclusively identify loci that are related to local adaptation. Therefore, to assess for a relationship between sea surface temperature (SST) and genetic differentiation, we used bayenv2 which performs regression between environmental variables and allele frequencies (Günther & Coop, 2013). For each sample site, we obtained daily satellite GHRSST (A Group for High Resolution Sea Surface Temperature) sea surface temperature (SST) data for the years 2003 to 2017, at a 0.05° x 0.05° (~5.5 km) spatial resolution [JPL MUR MEaSUREs Project (2015)]. Data were downloaded from the Jet Propulsion Laboratory’s Physical Oceanography Distributed Active Archive Centre’s website (https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR). We then calculated the maximum and minimum yearly temperatures for each site, and these variables were used during the bayenv2 runs. Bayenv2 creates a matrix of covariance in allele frequencies between populations using putatively neutral loci. This covariance matrix is then used as a null model, where bayenv2
assesses for correlation between environmental variables and allele frequencies by comparing Bayes factors. Due to potential issues of instability across independent runs of bayenv2 in making the covariance matrix (Blair et al., 2014), we took the mean values of 10 independent runs, each run consisting of 100,000 MCMC iterations. To assess whether the mean covariance matrix represented the true variance of allele frequencies, it was compared to a random subset of putative neutral loci using a Mantel test in R with 1,000 permutations. We then assessed the correlation between individual SNPs and the two environmental variables. This approach was repeated ten times for each combination of SNP-environmental variable tested for association to account for instability between independent runs (Blair et al., 2014), and the mean of the ten runs was then used to infer the final Bayes factors (BF). The Kass and Raftery (1995) criterion was then used to determine the probability of these SNPs under selection, and to control for false positives, $2\ln K$ values above six (BF > 20) were used to identify outlier loci. Loci identified by either method produced two datasets (“pcadapt dataset” and “bayenv dataset”), and these loci were removed from our original dataset to make a “neutral dataset”.

### 3.3.5 Population structure

Using individuals collected from all sites, we used the software ADMIXTURE v.1.3 (Alexander et al., 2009) to estimate the likelihood that an individual comes from one of a pre-defined number of putative sample populations ($K$). ADMIXTURE uses a maximum likelihood estimation from multilocus SNP genotype datasets, and calculates estimates using a faster numerical optimisation algorithm than the more commonly used programme STRUCTURE (Pritchard et al., 2000).

Using the neutral, pcadapt and bayenv2 datasets, we then performed a discriminant analysis of principal components (DAPC) to visualise between-population genomic variation (Jombart et al., 2010). DAPC transforms the data with a PCA before using PCA factors as variables to perform a discriminant analysis (DA), ultimately maximising the differences among groups while minimising variation within groups (Jombart et al., 2010). We used the package adegenet v.2.1.1 (Jombart, 2008) for R to perform the DAPC. We ran the DAPC with and without a priori knowledge of individual populations. Pairwise population genetic differentiation was examined by calculating $F_{ST}$ values following Weir & Cockerham (1984) in the R package hierfstat (Goudet, 2005), and assessing their significance by running 10,000 permutations after correcting for multiple comparisons using the Bonferroni correction. The descriptive population genomic statistics observed heterozygosity ($H_o$), expected heterozygosity ($H_e$) and inbreeding coefficient ($F_{is}$) were also calculated using the R package hierfstat (Goudet, 2005), with significance in $F_{is}$ values being inferred using bootstrapping over 10,000 permutations. We used Arlequin v.3.5 (Excoffier & Lischer, 2010) to perform an analysis of molecular variation (AMOVA) based on the geographical
groupings of the different sampling sites. We assessed for isolation by geographic distance for
samples within Antofagasta Bay by performing a Mantel test (Mantel, 1967) using the R package
dartR v.1.1.11 (Gruber et al., 2018). The package converts latitude and longitude coordinates into
Google Earth Mercator projections to create a Euclidean distance matrix, creates a pairwise $F_{ST}$
genetic matrix between samples, and then performs an isolation by distance analysis based on
log(Euclidean distance) against between population pairwise $F_{ST}/1-F_{ST}$ (Rousset, 1997).

3.3.6 Reconstructing invasion routes

To further understand the population history of *P. praeputialis*, Approximate Bayesian
Computation (ABC) analyses were conducted using DIYABC (Cornuet et al., 2014). Our first goal
was to infer the colonisation history of *P. praeputialis* from Australia to South America. For this,
we initially grouped sites based on their geographical location and the results of the above
population structure analyses. We attempted to determine whether Antofagasta Bay had been
settled by individuals from eastern Australia, south-eastern Australia, an admixture of both, or
whether Antofagasta Bay was ancestral itself (Figure 7.7). As specific population sizes and
potential admixture rates were unknown, we used a uniform distribution with a large interval
(population sizes: $10^{-6}$; admixture rates: $0.001 - 0.999$) when setting priors for these
parameters. As it was expected that *P. praeputialis* was only introduced to Antofagasta Bay in the
past ca. 150 years (Castilla et al., 2002a), we set the divergence time between the Chilean
population and Australian populations at $t = 10 - 400$, whilst we set the divergence time between
the two Australian populations at $t = 10 - 5 \times 10^5$, due to divergence estimations of ca. $1.5 \times 10^5$
years (95% CI = $7.9 \times 10^4 - 4.2 \times 10^5$ YA; Teske et al., 2011) As summary statistics, we used the
mean genic diversity, mean distribution of $F_{ST}$ values, mean distribution of Nei distances (Nei,
1973), and whenever an admixture event was included in the scenario, mean admixture
estimates. For all scenarios, we used the default $10^6$ simulated data per scenario to build
reference tables. Upon creation of the reference table, we pre-evaluated scenarios and prior
distributions by performing a PCA in the space of the summary statistics on 1,000 simulated
datasets for each scenario, adding the observed dataset to each plane (Cornuet et al., 2014). We
used a logistic regression on the 1% simulated datasets that were closest to the observed dataset
(using Euclidean distances between simulated and observed datasets) to calculate the posterior
probability of each scenario. This approach produces 95% confidence intervals for each scenario’s
posterior probability, with the most likely scenario defined as the highest estimate without
overlapping confidence intervals (Cornuet et al., 2008).

We then took the most likely scenario (Figure 7.7, see Results) and attempted to identify the
route of invasion from within eastern Australia to Antofagasta Bay (Figure 7.8), using the same
Chapter 3

priors and summary statistics as outlined previously. For the most probable scenario (scenario B, Figure 7.8) we calculated type I and type II error rates. Finally, we assessed the goodness-of-fit for the final chosen scenario by implementing the model-checking feature of DIYABC. One thousand datasets were simulated using posterior distribution values and compared with the observed dataset by considering different summary statistics than those used during the generation of the reference table, and simulated datasets were visualised using a PCA (Cornuet et al., 2014).

The second aspect that we studied assessed relative support for six scenarios concerning the number of individuals that founded the population in Antofagasta Bay, assuming that these originated from eastern Australia (see results). The following effective population sizes were investigated: 1 - 10, 10 - 100, 100 - 1,000, 1,000 - 10,000 and 10,000 - 100,000 individuals. Priors for the eastern Australian source population encompassed $10^6$ - $10^7$ individuals, and the time of colonisation ranged from 10 to 400 years [based on the founding of Antofagasta city in 1868]. Chosen summary statistics included mean genic diversities, $F_{ST}$, and Nei’s distance (Nei, 1973). Results were based on 5 x $10^6$ simulations. Posterior probabilities and their 95% confidence intervals for each scenario were calculated using a logistical approach based on a sub-set of 50,000 simulations.

3.3.7 Environmental matching

In order to ascertain the potential role of environmental matching between the native and introduced ranges on the success of *P. praeputialis*, we obtained daily satellite GHRSSST (the Group for High Resolution Sea Surface Temperature) sea surface temperature (SST) data for the study sites for the years 2003 to 2017, at a 0.05° x 0.05° (~5.5 km) spatial resolution [JPL MUR MEaSUREs Project (2015)]. Data were downloaded from the Jet Propulsion Laboratory’s Physical Oceanography Distributed Active Archive Centre’s website (https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR). To assess whether temperature can explain why the species has not successfully colonised outside of Antofagasta Bay, we also obtained daily SST data for seven additional sites (C7-C15, Table 7.10) along the coast of Chile where *P. praeputialis* is not present. These sites were historical ports where international shipping occurred during the onset of the nitrate trade between Chile and Australia (Arce, 1997), and therefore there is a probability that *P. praeputialis* may have been transported to these sites via anthropogenic transport. From these data we calculated monthly and yearly average SST temperatures time-series for the sites of interest, as well as extreme values. All oceanographic analyses were performed in R v.3.3.1 (R Core Team, 2016).
3.4 Results

3.4.1 Processing of raw GBS data

A total of 265,536,772 raw reads were generated, with an average of 1,412,430 reads per sample. After filtering and clustering using ipyrad and vcftools, we retained a total of 4,216 putatively unlinked SNPs in the sequence assembly. Eighteen individuals were removed from the dataset due to missing data (i.e. greater than 50% missing data), which was likely caused by poor DNA quality or secondary contaminants within the samples (Federman et al., 2018). This led to a final dataset of 167 individuals from 13 sampling sites (Table 3.1). Pcadapt identified a total of 74 $F_{ST}$ outlier loci, and 101 loci were recovered from bayenv2. In total, 151 loci were identified as outliers by either of the outlier loci analyses, and subsequently removed producing a dataset of 4,064 SNPs putatively neutral loci (known as the “neutral dataset”). In addition, the 74 and 101 loci identified by pcdadapt and bayenv2 were retained in two separate datasets known as “pcadapt dataset” and “bayenv dataset” respectively.

3.4.2 Global genomic diversity indices

The value of global $H_0$ was 0.076 with values per population ranging from 0.066 (A3) to 0.082 (A6, Table 3.2). Values of $H_E$ were higher than $H_0$, with a global $H_E$ of 0.106 and per population values ranging from 0.104 (A3) to 0.109 (C5, Table 3.2). This dearth of heterozygotes corresponds to the average within-population $F_{IS}$ value of 0.286, and all populations exhibited significantly positive $F_{IS}$ values (Table 3.2). Allelic richness ranged from 1.102 (C1) to 1.107 (C5) with 126 SNPs (ca. 3%) being private.
Table 3.2. Descriptive population genomic statistics of all sampling sites. Site abbreviations as Table 3.1.

<table>
<thead>
<tr>
<th>Region</th>
<th>Site abbreviation</th>
<th>(H_o)</th>
<th>(H_e)</th>
<th>(F_{IS})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile</td>
<td>C1</td>
<td>0.078</td>
<td>0.105</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>0.075</td>
<td>0.108</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>0.073</td>
<td>0.107</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>0.078</td>
<td>0.106</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>0.074</td>
<td>0.109</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>C6</td>
<td>0.076</td>
<td>0.108</td>
<td>0.208</td>
</tr>
<tr>
<td>Eastern Australia</td>
<td>A1</td>
<td>0.076</td>
<td>0.105</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>0.078</td>
<td>0.108</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>0.066</td>
<td>0.104</td>
<td>0.274</td>
</tr>
<tr>
<td>South-eastern Australia</td>
<td>A4</td>
<td>0.073</td>
<td>0.107</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>A5</td>
<td>0.078</td>
<td>0.104</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>A6</td>
<td>0.082</td>
<td>0.105</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>A7</td>
<td>0.078</td>
<td>0.104</td>
<td>0.178</td>
</tr>
</tbody>
</table>

3.4.3 Evolutionary history of the native range

Using genomic tools, we confirmed the presence of two genetic lineages within *Pyura praeputialis* found in Australia, the eastern Australia lineage and south-eastern Australia lineage. DAPC and ADMIXTURE (Figures 3.1A and B) and \(F_{ST}\) (Figure 7.9) analyses all supported the result of two groupings of populations within Australia when analysed using the neutral dataset. The composition plot produced by adegenet showed 100% group assignment probability to the two lineages. A similar result was also observed using the outlier loci datasets (Figures 7.1A and B).
Figure 3.1. Genomic differences between south-eastern Australian and Chilean sites inferred using putatively neutral loci: (A) Discriminant Analysis of Principal Components without population information as priors. (B) Admixture plot depicting the inferred ancestry of each *P. praeputialis* individual to pre-defined genetic clusters for $K = 2$. Site abbreviations as Table 3.1.

### 3.4.4 Population structure across all sample sites

The cross-validation procedure of the ADMIXTURE analysis recovered the best-supported number of ancestral populations ($K$) to be 2 for the neutral and bayenv dataset (Figures 7.11A and B), and 3 for the pcadapt dataset (Figure 7.11C). This clustered the populations into two clear groups (one group comprising Antofagasta Bay and eastern Australia sites, and the second comprising south-eastern Australia sites; Figures 3.1 and 7.12). The third cluster in the pcadapt dataset was only apparent in two individuals from A5, and this result still strongly supported the strong structuring observed between sample regions (Figure 7.12C).
Similarly, DAPC analyses including population identity as prior information recovered the same two genetic clusters when visualising between-population genomic variation for all datasets (Figures 7.13A, B, and C). When south-eastern Australian populations were removed from the analysis, the Antofagasta Bay and eastern Australian population still clustered together tightly, with site A3 (Eden) appearing slightly differentiated in the neutral and pcadapt dataset (Figures 7.13D and F).

Pairwise comparisons of $F_{ST}$ also suggested very strong genetic structuring between south-eastern Australian sites and all other sites, with a total of 37 out of 78 comparisons (47%) being significant (Figure 7.9), confirming that south-eastern Australian sites are highly unlikely to have played a role in the colonisation of Antofagasta Bay. Notably, there was no significant genetic differentiation between any sites within a given geographical area (i.e. Antofagasta Bay, eastern Australia, south-eastern Australia) with the exception of two south-eastern Australian sites (A5 v A7). Despite this, the AMOVA test using pre-defined geographical site clusters suggested significant genetic differentiation among groups, among sites within groups, and within sites (Table 7.11).

### 3.4.5 Reconstructing invasion routes

Pre-evaluation of priors used in the ABC analyses showed that our scenarios matched the observed data appropriately to continue the analyses (Figures 7.14 and 7.15). Our simulations indicated that the number of individuals introduced to Chile from eastern Australia to retain the considerable genetic diversity observed must have been in the order of thousands (Figure 7.16). We found that the Antofagasta population was sourced by individuals from eastern Australia (Probability $P=1.000$, CI=1.000, 1.000; Table 7.12; Figure 3.2), rather than south-eastern Australia or an admixture between the two (Table 7.12). The next set of scenarios found that the Antofagasta population had been sourced by Eden (Probability $P=0.8538$, CI= 0.7880, 0.9195; Figure 3.2) rather than individuals from Sydney or an admixture between the two (Table 7.13).
Figure 3.2. The most likely colonisation scenarios depicting the putative introduction route of *P. praeputialis* from Australia to Antofagasta Bay, as revealed in our analyses using the Approximate Bayesian Computation method. (A) Initial colonisation from the two Australian lineages to Chile. (B) The specific colonisation from sites within east Australia and Chile. The grouping “Chile” contains all sample sites from within Antofagasta. Site abbreviations in (B) as Table 3.1.

### 3.4.6 Environmental matching

The satellite-derived SST data showed that both the Australian and Chilean coastlines exhibited an increase in average SST with decreasing latitude (Figure 3.3). Finer-scale resolution SST within Antofagasta Bay however confirmed that the northern portion of the bay remained on average slightly warmer than a northerly neighbouring site (C12, Figure 3.3). However, between 2003 and 2017, the average SST for the whole of Antofagasta Bay closely followed that of C12 and was very similar to that of the Australian site A3 (Figures 3.3 and Figure 7.17). Interestingly, the average annual SST of A1 (20.35 ± 0.45°C) was higher than all sites in Chile, whereas the average annual SST in south-eastern Australia (15.97 ± 0.19°C) was lower than all sites in Chile (Figure 3.3).
Figure 3.3. Average sea surface temperature (SST; back dots), maximum and minimum SST (blue crosses), and range between the maximum and minimum SST obtained from the satellite derived GHRSSST dataset (A Group for High Resolution Sea Surface Temperature; JPL MUR MEaSUREs Project, 2015) over a fifteen-year period (2003-2017) at the sample sites along the southeast coast of (A) Australia and historical ports along the coast of (B) Chile. Note, C13 and C14 are sites in the Antofagasta Bay, where the Chilean samples analysed in this study come from. (C) Photographs of *P. praeputialis* in the field from (i) Wollongong, south-eastern Australia, and dense clusters within (ii) La Rinconada, Antofagasta Bay (C1, see site details in Table 3.1). Photos courtesy of Marc Rius.
When comparing daily SST between C1 (i.e. La Rinconada, see Table 1, the site furthest north within Antofagasta Bay) and C6 (the site slightly out of Antofagasta Bay), we found that the temperature was warmer within the bay ~75% of the time between 2003 - 2017, with the maximum difference in temperature between these two sites being 2.48°C on 09 December 2012 (Figure 7.18). This difference did not however reach the range observed by a previous study using in situ temperature data loggers (Piñones et al., 2007). Interestingly, SST experienced at historic ports along the Chilean coastline where *P. praeputialis* is currently not found is within the SST range of the native range (Figure 3.3). Furthermore, extreme temperature events (i.e. maximum and minimum SST) throughout the native range also encompass a wider temperature window than those across the Chilean coastline.

### 3.4.7 Fine-scale population genomic analysis within the introduced range

We found no pattern of isolation by distance within Antofagasta Bay ($r = 0.2399, P = 0.190$). Performing a DAPC using only sites within Antofagasta Bay recovered an interesting pattern when sample sites were used as priors (Figure 3.4) on the neutral dataset. The two sites closest to the mouth of the bay (C2 and C6, see Table 3.1) cluster closely, as do sites found in the middle of the Bay (C4 and C5, see Table 3.1).
Figure 3.4. Discriminant Analysis of Principal Components using unlinked loci with population information provided as a prior and sites from Australia excluded. The inset map represents sea surface temperature (SST) within Antofagasta Bay on the 9th of December 2012. This day corresponds to the largest range in SST within the bay between the years 2003 and 2017. The warmer water found in the northern part of the bay are evident. Site abbreviations as Table 3.1.

When performing DAPC on the outlier datasets for samples within Antofagasta (Figures 7.13H and I), we did not find any obvious pattern between genomic structuring and the unique SST patterns observed within Antofagasta (Figure 3.4), suggesting that our datasets were not identifying evidence of local adaption to temperature within Antofagasta Bay.

3.5 Discussion

Non-indigenous species with restricted introduced ranges provide a unique opportunity to understand disparities between realised and potential ranges. In this study we found high genomic diversity in the introduced range, suggesting high adaptive potential. In addition, we checked environmental matching between native and introduced ranges and found that invasive species with restricted introduced ranges can have great potential to spread and enlarge their ranges. Our results urge researchers and managers to assess environmental similarity of important abiotic conditions across the species range, as well as population genomic attributes, to
fully understand the potential for spread of these species. Understanding why a species has not spread further from its introduced region can aid unravelling the mechanisms ruling species invasions. We argue that recognising the potential range is essential to forecast and mitigate invasion risk.

The introduction of *P. praeputialis* to a single bay in Chile provides a rare natural experiment in invasion biology. Our results suggest that environmental matching between native and introduced ranges could explain a more extensive introduced range of *P. praeputialis*. Despite *P. praeputialis* presumably being transported to other sites along the Chilean coastline (i.e. failed introductions), we suggest that peculiar conditions characterising Antofagasta Bay (Figure 3.4; Castilla *et al.*, 2002b, 2004; Piñones *et al.*, 2007), coupled with the biology of *P. praeputialis* (Clarke *et al.*, 1999; Castilla *et al.*, 2007), facilitated the retention of *P. praeputialis* from the eastern Australian lineage inside the bay.

### 3.5.1 Evolutionary history within Australia

Previous research using multiple genetic markers (i.e. COI, 18S, ATPSα, and ANT loci) have reported the occurrence of two lineages of *P. praeputialis* within its native range of Australia (Teske *et al.*, 2011; Rius & Teske, 2013). We confirmed, using high throughput genomic data, the presence of these two lineages without any further cryptic population structure. The curious presence of two lineages within species’ native ranges has been identified in other study systems (Rius *et al.*, 2014a) and has been posited to be a result of intra-species physiological differences or limited connectivity across the native range. Although we found no evidence for regional population structure within each lineage, one pairwise population comparison had a significant $F_{ST}$ value [Cowes vs Marengo Bay; one of the same pairwise values that were significant in a previous study by Teske (2014)]. Together, our results suggest limited gene flow between the two lineages separated by a biogeographic break, as reported in other marine invertebrates (Waters, 2008; Ayre *et al.*, 2009).

### 3.5.2 Invasion history

The first introduction of *P. praeputialis* to Chile would most likely have occurred at a time when maritime traffic was not only considerable between Australia and Chile, but when fouling organisms were rarely removed from ships (Carlton, 1999). The most likely introduction would have been during the late 19th century, when ships had wooden hulls to which this ascidian could readily attach. We found little genetic structure between Chile and eastern Australia (Figure 3.1), providing further evidence that this is the source lineage into Antofagasta (Castilla *et al.*, 2002a;
Teske et al., 2011), and we found no evidence of input from southern Australia populations to Chile. Until now, no study has shown evidence for a single lineage introduction of P. praeputialis from Australia to Chile. Indeed, the south-east Australian lineage of P. praeputialis too inhabits in a region containing busy ports, and one would expect that this region would have been linked to South America through historic shipping. Our oceanographic data suggest that the maximum SST of recipient locations may inhibit the establishment of this south-eastern lineage of P. praeputialis in Chile, including sites within Antofagasta bay. The highest SST observed in south-eastern Australia between 2003-2017 was 22.69°C, whereas each sampled (for temperature) site along the Chilean coastline experienced higher temperatures (Figure 3.3; the lowest value was 23.96°C at site C15). Previous studies have shown that lineages inhabiting different biogeographic zones are often locally adapted to regional conditions and movement to areas of differing environments can disrupt growth, reproduction, development, and survival (Teske et al., 2007, 2008; Papadopoulos & Teske, 2014). Additionally, increased temperature has been shown to promote disease in marine environments (Bruno et al., 2007; Sato et al., 2009), and therefore even if the high temperatures of Chile do not directly affect the ecophysiology of the south-eastern Australia lineage, it may promote mortality and preclude successful establishment indirectly. In a similar vein, the higher temperature experienced by the eastern Australia lineage may have aided the introduction to Antofagasta through environmental matching.

Previous research has shown that invasive species have a wider tolerance to environmental stresses (Lenz et al., 2011) and tend to come from habitats with higher maximum temperatures (Zerebecki & Sorte, 2011). To further study this, experiments could be performed to assess the thermal tolerance of P. praeputialis individuals sampled throughout its range. The effect of temperature on early life history development has been researched recently using the sister species P. herdmani and P. stolonifera within southern Africa (Hudson et al., in press), and a similar approach could empirically assess the role abiotic factors have played in both the successful and failed introduction of P. praeputialis. It is interesting to note very little change in expected heterozygosity between Chile and eastern Australia. It has been suggested that P. praeputialis is a relatively recent introduction (~150 years ago; Castilla et al., 2002a), and the observed high genetic diversity suggests evidence against a genetic bottleneck. Surprisingly, private alleles were found in all population groupings (eastern Australia, south-eastern Australia, and Chile). Normally, for a recent introduction, one would expect many shared alleles and few private alleles within recipient populations. The vast abundance of P. praeputialis and high genetic diversity may lead to a high number of private alleles, due to the positive relationship between population size and genetic diversity (Hague & Routman, 2016). However, the most likely reason for the presence of a high proportion of private alleles in the introduced range is limited sampling
within eastern Australia. Previous studies have highlighted the importance of sampling as much of the native range as possible (Geller et al., 2010; Viard et al., 2016), and an unsampled source population in Australia may be the cause of the private Chilean alleles in our study.

Reconstructing the introduction route using the ABC method confirmed that the Antofagasta population originated from individuals from east Australia whilst somewhat surprisingly suggested the smaller port, Eden (A3, Table 1), rather than Sydney (A1) was the source. The average yearly temperature of Eden (A3) closely follows that of Antofagasta (Figure 7.17), so perhaps a nuanced effect of temperature has promoted introduction from one port to the other.

3.5.3 Patterns within the introduced range

Our satellite-derived SST data from a range of sites along the coast of Chile and Australia suggests temperature cannot solely explain the limited distribution of P. praeputialis along the coast of Chile. This observation is consistent when taking into account both mean values and extreme values of SST (Figure 3.3). If thermal niche matching was the sole explanation for the ability for P. praeputialis to be introduced to Antofagasta, it would be assumed to be able to survive in all other sites north of Antofagasta too, as the SST of these sites sit between that average SST of Eden and Sydney in Australia. Whilst our high-resolution temperature data are derived from satellite offering data a wide geographical range, it should be noted that in some regions, in situ measurements of SST have been shown to be inconsistent with satellite-derived data (Smit et al., 2013). Indeed, our measurements of SST within Antofagasta Bay are slightly different to those previously reported (Figure 7.18), but the general patterns observed remain the same. Studying failed introductions is inherently difficult as there is often no footprint left behind of the introduction event. It is known that Chile and Australia were well-connected via shipping in the 19th century (Maino, 1985; Arce, 1997), and it would be unreasonable to assume the species was only attached on ships between Chile and Antofagasta. Van Name (1945) described a species of the family Pyuridae from Peru and Ecuador as Pyura bradleyi. Interestingly, the description of this species includes a double spiral cone dorsal tubercle, and a similar distribution of the gonads, both useful morphological characters to identify P. praeputialis (Rius & Teske, 2011). However, the description in Van Name (1945) can only be seen as possible evidence of the introduction of P. praeputialis to additional South American sites, with more taxonomic or genetic work needed to clarify differences between P. praeputialis and P. bradleyi. Interestingly, despite no evidence of naturalisation or spread of P. praeputialis to sites outside of Antofagasta Bay, work conducted by Castilla et al., (2004) has shown that juvenile P. praeputialis can be transplanted outside of Antofagasta, and continue to survive and grow. Additionally, the authors conclude that there appears to be no biotic rationale behind the lack of expansion from the bay, such as predator
intensification (Castilla et al., 2004). This is in contrast to another highly localised non-indigenous species, the cyprinid *Rhinichthys osculus*. This fish is limited in its distribution within the introduced range due to predation (Harvey et al., 2004), a possible result of a genetic bottleneck (Kinziger et al., 2011). Successful transplantation of species outside of their ranges, overcoming niche constraints, is not rare (Hargreaves et al., 2014) and show that dispersal is an important explanatory variable of range limits. Therefore, it may be possible that Allee effects (Stephens et al., 1999), the positive relationship between mean fitness and population density, additionally limit the ability for *P. praeputialis* to naturally disperse from Antofagasta. Within Antofagasta, our genomic data suggested reduced structuring, with sites closest to the mouth of the bay (C2 and C6, Table 1) seemingly genetically separated in our DAPC analysis (Figure 4). The warmest site within Antofagasta (C1) is also the region with the largest individuals and highest biomass, whilst the site furthest to the mouth of the bay (C6) has the lowest population density (MR and JCC personal observations). In addition, larvae of *P. praeputialis* can be retained by bio-foam produced during reproduction in Antofagasta Bay (Castilla et al., 2007), which combined with the short larval duration in the species promotes retention from the bay. Taken together, we suggest the introduction of the lineage is only possible if the biotic (i.e. predation, competition), abiotic (upwelling trap, sub-surface water bringing in water rich in chlorophyll, bay retention mechanisms etc), and shipping routes are just right (as in the case of Antofagasta Bay).

### 3.5.4 Conclusions and future directions

To our knowledge, no other marine invasive species with restricted introduced range has been so extensively investigated to understand the realised and potential size of the introduced range. By using a combination of genomic and oceanographic data we showed that the unique characteristics of Antofagasta bay has precluded the expected expansion of this species. If conditions change in the future though (e.g. wind patterns, currents), this invasive species may be able to expand to other areas. In addition, whilst temperature may explain failed establishment of one native lineage (south-east Australia), our study suggests a great potential to spread of this highly invasive species in its introduced range.

### 3.6 Acknowledgements

We thank Dr Mark de Bruyn, Rhiarn Hoban and Dr Katie Dafforn for their help providing invaluable samples from Australia. MR received funds from ASSEMBLE (an EU FP7 research infrastructure initiative comprising a network of marine research stations) to visit the Estación Costera de Investigaciones Marinas de Las Cruces and the Pontificia Universidad Católica de Chile. PRT was supported by the University of Johannesburg (FRC/URC grant). JCC acknowledges
financial support from the Nucleo Milenio en Conservación Marina. JCC sincerely acknowledges long-term financial support (PUC- Escondida Project) from Minera Escondida Ltda. and the important assistance from Dr Patricio Manríquez. CDM received funding from the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation.
Chapter 4 Contemporary climate change hinders hybrid performance of ecologically dominant marine invertebrates

This chapter is a reproduction of text published with the *Journal of Evolutionary Biology* in the special issue titled “Speciation in Marine Environments”. As such, this chapter is written in the style of the journal.

4.1 Abstract

Human activities alter patterns of biodiversity, particularly through species extinctions and range shifts. Two of these activities are human mediated transfer of species and contemporary climate change, and both allow previously isolated genotypes to come into contact and hybridize, potentially altering speciation rates. Hybrids have been shown to survive environmental conditions not tolerated by either parent, suggesting that, under some circumstances, hybrids may be able to expand their ranges and perform well under rapidly changing conditions. However, studies assessing how hybridization influences contemporary range shifts are scarce. We performed crosses on *Pyura herdmani* and *Pyura stolonifera* (Chordata, Tunicata), two closely related marine invertebrate species that are ecologically dominant and can hybridize. These sister species live in sympatry along the coasts of southern Africa, but one has a disjunct distribution that includes northern hemisphere sites. We experimentally assessed the performance of hybrid and parental crosses using different temperature regimes, including temperatures predicted under future climate change scenarios. We found that hybrids showed lower performance than parental crosses at the experimental temperatures, suggesting that hybrids are unlikely to expand their ranges to new environments. In turn, we found that the more widespread species performed better at a wide array of temperatures, indicating that this parental species may cope better with future conditions. This study illustrates how offspring fitness may provide key insights to predict range expansions and how contemporary climate change may mediate both the ability of hybrids to expand their ranges and the occurrence of speciation as a result of hybridization.
4.2 Introduction

Changes in gene flow between genetically distinct populations or species are known to alter speciation rates. On the one hand, increased gene flow can accelerate speciation through reinforcement (Hoskin et al., 2005; Abbott et al., 2013) or lead to the formation of novel genetic entities via hybrid speciation (Mallet, 2007). On the other hand, hybridization may slow speciation by allowing gene flow among diverging populations (Abbott et al., 2013). Consequently, hybridization can have a variety of effects on populations experiencing increased gene flow. Other effects of hybridization include an increase in genetic variation within populations, creating a larger pool of genotypes on which natural selection can act (Hegarty, 2012), or the purge of deleterious recessive alleles that have accumulated in parental populations (Keller & Waller, 2002). Hybridization can also be detrimental to parental populations, either directly, by leading to the extirpation of one or both parental populations via introgression (Rhymer & Simberloff, 1996; Arcella et al., 2014; Muhlfeld et al., 2014), or indirectly, by providing first-generation hybrids with phenotypic superiority over their parents (i.e., heterosis). Hybridization may also lead to the genesis of new phenotypes as a result of transgressive segregation, offering later generations a selective advantage (Lexer et al., 2003). For example, studies on cichlid fish, sunflowers, copepods and water fleas have shown the generation of hybrid phenotypes with traits that are extreme compared to the parental phenotypes (Rieseberg et al., 2003; Stelkens et al., 2009; Pritchard et al., 2013; Griebel et al., 2015). Finally, hybridization can erode accumulated genetic differentiation when reproductive barriers are eventually removed (Taylor et al., 2006). Taken together, this evidence indicates that hybridization can have several genetic and phenotypic effects with outcomes that are difficult to predict, especially at a time when contemporary climate change (CCC) is extensively reshaping species distributions and abiotic conditions (Potts et al., 2014).

It is well established that range shifts allow previously isolated populations to come into contact and hybridize. Natural range expansions may arise due to changes in both biotic and abiotic factors (e.g., sudden removal of predators or gradual changes in temperatures) occurring over a wide range of temporal and/or spatial scales (Sexton et al., 2009). For example, historic changes in climate led to the expansion of the European common frog, *Rana temporaria*, due to glacial retreat across Europe ca. 20,000 years ago (Vences et al., 2013). Another example of natural range expansions occurs when two habitats that were previously isolated because of a physical barrier come into contact, such as the formation of the Isthmus of Panama ca. $3 \times 10^6$ years ago, which led to the Great American Interchange (Marshall, 1988). Similarly, CCC greatly affects species distributions and often leads to range shifts. Indeed, warming winter temperatures associated with CCC has promoted a range expansion in the butterfly *Atalopedes campestris*.
across western North America (Crozier, 2004). In addition to major changes in abiotic factors, species interactions such as predation (Huang et al., 2012) and competition (van der Knaap et al., 2005) can also shape the nature of range shifts (Svenning et al., 2014). Range expansion of the Humboldt squid, *Dosidicus gigas*, in the eastern Pacific has been attributed in part to the removal of competing top predators, which prey on juvenile *D. gigas* (Zeidberg & Robison, 2007). Other types of range shifts are those associated with human activities, such as anthropogenic transport of species that has resulted in unprecedented increases in the speed and magnitude of species translocations (Carlton & Ruiz, 2015; Seebens et al., 2016). Anthropogenic transport is particularly problematic in the marine environment, where shipping provides an unparalleled transoceanic vector for the translocation of organisms from their native ranges to new locations, allowing the transport of large numbers of propagules to distant regions (Carlton & Geller, 1993). Range expansions as a result of anthropogenic transport may be exacerbated by CCC (Occipinti-Ambrogi, 2007; Hellmann et al., 2008; Sorte et al., 2010; Rius et al., 2014a; Canning-Clode & Carlton, 2017; Chan et al., 2019), with rising temperatures enhancing settlement success (Raitsos et al., 2010) or causing phenological shifts that favour range shifting species (Stachowicz et al., 2002; Wolkovich & Cleland, 2011; Chefaoui et al., 2019).

As CCC and human-mediated transport of species bring previously isolated genotypes into contact, unprecedented levels of hybridization have been reported in recent times (Vallejo-Marín & Hiscock, 2016; Canestrelli et al., 2017; Makino et al., 2018). As some hybrids have been reported to survive conditions that the parents cannot, hybridization is often associated with recent range expansions (Hegarty, 2012; Hovick et al., 2012; Rius & Darling, 2014). Consequently, complex interactions among CCC, hybridization, anthropogenic transport of species and natural range shifts are potentially shaping speciation patterns. A clear example of this is the release of *Ambystoma tigrinum* (the barred tiger salamander), which was intentionally released from southern to western regions of North America (Riley et al., 2003), where it has now become invasive. Part of the range of *A. tigrinum* now overlaps with that of the native *Ambystoma californiense* (the California tiger salamander), allowing interspecific hybridization to occur. Experimental crosses have shown that temperature has a positive effect on juvenile salamander dispersal distances, suggesting that CCC may facilitate a rapid range expansion of the hybrid swarm and hence the spread of non-native genotypes (Johnson et al., 2010).

Research to date has assessed how CCC affects the potential for hybridization, which can alter both pre-mating and post-mating reproductive barriers (Chunco, 2014). CCC has been shown to promote maladaptive hybrids, altering life-history traits in toads (*Bufo* spp.) (Canestrelli et al., 2017). Similarly, CCC can foster range shifts and increase the likelihood of hybridization in insects (Sánchez-Guillén et al., 2013). In line with this, studies have shown alteration of spatiotemporal
patterns of hybridization between the salmonids *Oncorhynchus clarkii lewisi* and *Oncorhynchus mykiss* (Muhlfeld et al., 2014). Despite all this research, little is known about how hybridization facilitates range expansions and how hybrids react to rapidly changing climatic conditions. Anthropogenic transport of species and CCC drive range shifts and species invasions, involving both hybrids or parental genotypes (Hegarty, 2012), potentially altering ecosystem structure and function across large geographic areas (Miehls et al., 2009; Katsanevakis et al., 2014). Thus, understanding how hybridization influences range expansions is an important aspect for predicting the effects of CCC on species distributions.

Here, we examined how conditions expected under CCC affect the performance of a range of ontogenetic stages of both hybrids and parental individuals of sympatric marine ascidian species, and how hybridization shapes the probability of future range shifts and speciation. Our objectives were to: (a) Quantify the performance of hybrid and parental crosses; (b) Assess differences in performance at different temperatures between hybrid and parental crosses; (c) Find links between offspring performance and species distributions. We hypothesized first that hybrids would perform similarly to parental species under conditions matching the area where their species distributions overlap, and second, hypothesized that under extreme temperatures, hybrids would show different performance than parental crosses. To achieve this, we tested the effects of different temperature regimes across multiple life history stages, as each ontogenetic stage has the potential to act as a bottleneck for species persistence (Byrne, 2011). We predicted that later life history stages would be tolerant to a wider range of temperatures than earlier stages, in line with previous studies (Pineda et al., 2012).
4.3 Methods

4.3.1 Study species

![Image of distribution map and study species photos]

Figure 4.1. The distribution of two *Pyura herdmani* lineages and *Pyura stolonifera* along the coasts of southern Africa (A) based on Teske et al. (2011) and Rius and Teske (2011). The two sample sites used for this study were SRP (Shark Rock Pier) and SLH (Summerstrand Lighthouse) along the south coast of South Africa. B) *Pyura herdmani* has spread its range to include northwest Africa and southwest Europe (see details in main text). C) Three *P. herdmani* temperate lineage individuals collected in Port Elizabeth harbour (PEH). D) One *P. herdmani* subtropical lineage individual collected in Park Rynie (north east coast of South Africa) and housed in an aquarium at Rhodes University. E) Three *P. stolonifera* individuals collected at SRP in an aquarium at Rhodes University.

We selected two closely related bioengineer marine invertebrates [*Pyura herdmani* (Drasche, 1884) and *Pyura stolonifera* (Heller, 1878), Chordata, Tunicata, Asciidiacea] that coexist along extensive stretches of coastline in southern Africa (Rius et al., 2017). *Pyura herdmani* has also been described in north Africa (Monniot & Bitar, 1983; Lafargue & Wahl, 1986 - though identified in both references as *Pyura stolonifera*) and has recently been reported in southwestern Europe.
(X. Turon, personal communication). Pyura spp. are solitary ascidians that are broadcast spawners with a very short pelagic larval duration of <24 hr (Svane & Young, 1989). This, combined with its disjunct distribution, suggests that P. herdmani is either undergoing current, or has previously undergone historic, range expansion due to anthropogenic transport. In southern Africa, the two species inhabit the lower intertidal and subtidal zones, where they form dense aggregates (Rius & Teske, 2011). A phylogenetic study based on mitonuclear and nuclear loci revealed that P. herdmani consists of two presumably temperature-defined lineages in South Africa: a temperate lineage inhabiting the west and south coasts, and a tropical/subtropical lineage inhabiting the east coast (Teske et al., 2011). The temperate lineage of P. herdmani lives in sympatry with P. stolonifera (Figure 4.1), and thus, there is potential for hybridization in the field. Previous work has suggested that P. herdmani and P. stolonifera can hybridize in the laboratory (Rius & Teske, 2013), but no empirical data are available on how these hybrids perform compared to parental crosses or how they are affected by temperature.

4.3.2 Field sampling

Individuals of P. herdmani and P. stolonifera were collected from natural hard substrata at spring low tides between September and November 2017 from Shark Rock Pier (only P. stolonifera; 33°59′28″S, 25°40′37″E) and Summerstrand Lighthouse (both P. herdmani and P. stolonifera; 33°58′47″S, 25°39′29″E; Figure 4.1) on the south coast of South Africa. Care was taken not to damage the inner body of the ascidians during collection, and to remove any damaged epibionts from the tunic as dead tissue leads to bacterial infection and causes causalities among the collected individuals (Monniot, 1990). Sampled ascidians were placed inside insulated cooler boxes filled with seawater and returned to the laboratory as soon as possible (within approximately 2 hr).

4.3.3 Sea surface temperature data

We obtained daily sea surface temperature (SST) data for the study site from the JPL MUR MEaSUREs Project (2015) for the years between 2003 and 2017, at a 0.01 (latitude) × 0.01 (longitude) spatial resolution. From this, we calculated monthly and yearly average SST temperatures. All data extraction and analyses were performed in R version 3.3.1 (R Core Team, 2016).
4.3.4 Laboratory housing of animals

Individuals were maintained in 50 L aquaria in a constant temperature room, with a 12-hr light/dark cycle. The aquaria were oxygenated using air pumps and seawater was replaced every other day using water from either the Swartkops River estuary or from Kenton-on-Sea (33°41′1.71″S, 26°41′8.52″E). Each day, the ascidians were fed 200 ml of either Isochrysis galbana, Dunaliella primolecta or a mixture of the two algae. The ascidians were checked daily for signs of bacterial infection and any individual showing signs of infection (identified by either the presence of a white bacterial mat growing on the tunic or a reduced response of the siphons to gentle physical stimuli) were immediately removed.
4.3.5 Fertilization methods

Figure 4.2. Experimental design used to test the effects of temperature on inter/intraspecific crosses. A) Two individuals of species were used per cross. B) Surgical collection of sperm (♂) and eggs (♀). C) Eggs from each individual were outcrossed with sperm from either one (control cross- same species) or two (hybrid cross- different species) individuals in 5ml of filtered seawater. The symbols above each dish represent the source individual for eggs (♀) and sperm (♂) in each dish. D) Fertilised eggs from each dish from C were washed and grouped into 500 ml beakers of filtered seawater and incubated at 20°C in darkness to hatch. E) Hatched larvae were pipetted into Petri dishes and designated a control environment room (either 12°C, 16°C, 20°C, 24°C, or 28°C) where they were assessed after 24, 72 and 120 hours. Twenty larvae were pipetted into each Petri dish, and the number of Petri dishes at each temperature for each cross is shown in Table 7.14.

Two individuals of each species were used for each cross (Figure 4.2). The tunics of all individuals were first removed, revealing the inner soft body. The dorsal tubercle was used to confirm the
identity of each species (Rius & Teske, 2011). The strip-spawning method of Marshall et al. (2000) was followed to dissect out gametes. The mixture of sperm and ova was poured through a 160 μm mesh filter so the ova were retained on the mesh but the sperm were allowed to wash through to another Petri dish. The ova were then washed off the mesh using 15 ml filtered seawater (FSW) into a final Petri dish. This process was then repeated for the remaining individuals. Filtered seawater was obtained using a vacuum pump to filter the seawater collected from the field through a 0.7 μm filter. In all crosses, sperm concentration was kept as high as possible to limit the effect of sperm ageing (Marshall et al., 2000) and the time between gamete extraction and gamete mixing was kept to a minimum. The ova of each individual were then aliquoted into two Petri dishes, with the ova in one dish receiving conspecific sperm and the ova in the second dish receiving interspecific sperm (Figure 4.2). A total volume of 1 ml of sperm (either from one individual for control crosses, or multiple individuals for hybrid crosses) was added to each Petri dish with ova. The sperm and eggs were kept at 20°C to allow fertilization to occur as this temperature has previously been shown to promote successful egg development in both parental species (Rius et al., 2014a). Once egg cleavage was observed, one more round of filtering ensured excess sperm would be washed away, reducing the probability of polyspermy. Fertilized eggs were then grouped by cross and kept in 500 ml of aerated FSW in total darkness at 20°C (Figure 4.2).

Once motile larvae had hatched (after c. 12 hr), 20 randomly selected larvae from each cross were transferred into one of five pre-roughened Petri dishes with 15 ml FSW (Figure 4.2). The Petri dishes had been left in unfiltered seawater for a few days to promote the formation of a biofilm, which is known to facilitate larval settlement (Wieczorek & Todd, 1997). Each of the five Petri dishes was then randomly allocated to one of five temperature-controlled rooms set at 12, 16, 20, 24 or 28°C (Figure 4.2). Due to different numbers of hatched larvae in the various crosses, the number of replicates for testing settlement success and post-metamorph performance at each temperature varied, ranging from four to eight (Table 7.14).

After 24, 72 and 120 hr, the larvae in each temperature-controlled room were examined under a microscope and the stage of development of each individual was noted. These stages of development comprised floating tadpole larvae, attached tadpole larvae, attached settlers, non-attached settlers, settlers with obvious tail reabsorption, pre-metamorphs and post-metamorphs (Table 7.14). As the duration of metamorphosis determines the length of time individuals are exposed to sources of mortality however (O’Connor et al., 2007), it was assumed that if individuals had not reached the post-metamorph stage within 120 hr in the laboratory, they would be unlikely to survive in the field due to the pressures of smothering and predation.
Therefore, the percentage of post-metamorphs at 120 hr was used as a proxy for species performance under the different temperature treatments.

4.3.6 Data analysis and statistics

Due to the proportional nature of our datasets, we analysed the data using generalized linear models (GLM) with a binomial error distribution and a logit link function. We tested for overdispersion in our data and where present we included a random factor (individual Petri dish) using the glmer package (Bates et al., 2015). In order to determine whether there were any interactive effects of temperature and cross on development or not, we first investigated the effect of cross under in situ temperatures (i.e., 20°C, Figure 7.20), and then, assessed pre-metamorph and post-metamorph performance at 120 hr. As performance values were zero at certain temperatures, we removed 28°C from the pre-metamorph analysis at 120 hr, and similarly removed 12 and 28°C from the post-metamorph analysis at 120 hr. Post hoc Tukey tests were used to determine the pairwise comparisons that drove significant differences. Repeated-measured analyses could not be performed as offspring performance through time was measured at the level of the Petri dish rather than the individual. All statistical analyses were performed in R (R Core Team, 2016).

4.4 Results

4.4.1 Temperature results

Sea surface temperature records indicate that the average SST for Port Elizabeth waters was 19.23°C ± 0.13°C (SE) between 2003 and 2017, with average summer and winter fluctuating around 22 and 17°C, respectively (Figure 7.20).
4.4.2 Development at in situ temperature

Figure 4.3. Violin plots depicting the development of control and hybrid crosses at 20°C. Percentage of (A) settlers after 24 hours, (B) pre-metamorphs after 72 hours, (C) post-metamorphs after 72 hours, (D) pre-metamorphs after 120 hours, and (E) post-metamorphs after 120 hours. Percentage values are means; error bars denote standard error (SE). Dots represent raw data points. Letters indicate homogenous groups identified by post-hoc Tukey tests.

All reciprocal crosses at 20°C produced well-developed motile larvae (see Figure 7.19), with hatching occurring ~12 hr after fertilization. There were significant differences in settlement success (Table 4.1) between crosses under in situ temperatures (i.e., 20°C), with control *P. stolonifera* crosses being more successful than control *P. herdmani* (p < .05, Tukey HSD test; Table 4.1, Figure 4.3A). However, there was no significant difference in pre-metamorph success after 72
or 120 hr or in post-metamorph success after 72 hr (Table 4.1B–D). Despite this, we found that after 120 hr there was a significant difference between the proportion of larvae that developed into post-metamorphs from control *P. stolonifera* crosses and control *P. herdmani* crosses (p < .05, Tukey HSD test, Table 4.1E, Figure 4.3E).

### 4.4.3 Development at experimental treatment temperatures

![Stacked bar plots showing the percentage of post-metamorphs, pre-metamorphs, and other/dead stages of larvae at different times across temperature treatments. First letter of cross abbreviation represents mother species and second letter represents father species.](image)

Figure 4.4. Stacked bar plots showing the percentage of post-metamorphs, pre-metamorphs, and other/dead stages of larvae at different times across temperature treatments. First letter of cross abbreviation represents mother species and second letter represents father species.
Pre-metamorphs developed under a wide range of temperatures for all crosses, only failing to develop after 120 hr at 28°C (all crosses), 12°C (hybrids crosses only) and 16°C (hybrid cross with eggs from *P. stolonifera* only; Figures 4.4 and 4.5). Post-metamorphs developed in all crosses after 120 hr, with this development only apparent at 16, 20 and 24°C.

Figure 4.5. Effects of temperature on the pre-metamorph development after 120 hours of control and hybrid crosses. Percentage values are means; error bars denote standard error (SE). Letters indicate homogenous groups identified by post-hoc Tukey tests.

4.4.4 Effects of temperature and cross on pre-metamorphic development

There was no significant interaction between the effects of temperature and cross on pre-metamorph development after 120 hr (chi-square = 8.429, df = 6, p = .208; Table 4.2A), but there were significant effects of both temperature and cross individually on pre-metamorph development (temperature: chi-square = 17.433, df = 3, p < .001; cross: chi-square = 8.368, df = 3, p < .05). The percentages of control *P. herdmani* and control *P. stolonifera* pre-metamorphs after 120 hr were similar at each temperature treatment except for 24°C (Tukey post hoc test, p < .05, Figure 4.5). For all crosses, the lowest percentage of pre-metamorphs developed at 12°C (except 28°C, where no pre-metamorphs developed for any cross).
4.4.5 Effects of temperature and cross on post-metamorphic development

Figure 4.6. Effect of temperature on the development of different control and hybrid crosses on post-metamorphic development after 120 hours. Error bars denote standard error (SE). Letters indicate homogenous groups identified by post-hoc Tukey tests. Note absence of post-metamorphs at 12 and 28°C.

There was a low percentage of post-metamorphs at all temperatures (mean < 30%). In contrast to pre-metamorphs, there was a statistically significant interaction between the effects of temperature and cross on post-metamorphic development after 120 hr (chi-square = 13.384, df = 4, p ≤ .01; Table 4.2B). The highest percentage of post-metamorphs for control *P. herdmani* crosses was found at 24°C, whereas for control *P. stolonifera* this was found at 20°C (Figure 4.6). Interestingly, maximum values for hybrid post-metamorphs were recorded at 20°C and were intermediate between the values for the two control crosses at that temperature (Figure 4.6). Although some post hoc comparisons were unable to detect significant differences (Figure 4.6), the overall pattern showed no survival of the post-metamorph stage at the lowest and highest temperatures, with limited differences among crosses at intermediate temperatures.
Table 4.1. Results of generalised linear models with a binomial error distribution and a logit link function testing the effects of cross at 20°C, on the proportion of (A) settlers after 24 hours, (B) pre-metamorphs after 72 hours, (C) post-metamorphs after 72 hours, (D) pre-metamorphs after 120 hours, and (E) post-metamorphs after 120 hours.

<table>
<thead>
<tr>
<th>Source</th>
<th>Chi-square</th>
<th>d.f.</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Proportion of settlers at 24 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>19.481</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(B) Proportion of pre-metamorphs at 72 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>5.540</td>
<td>3</td>
<td>0.063</td>
</tr>
<tr>
<td>(C) Proportion of post-metamorphs at 72 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>2.663</td>
<td>3</td>
<td>0.264</td>
</tr>
<tr>
<td>(D) Proportion of pre-metamorphs at 120 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>0.348</td>
<td>3</td>
<td>0.951</td>
</tr>
<tr>
<td>(E) Proportion of post-metamorphs at 120 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross</td>
<td>9.857</td>
<td>3</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Table 4.2. Results of generalised linear models with binomial error distributions and a logit link function testing the effects of temperature and cross on the proportion of (A) pre-metamorphs and (B) post-metamorphs.

<table>
<thead>
<tr>
<th>Source</th>
<th>Chi-square</th>
<th>d.f.</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Proportion of pre-metamorphs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td>17.433</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cross</td>
<td>8.368</td>
<td>3</td>
<td>0.039</td>
</tr>
<tr>
<td>Temp x Cross</td>
<td>8.429</td>
<td>6</td>
<td>0.208</td>
</tr>
<tr>
<td>(B) Proportion of post-metamorphs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td>4.130</td>
<td>2</td>
<td>0.127</td>
</tr>
<tr>
<td>Cross</td>
<td>4.699</td>
<td>3</td>
<td>0.195</td>
</tr>
<tr>
<td>Temp x Cross</td>
<td>13.384</td>
<td>4</td>
<td>0.010</td>
</tr>
</tbody>
</table>
4.5 Discussion

Our results showed that ontogenetic stages of hybrids survived a narrower range of temperatures than either parental species, suggesting that hybrids are unlikely to both expand their ranges to dissimilar environments and perform better under future conditions. We also found that the more widespread parental species performed better at higher temperatures. Thus, our results indicate that the more tolerant parental species (*P. herdmani*) may perform better under warming conditions. Finally, our study provides insights into how CCC may inhibit both the ability of hybrids to expand their ranges and the occurrence of speciation due to hybridization.

We confirmed that reciprocal fertilization between *P. herdmani* and *P. stolonifera* produces viable offspring that can develop to the post-metamorph stage. We hypothesized that hybrids would perform differently than parental species at extreme temperatures (e.g., Welch & Rieseberg, 2002) and found that neither hybrid cross showed broader temperature tolerances than the parental crosses. In southern Africa, evidence of hybridization between *P. herdmani* and *P. stolonifera* had, until now, been anecdotal (Rius & Teske, 2013). Our study confirms that fertilization can occur between these species and provides empirical evidence of the relative success of hybrids in a laboratory setting. However, future studies are needed to study the fertility of these hybrids and their viability in the field. Even though the study species live in sympathy (Figure 4.1), whether hybrids occur naturally remains unclear. *Pyura stolonifera* is common along rocky shores with high wave-exposure, whereas *P. herdmani* often inhabits more sheltered regions (Rius & Teske, 2011). Consequently, opportunities for hybridization at locations where both species are found may be fewer than initially expected. Mosaic-style hybrid zones have been reported along the European Atlantic coast, where salinity, wave exposure and tidal height explain the spatial distribution of alleles within the hybrid zones of the mussels *Mytilus edulis* and *Mytilus galloprovincialis* (Gardner, 1994). Niche segregation may, therefore, contribute to a low prevalence of hybrids, minimising or even preventing gene flow between species. Gene flow can also be reduced by spawning asynchrony, something that could occur among individuals separated by as little as 10s of metres (Marshall, 2002).

Hybrids performed generally well but in a narrower range of temperatures than either parent species. Considering the direction of predicted global SST changes, it appears unlikely that hybrids of *P. herdmani* and *P. stolonifera* will expand their range to locations with novel environmental conditions. While hybridization has been suggested to contribute to range shifts (Chown *et al.*, 2015; Pfennig *et al.*, 2016), especially considering expected CCC conditions (Chunco, 2014), our
results suggest that this pattern may not be as general as previously reported, as many failed hybridization events occur in the field and remain unreported.

Irrespective of parental source, all crosses showed a wider range of thermal tolerance at the pre-metamorph stage than the post-metamorph stage (Figures 4.4–4.6). This contrasts other studies using ascidians that found that later life history stages are less sensitive to environmental stress than earlier ones (Pineda et al., 2012), but is in line with studies showing that ontogenetic stages of copepods and gastropods are more tolerant to heat stress than later developmental stages (Diederich & Pechenik, 2013; Tangwancharoen & Burton, 2014). A possible explanation to this pattern is that early developmental stages of intertidal organisms experience a high variability of conditions and selection favours larvae with high thermal tolerance (Tangwancharoen & Burton, 2014). Both *P. herdmani* and *P. stolonifera* can be found in the low intertidal, where ontogenetic stages have to survive highly variable temperatures, whereas previous studies of ascidians have focussed on subtidal species (Pineda et al., 2012) that experience more stable conditions than in the intertidal zone.

While hybrid crosses did not outperform both parental crosses at any temperature (Figures 4.5 and 4.6), the percentage of post-metamorph hybrids at the temperature treatment matching in situ conditions at our sampling sites (i.e., 20°C, approximate yearly mean SST of sample sites, Figure 7.20) was intermediate to values for the parental crosses at this temperature (Figure 4.6). This is in line with previous studies that have reported hybrids possessing similar fitness to parent species (Arnold & Hodges, 1995). This suggests that while temperature does not preclude the ability of hybrids and parental species to live in an area of sympatry, hybrids are unlikely to spread to locations with an environmental (temperature) mismatch. There are two caveats to our findings. First, we inferred in situ temperature using satellite data rather than field measurements. In situ field measurements of temperature are clearly the best way to assess the effect of temperature in the intertidal zone. However, when field and satellite SST data have been compared in the study area, studies have found an almost complete matching [see figure 4 in Smit et al. (2013)]. A second caveat is that we only used 20°C to perform fertilization, and thus, poor performance in some treatments could be a result of thermal shock from fertilization to when the petri dishes were placed at the different temperature treatments. Performing the crosses at a range of temperatures would have tackled this and allowed testing the effects of temperature from fertilization to subsequent development stages.

The ability of species to develop successfully under conditions dissimilar to their native environment suggests the potential for future range expansions (Sorte et al., 2010; Rius et al., 2014a). Although some of the differences among crosses were not significant, the highest survival
of post-metamorphs at 24°C were from *P. herdmani* control crosses. Interestingly, the areas to which *P. herdmani* has supposedly expanded to (northern Africa and southern Europe) exhibit temperatures that are either similar to or cooler than those in Port Elizabeth. It has been suggested that the current disjunct range of *P. herdmani* is a relic of a historical Gondwanan distribution (Kott, 1985, 2006), but another possibility is that this distribution is due to a combination of modern anthropogenic transport and an ancient long-distance dispersal event (Teske et al., 2011; Rius et al., 2017). Given the close geographical proximity between a possible Moroccan source and the recent southern European population and the fact that our results indicate that *P. herdmani* can survive at lower temperatures, it seems likely that these regions are in gene flow contact. As seen in *P. herdmani*, the ability of *P. stolonifera* to develop to the post-metamorph stage successfully at 16, 20 and 24°C implies a wide range of temperature tolerance. Indeed, *P. stolonifera* has been recorded along an extensive stretch of the southern Africa coastline (Monniot et al., 2001), encompassing a wide range of temperature conditions (Rius et al., 2014a).

Anthropogenically induced CCC is causing major alterations in the marine environment at an unprecedented rate. Mean global SST has increased at 0.07°C per decade since 1960 and at 0.11 ± 0.02°C per decade since the 1970s (Burrows et al., 2011). This upwards trend is predicted to continue, with mean global SST projected to increase by up to 2°C by 2060 (Kirtman et al., 2013). Despite this global trend of SST warming, there is increasing evidence that SST values are becoming more extreme in certain regions (Taboada & Anadón, 2012). Between 1960 and 2010, coastal waters around South Africa exhibited both warming and cooling of SST (Rius et al., 2014a). The SST of our study area, the Port Elizabeth region, cooled between 1982 and 2009, with especially strong cooling during austral winter months, while farther east, the coast experienced strong warming of SST over the same period (Rouault et al., 2010). The sampled region of this study is ~300 km away from the documented eastern-most limit of the temperate lineage and the western limit of the subtropical lineage of *P. herdmani* (Teske et al., 2011). If the sub-tropical lineage had expanded southwards since the collection of the samples analysed in Teske et al. (2011), either naturally or through human-mediated transport, then, it means that this lineage is now present in Port Elizabeth. As a result, the possibility exist that sub-tropical individuals were collected and crosses, which may explain the highest success of *P. herdmani* post-metamorphs at 24°C.

Previous studies on marine invertebrates inhabiting the south-east coast of Africa have suggested the presence of a biogeographic break reflecting oceanographic conditions and dispersal rather than physiological tolerance (Teske et al., 2008; Zardi et al., 2011). For example, the invasive mussel *M. galloprovincialis* is prevented from further spread towards the east of the South
African coastline by a steep transition between cool-temperate and subtropical waters along the south-eastern coast (Assis et al., 2015). Therefore, it appears unlikely that either lineage of *P. herdmani* could have spread across this biogeographic break naturally. Nevertheless, our results suggest that divergence between these divergent lineages is driven by prezygotic barriers (e.g., oceanography and dispersal) rather than thermal tolerance. It is likely that the dispersal of *P. stolonifera* to the east is similarly limited, despite the ability of its larvae to develop at the higher temperatures exhibited there (Figures 4.5 and 4.6). Extreme temperature treatments (12 and 28°C) precluded post-metamorph development in all crosses, indicating that these temperatures are outside the thermal thresholds of both species and all hybrids and that the species' ranges are unlikely to expand to cooler temperate or tropical regions (or at least regions where these temperatures coincide with spawning periods). Two points temper this interpretation. First, we sampled animals from a single area, and thus, we did not consider that individual responses may vary throughout a species range (Neuheimer et al., 2011), and second, future temperature change is likely to be gradual, raising the possibility of rapid adaptation helping to cope with CCC in both species.

Early ontogenetic stages are often particularly sensitive in marine invertebrates (Verween et al., 2007; Pineda et al., 2012), and consideration of multiple life-history stages is key when exploring thermal tolerances, as these can vary considerably across the life-cycle (Rius et al., 2010a). Changing SST will render some previously inhospitable environments habitable (Poloczanska et al., 2016), and this could occur in our study system. Mass mortalities of *P. stolonifera* have occurred along the southern coast of Africa in both 1991 (Hanekom et al., 1999) and 2012 (Hanekom, 2013) and have been attributed to rapidly changing air and sea temperatures. The coastline where these mass mortalities took place is within the area of sympatry for *P. herdmani* and *P. stolonifera* (Figure 4.1). It is unknown whether the species identified as *P. stolonifera* in Hanekom et al. (1999) and Hanekom (2013) were *P. stolonifera*, *P. herdmani* or hybrids. If only *P. stolonifera* is affected by these mortalities, there may be the potential for *P. herdmani* or hybrids to quickly occupy new available substratum (Rius et al., 2017). Rates of change in SST are not consistent throughout the year in southern Africa (Rouault et al., 2010), and fluctuations in maximum and minimum SST in the study region have become more extreme in recent years (Rius et al., 2014a). Therefore, while fluctuating climatic conditions may not promote the expansion of hybrids to novel locations per se, the opening of an ecological niche may ultimately affect community composition (Sagarin et al., 1999).

To conclude, we found that: (a) In situ temperature conditions did not preclude hybridization between *P. herdmani* and *P. stolonifera* and subsequent development from larvae to the post-metamorph stage; (b) Hybridization did not enhance survival under a wider range of
temperatures; (c) Changes in SST as a result of CCC may enhance range expansions of the parental species but not the hybrids. Our results indicate that offspring performance at a variety of temperatures may be a good predictor of range expansions, and that ongoing CCC may inhibit, rather than promote range expansions by hybrids.

4.6 Acknowledgements

We would like to thank Jaqui Trassierra, Aldwyn Ndhlovu and Dr Cristian Monaco for their assistance in sample collection. We acknowledge Dr Carlota Fernández Muñiz for her invaluable assistance in the culturing of algae for the experiments, Dr Ivan Haigh for assistance collecting and analysing seawater temperature data and Prof Dustin Marshall for his help in methodological aspects. Funding for JH’s stay at Rhodes University was provided for by the Research Training and Support Grant from the Southampton Marine and Maritime Institution. This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation (Grant number 64801).
Chapter 5     Genomic signatures of local adaptation to different depths in a dominant marine invertebrate

5.1     Abstract

Identifying genetic signatures of local adaptation are key for understanding the mechanisms behind adaptive divergence. Most population genomics studies assume that species are distributed across a two-dimensional space (e.g. latitude and longitude). However, the vertical plane of altitude or of depths and associated environmental gradients are also expected to be important. In aquatic ecosystems, thermoclines and pycnoclines in the water column or redox horizons in sediments are known to influence species distributions and are likely to influence patterns of local adaptation. Environmental gradients associated with depth are much stronger and occur across smaller spatial scales than those with altitude. We studied a dominant marine epibenthic invertebrate for which previous studies have shown the presence of depth-defined genomic lineages, and thus hypothesised the existence of depth-related signatures of local adaptation. We first conducted localised population genomic analyses and focussed on analysing outlier loci that might be under selection or linked to selected loci. We found that 8.9% of 2,626 SNPs were highly correlated with key environmental variables such as temperature, salinity, and chlorophyll $a$ concentration. We mapped these SNPs to a genome of a closely related species, and functionally annotated genes close to these loci. We found 28 enriched gene ontology terms that were over-represented in the loci, which suggest that these genes are associated with adaptation to depth. We propose that local adaptation at different depths and the associated divergent selection support the observed population divergence. Our study emphasises the need for incorporating this vertical plane in genomic studies of local adaptation.

5.2     Introduction

The use of high throughput sequencing enables the study of thousands of loci, with differences in genomic variation amongst populations being used to infer historical evolutionary processes (Stapley et al., 2010; McCormack et al., 2013; Gagnaire et al., 2015; Rius et al., 2015a). Often of interest is the study of the subset of these loci that display greater genetic differentiation than expected by neutral models (outlier loci; Narum & Hess, 2011). It is often assumed that selection is acting directly on these outlier loci, or at loci in close proximity to the outlier loci (Nosil et al., 2009). However, this inference relies on equilibrium between migration, selection, mutation, and
Chapter 5

Drift amongst populations (Whitlock & McCauley, 1999). As such, other explanations for patterns of outlier loci have been put forward including demographic history and population structure (Hoban et al., 2016), gene surfing through mutations at the front of range expansions (Excoffier et al., 2009), background selection against deleterious mutations (Cruickshank & Hahn, 2014), coupling between intrinsic and extrinsic reproductive barriers (Bierne et al., 2011), and spatial autocorrelation of alleles (Hoban et al., 2016). Therefore, careful consideration is required to confidently disentangle the different roles these mechanisms play in heterogeneous genomic differentiation among populations.

Divergent selection often acts on populations living under different environmental conditions (Abebe et al., 2015). As a result, individuals evolve to be fitter to their local habitat than individuals originating from elsewhere, a mechanism known as local adaptation (Kawecki & Ebert, 2004). Studies that assess the roles of abiotic factors and selective forces across areas of environmental heterogeneity mostly focus on signatures of local adaptation across horizontal planes (Sanford & Kelly, 2011), such as selective pressures associated with latitude and longitude. As a result, studies of local adaptation often only consider two spatial dimensions and overlook the vertical plane (i.e. elevation or depth). However, there is a growing number of studies assessing the effects of local adaptation in the context of altitude in terrestrial systems, as regions at different altitudes experience vastly different environmental factors including rainfall, temperature, atmospheric pressure, and solar radiation (Körner, 2007). These environments represent appropriate systems to study local adaptation as changes in elevation provide shorter and steeper gradients than latitude and longitude (Körner, 2007; Hargreaves et al., 2014).

Signatures of local adaptation to altitude have been identified in plant and tree species (Kim & Donohue, 2013; Marcora et al., 2017; Hämälä et al., 2018), as well as vertebrate (Fischer et al., 2011; Muir et al., 2014) and invertebrate animal species (Zhang et al., 2017; Gamboa & Watanabe, 2019). In some cases, local adaptation to altitude has also been identified as a driver of speciation with gene flow (Chapman et al., 2013). Regarding aquatic environments, whilst it is well known that different species experience adaptations to different depths (e.g., Sebert, 2002), little is known about the effects of depth on local adaptation across small spatial scales. The steep physical gradients associated with depth, [e.g. changes in light, pressure, nutrients, temperature, and salinity (Rosenberg et al., 1992; Somero, 1992)] are much stronger and occur across smaller spatial scales (Terlizzi et al., 2003) than those with altitude, and differences in gradient steepness have been shown to affect adaptive population divergence (Bachmann et al., 2020). Marine environments therefore provide an appropriate opportunity to study local adaptation across seascapes, as depth adds an additional component of variation that is often ignored in local adaptation studies.
Seascapes have been considered to have fewer physical barriers than landscapes (May et al., 1994). This, together with the high capacity for dispersal of planktonic larval stages due to ocean currents and tidal fronts (Siegel et al., 2003), has historically led to the belief that marine populations are demographically open (Caley et al., 1996) with lower levels of local adaptation than terrestrial environments (Sanford & Kelly, 2011). However, there is a growing appreciation that local adaptation occurs in marine species across a range of spatial scales (Sotka, 2005; Sanford & Kelly, 2011). For example, evidence for local adaptation has been identified in species from regional (Hess et al., 2013; Milano et al., 2014) down to local scales with patterns identified within individual rocky shores, where environmental gradients are exceptionally steep (Sherman & Ayre, 2008). In the past decade however, a combination of studies investigating both nuclear and mtDNA markers and performing transplant experiments have identified depth-defined divergence in both Australian and Caribbean corals (Bongaerts et al., 2011; Prada & Hellberg, 2013). In addition, studies using high throughput sequencing have suggested the presence of depth-association adaption that may lead to genomic divergence (Bongaerts et al., 2017; Pratlong et al., 2018). Despite all this research progress, no functional annotation of outlier loci have been reported in these studies, and therefore it remains unclear whether this depth-associated divergence is the result of neutral genetic processes, local adaptation to different environmental pressures, or a combination of both. Thus, there is a need to specifically address how local adaptation is associated with depth-defined environments.

Along the east coast of the North Sea, the outflow of large amounts of brackish water from the Baltic Sea leads to strongly stratified waters and a well-developed pycnocline, with surface waters down to 10-15 m experiencing lower salinities and more variable temperatures than the more stable, cooler, and more saline deeper water. As a result, low motility benthic species are expected to exhibit adaptation to the particularities of this sharp variation in abiotic conditions (Decker et al., 2003; Renborg et al., 2014). The ascidian *Ciona intestinalis* (Asciidiacea, Tunicata, Chordata) can be found on both sides of this pycnocline, from surface waters to depths of >50 m (Dybern, 1965). Individuals from shallow and deep waters exhibit differing spawning rates and generation lengths (Dybern, 1965) and recent work using microsatellites and neutral single nucleotide polymorphisms (SNPs) has shown strong genetic differentiation across the pycnocline at an extremely fine vertical spatial scales (metres; Johannesson et al., 2018; Hudson et al., 2020). In addition, these studies have shown little differentiation among sites separated by horizontal distances of up to ~100km, suggesting that this pycnocline may be a strong barrier to gene flow.

Here, we used population genomic data and environmental parameters to identify outlier loci from *C. intestinalis* populations found at different depths within the eastern North Sea. We also conducted experimental crosses within and between deep and shallow lineages to assess how
different environmental conditions influence offspring performance. We hypothesised that outlier loci, whilst not necessarily being the target of selection, would be in genomic regions containing the targets, and therefore functional annotation of genes in these regions would be enriched for processes underlying local adaptation. This will help unravel key abiotic factors associated with depth and enable us to separate the signatures of depth-associated selection from other signatures leading to genomic differentiation. Regarding the crosses, we expect that offspring will perform better under the conditions where adults can be found and that there will be no evidence of reproductive isolation.

5.3 Methods

5.3.1 Sampling

Adult *C. intestinalis* were sampled from a region within the eastern North Sea (Figure 5.1) where two genetically differentiated lineages coexist (Johannesson *et al.*, 2018; Hudson *et al.*, 2020). We collected a total of 160 individuals from shallow and deep sites (Table 5.1) that were used for our population genomic analysis. As *C. intestinalis* is known to secrete mucus strings that facilitate larval retention (Svane & Havenhand, 1993), we sampled individuals separated by at least one metre to minimise the chance of collecting closely-related individuals.

Table 5.1. Sampling information for *Ciona intestinalis*, including geographical region, site abbreviation code, coordinates of sampling sites, depth (shallow [<10m] or deep [>15m]), substratum type, and the number of individuals used in genomic analyses.

<table>
<thead>
<tr>
<th>Site name</th>
<th>Site number</th>
<th>Code</th>
<th>Latitude (N)</th>
<th>Longitude (E or W)</th>
<th>Depth (category)</th>
<th>Substratum</th>
<th>No. of individuals analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burholmen</td>
<td>1</td>
<td>BUH</td>
<td>58.89°</td>
<td>11.13°</td>
<td>5m (Shallow)</td>
<td>Natural</td>
<td>16</td>
</tr>
<tr>
<td>South Koster</td>
<td>2</td>
<td>KOS</td>
<td>58.88°</td>
<td>11.05°</td>
<td>3-4m (Shallow)</td>
<td>Natural</td>
<td>15</td>
</tr>
<tr>
<td>Lindholmen</td>
<td>3</td>
<td>LIN</td>
<td>58.88°</td>
<td>11.15°</td>
<td>0-1m (Shallow)</td>
<td>Artificial</td>
<td>14</td>
</tr>
<tr>
<td>Vattenholmen</td>
<td>4</td>
<td>VAT</td>
<td>58.87°</td>
<td>11.09°</td>
<td>60m (Deep)</td>
<td>Natural</td>
<td>16</td>
</tr>
<tr>
<td>Brattskär</td>
<td>5</td>
<td>BRA</td>
<td>58.86°</td>
<td>11.07°</td>
<td>1-4m (Shallow)</td>
<td>Artificial</td>
<td>15</td>
</tr>
<tr>
<td>Kåva</td>
<td>6</td>
<td>KAV</td>
<td>58.33°</td>
<td>11.36°</td>
<td>18-22m (Deep)</td>
<td>Natural</td>
<td>16</td>
</tr>
<tr>
<td>Gåseklåvan</td>
<td>7</td>
<td>GUL</td>
<td>58.31°</td>
<td>11.54°</td>
<td>20-25m (Deep)</td>
<td>Natural</td>
<td>15</td>
</tr>
<tr>
<td>Jämningarna</td>
<td>8a</td>
<td>JAM_D</td>
<td>58.26°</td>
<td>11.39°</td>
<td>17-20m (Deep)</td>
<td>Natural</td>
<td>8</td>
</tr>
<tr>
<td>Jämningarna</td>
<td>8b</td>
<td>JAM_S</td>
<td>58.26°</td>
<td>11.39°</td>
<td>5-7m (Shallow)</td>
<td>Natural</td>
<td>16</td>
</tr>
<tr>
<td>Fiskebäckskil</td>
<td>9</td>
<td>FIS</td>
<td>58.24°</td>
<td>11.46°</td>
<td>0.5 - 2m (Shallow)</td>
<td>Artificial</td>
<td>15</td>
</tr>
<tr>
<td>Porsholmen</td>
<td>10</td>
<td>POR</td>
<td>58.23°</td>
<td>11.40°</td>
<td>2-4m (Shallow)</td>
<td>Natural</td>
<td>14</td>
</tr>
</tbody>
</table>
Figure 5.1. Eastern North Sea coastline where samples of *Ciona intestinalis* were collected. Purple circles represent shallow sampling sites and orange circles represent deep sampling sites. Site numbers are 1: Burholmen, 2: South Koster, 3: Lindholmen, 4: Vattenholmen, 5: Brattholmen, 6: Kävra, 7: Gåseklåvan, 8: Jämningarna, 9: Fiskebäckskil, 10: Porsholmen.
Chapter 5

5.3.2 DNA extraction and genotyping

Individuals used for genomic analyses had a small section of the branchial sac tissue preserved in 95% ethanol and stored at -20°C. Ethanol was periodically changed until no more pigment leached into the preservative. DNA was extracted using the Qiagen DNeasy® Tissue Kit (Qiagen) according to the manufacturer’s protocol, and DNA quality and quantity were assessed by gel electrophoresis and the Quantifluor® dsDNA System (Promega) respectively. Extracted DNA was genotyped using a genotyped-by-sequencing (GBS) approach (Elshire et al., 2011) at the University of Wisconsin Biotechnology Center.

Reads were assembled using ipyrad v. 0.9.13 (Eaton, 2014). Ipyrad uses seven sequential steps to perform GBS assembly, and we used parameters based on those recommended for single-end GBS data (http://ipyrad.readthedocs.io/). The default level of sequence similarity for clustering (85%) was quite conservative for a population genomic analysis and appeared to cause excess clustering of genomic regions that should be distinct. We therefore set this value to 90%, but the remaining parameter settings for de novo assembly remained as default. We filtered for loci with a minimum genotype depth of less than five reads, a mean minimum depth across individuals of less than 15 reads, and removed loci with a minor allele count of less than three. This final filtering step was used in place of the commonly used minor allele frequency cut off of 5%, as this may remove informative but rare alleles (O’Connor et al., 2015). We then used an iterative filtering framework (O’Leary et al., 2018) so that our final dataset contained loci with at least a 50% call rate, and a maximum of 50% missing data per individual, using vcftools v.0.1.13 (Danecek et al., 2011).

5.3.3 Identification of outlier loci

We created three separate datasets based on the identification of outlier loci, detailed below. Firstly, we created a dataset of loci identified by pcadapt (hereafter known as “pcadapt dataset”); secondly, we created a dataset of loci identified by Bayenv2.0 (“bayenv dataset”), and finally we conservatively created a putative neutral dataset of loci identified by either pcadapt or Bayenv2.0 (“neutral dataset”).

5.3.4 Identification of outlier loci associated to population structure

Previous studies have shown that outlier loci can recover population structure uncaptured by neutral loci markers (Gagnaire et al., 2015; Tigano et al., 2017). We therefore used pcadapt v.4.1.0 (Luu et al., 2017), which ascertains population structure using a principal component analysis (PCA), to identify candidate loci excessively related to population structure using a false
discovery rate of 5%. However, whilst pcdadapt can provide a list of outlier loci associated with population structure, it cannot explain what forces are driving this structure, and therefore was only used in conjunction with other methods.

5.3.5 Identification of outlier loci associated with environmental variables

We obtained monthly sea temperature, salinity, and chlorophyll $a$ data for each site from the Swedish Meteorological and Hydrological Institute (SMHI) Svenskt HavsARKiv (SHARK) database. Using these data, we calculated for each parameter the average summer (June – August), winter (December – February), and annual value at each site. All of these analyses were performed in R version 3.3.1 (R Core Team, 2016).

Assessments of the degree of correlation between environmental variables and genomic data were carried out using Bayenv2.0 (Günther & Coop, 2013). Briefly, Bayenv2.0 estimates a matrix of covariance in allele frequencies among populations. This covariance matrix is then used as a null model, where Bayenv2.0 assesses correlation between environmental variables and allele frequencies by comparing Bayes factors (the measure of support for alternating models). Due to potential issues of instability across independent runs of Bayenv2.0 in making the covariance matrix (Blair et al., 2014), we took the mean values of 10 independent runs, each run consisting of 100,000 MCMC iterations. To assess whether the mean covariance matrix represented the true variance of allele frequencies, it was compared to the pairwise $F_{ST}$ matrix (derived from 1,653 putatively neutral SNPs for the same sites) from Hudson et al. (2020) using a Mantel test in R with 1,000 permutations. We then assessed correlation between individual SNPs and the nine environmental variables mentioned above. This approach was repeated ten times for each combination of locus-environmental variable to account for instability between independent runs (Blair et al., 2014), and the mean of the ten runs was then used to infer the final Bayes factors (BF). The Kass and Raftery (1995) criterion was then used to determine the probability of these loci being under selection, with BF values above 20 classified as strong probability. Whilst this BF value is large, including loci identified with a more relaxed BF criterion resulted in many more being retained suggesting a high probability of false positives. Finally, after running Bayenv2.0 on seasonal parameters (i.e. summer, winter, and annual means) to identify loci correlated to specific aspects of environmental variability, the loci recovered were grouped together as being correlated to temperature, salinity, or chlorophyll in general.
5.3.6 Population genomic analyses

We calculated global $F_{ST}$ values and pairwise $F_{ST}$ comparisons between populations using the R package *hierfstat* (Goudet, 2005; R Core Team, 2016) and corrected for multiple comparisons using the Benjamini-Yekutiel method (Yekutieli & Benjamini, 1999). We performed clustering analyses using the program ADMIXTURE (Alexander et al., 2009) and a discriminant analysis of principal components (DAPC; Jombart et al., 2010) using the R package adegenet v.2.1.1 (Jombart, 2008). ADMIXTURE estimates the likelihood that an individual comes from one of a pre-defined number of putative sample populations ($K$) using a maximum likelihood estimation from multilocus SNP genotype datasets, and calculates estimates using a faster numerical optimisation algorithm than the more commonly used programme STRUCTURE (Pritchard et al., 2000). We performed a twenty-fold cross-validation procedure within ADMIXTURE to infer the best supported value of $K$. DAPC transforms data using PCA before using PCA factors as variables for a discriminant analysis (DA), ultimately maximising the differences among groups while minimising variation within groups (Jombart et al., 2010). For DAPC we used the `xvalDapc` function to calculate an appropriate number of principal components to be retained.

We assessed for potential hybridisation between the two lineages using *parallelnewhybrid* (Wringe et al., 2017), which runs the Bayesian model-based clustering framework found in NEWHYBRIDS v1.1 (Anderson & Thompson, 2002) but is able to utilise the additional power of multicore functioning. *Parallelnewhybrid* computes a posterior probability that individuals belong to distinct hybrid classes (or parental classes) by Markov chain Monte Carlo (MCMC). Due to computational limitations of NEWHYBRIDS and *parallelnewhybrid*, rather than using the full loci dataset, we ran *parallelnewhybrid* using both the bayenv and pcadapt dataset (see below for descriptions of these datasets). To confirm the ability of *parallelnewhybrid* to identify simulated hybrid genotype classes correctly, we selected the 30 individuals that had the highest ancestry coefficient ($Q$ value; $Q > 0.9$) for each lineage as calculated using ADMIXTURE, and simulated 30 individuals within F1, F2, and backcross hybrid classes using HYBRIDLAB (Nielsen et al., 2006), before running this new dataset containing 60 empirical and 120 simulated individuals in *parallelnewhybrid*. For each dataset, we performed five independent runs, and took the mean posterior probability per individual. Independent runs were performed with a 10,000 burn-in period followed by 50,000 sweeps using Jeffrey’s-like priors for estimating allele frequencies and mixing proportions. Individuals were considered to belong to a simulated genotype class if their posterior probability was $> 0.5$, following the suggestions of Anderson & Thompson (2002).
5.3.7 Annotation of outlier loci

As one can assume the loci in the pcdadapt dataset are associated with population structure, these may have differentiated between populations due to neutral evolutionary processes and therefore may not actively play a role in promoting local adaptation. Therefore, we instead used genes associated with the loci within the bayenv dataset to identify candidate loci for natural selection. Firstly, we created a consensus sequence for each locus using the R package DECIPHER (Wright, 2016). The BLAST tool on Ensembl server (Ensembl release 98; Hunt et al., 2018) was then used to match for hits against the genome of the sister species Ciona robusta (GCA_000224145.2; note though that this is mislabelled as C. intestinalis on Ensembl). For this search, we set a maximum e-value of $e^{-10}$ to limit poor matches. We then extracted genomic location coordinates for each matched GBS locus, mapped them to an appropriate chromosome, and identified genes within ±39.5 Kb of each GBS loci (i.e. a 79 Kb window) to ascertain genes which may putatively be linked. A value of 39.5 Kb was chosen as a 1 cM value of 39.5 Kb has been reported in the C. robusta genome (Caputi et al., 2008). For each gene in this region, we ascertained how many were annotated (i.e. not labelled “uncharacterised locus”) within the Ensembl database. To test if any biological processes were enriched more than expected based on background levels of gene function, these characterised genes were then run through the Gene Ontology (GO) enrichment analysis comparing against genes in the Ciona intestinalis reference dataset using Fisher’s exact test with significance corrected using False Discovery Rate [FDR < 0.05] (Ashburner et al., 2000; Mi et al., 2019; The Gene Ontology Consortium, 2019). Finally, to assess whether there was clustering of outlier loci across the genome, we split the genome into 1 Mb ‘bins’ and used Fisher’s exact test to compare the number of outlier loci located within each bin and the number of expected outlier loci in each bin based on the distribution of the total 2,626 GBS loci recovered.

5.3.8 Laboratory housing of animals, fertilisation methods and data analysis of experimental crosses

Individuals were collected from a deep site (Gåsekåvan, site 7 in Figure 5.1) and a shallow site (Lindholmen, site 3 in Figure 5.1). Collected ascidians were maintained inside aquaria at the Tjärnö Marine Laboratory, Sweden, and kept at either 12°C and 34 PSU (deep individuals) or 16°C and 30 PSU (shallow individuals) for no more than two days to limit acclimation to laboratory conditions. These temperatures and salinities were chosen as they closely matched field conditions experienced during the experiment [as assessed using live salinity and temperature probes from Lindholmen (site 3) and a site near Gåsekåvan (site 7; Figure 5.1)]. Natural food supply brought in via the flow through seawater system was supplemented daily by adding food
in the form of *Artemia* spp cultured on site. The ascidians were checked daily for signs of poor health, and if an individual showed a sign of bacterial infection, they were immediately removed to limit the spread of infection to other individuals.

Figure 5.2. Experimental set up for laboratory crosses. Briefly (A) Six individuals were sampled from deep and shallow environments, with three individuals from each environment being treated as female and three as male. (B) Eggs and sperm were dissected out from each individual and pooled together based on the source of the individual. (C) Eggs were then aliquoted into two separate Petri dishes, before (D) sperm was added to both intra- and inter-lineage eggs. (E) Eggs were left to fertilise, and larvae were subject to differing treatments.
Six individuals from Lindholmen and six from Gåseklåvan were used for crosses. As *C. intestinalis* is hermaphroditic, three individuals from each population were treated as female and three treated as male (Figure 5.2A). Filtered seawater (FSW; obtained from inflow laboratory taps at the Tjärno Marine Centre) was obtained via a vacuum pump using 0.7 µm Whatman glass microfibre filters. We extracted eggs and sperm from the oviduct and spermiduct respectively following Young & Chia (1985) (Figure 5.2B). We pooled together eggs from all shallow or deep mothers in separate dishes (Figure 5.2C), and further aliquoted these pooled eggs into two 90mm Petri dishes corresponding to whether they would be used as intra- or interpopulation crosses (Figure 5.2D). Sperm from all three males for each population were pooled (Figures 5.2A and B), before being diluted in FSW to a concentration of ca. $10^4$ cells μL$^{-1}$ (see Appendix D) using a haemocytometer. Approximately 160 μL of sperm dilution was pipetted into each Petri dish and the eggs were left to fertilise under shallow conditions (16°C and 30 PSU). Eggs were then rinsed of excess sperm using FSW once cleavage was apparent (~1 hour after sperm addition), and fertilised eggs were left at 16°C in the dark to develop and hatch. Larvae hatched after ca. 21 hours, and 20 random healthy [c.f. viable larvae in Pineda et al. (2012)] larvae were pipetted into replicate Petri dishes with FSW, which were randomly assigned to either shallow (16°C and 30 PSU) or deep water (12°C and 34 PSU) conditions (Figure 5.2E). These Petri dishes were pre-roughened and left in unfiltered seawater overnight to form a biofilm, which promotes settlement in ascidians (Wieczorek & Todd, 1997). Larvae and subsequent early life history stages were checked after 24, 72, and 120 hours at treatment conditions to assess settlement and metamorphosis success.

A general linear model using a logit link function was performed to test the effects of temperature and cross type on settlement success and post-metamorph performance at 24 and 120 hours respectively. Tukey HSD post-hoc tests were used to determine the pairwise comparisons that exhibited significant differences. All statistical analyses were performed in R (R Core Team, 2016).

### 5.4 Results

#### 5.4.1 Loci assembly

GBS generated a total of 364,193,994 raw reads, with an average of 2,167,821 reads per sample. After filtering and clustering using ipyrad and vcftools, we retained a total of 2,626 putatively unlinked SNPs in the sequence assembly. Eight individuals were removed from the dataset due to missing data (i.e. greater than 50% missing data), which was likely caused by poor DNA quality or secondary contaminants within the samples (Federman et al., 2018). This led to a final dataset of 160 individuals from 11 sampling sites.
Chapter 5

5.4.2 Identification of outlier loci

We recovered a total of 107 $F_{ST}$ outlier loci from the original 2,626 loci (~4.1% of original loci) using pcadapt. These loci formed the “pcadapt dataset”.

We confirmed that our mean matrix covariance produced by Bayenv2.0 represented the true variance of allele frequencies across populations (i.e. $F_{ST}$ matrix from Hudson et al. (2020)) using the Mantel test ($R^2 = 0.715$, $p < 0.001$). The correlation matrix of environmental variables found correlation amongst parameters (Table 7.15). Bayenv2.0 is univariate in nature so whilst no parameter was removed due to this correlation, care was taken in the interpretation of loci correlated to multiple environmental variables. In total, we identified a total of 223 outlier loci (ca. 8.9% of original loci) with strong Bayes Factors (>20) using Bayenv2.0 (Table 5.2). We therefore removed loci identified by either method to obtain the “neutral dataset” of 2,349 putatively neutral loci (for results of neutral loci, see Appendix D).

Table 5.2. Number of loci of Ciona intestinalis identified using Bayenv2.0 at different Bayes Factors, for each environmental variable.

<table>
<thead>
<tr>
<th>Environmental variable</th>
<th>Bayes Factor &gt;20</th>
<th>&gt;10</th>
<th>&gt;5</th>
<th>&gt;2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>115</td>
<td>176</td>
<td>265</td>
<td>488</td>
</tr>
<tr>
<td>Winter</td>
<td>25</td>
<td>53</td>
<td>154</td>
<td>455</td>
</tr>
<tr>
<td>Annual</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>221</td>
</tr>
<tr>
<td>Summer salinity</td>
<td>110</td>
<td>168</td>
<td>304</td>
<td>671</td>
</tr>
<tr>
<td>Winter salinity</td>
<td>100</td>
<td>173</td>
<td>316</td>
<td>628</td>
</tr>
<tr>
<td>Annual salinity</td>
<td>95</td>
<td>164</td>
<td>318</td>
<td>674</td>
</tr>
<tr>
<td>Summer Chl-a</td>
<td>16</td>
<td>46</td>
<td>117</td>
<td>444</td>
</tr>
<tr>
<td>Winter Chl-a</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>311</td>
</tr>
<tr>
<td>Annual Chl-a</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>421</td>
</tr>
</tbody>
</table>
5.4.3 Genetic structuring inferred by outlier loci

Similar genomic patterns, as visualised using DAPC, amongst deep populations were observed in the pcadapt dataset (Figure 5.3A) and the neutral dataset (Figure 7.24). However, whereas the shallow populations were tightly clustered together in the pcadapt dataset, neutral loci found all shallow sites had moved closer to the GUL site, with JAM_S and FIS deviating from the remaining sites. The pcadapt dataset maximised the differentiation both between shallow and deep, and also within deep sites, suggesting a geographical component among deep sites which is absent in the shallow sites. In contrast, the DAPC using the bayenv dataset did not separate the deep populations to such an extent, with the exception of the fjord site GUL, which was differentiated to the coastal deep sites (Figure 5.3B).
Figure 5.3. Discriminant analysis of principal components showing clear genetic differentiation by depth based on loci under selection according to (A) the pcadapt and (B) the Bayenv2 software. Blue colours represent deep populations and orange colours represent shallow populations. Sites are abbreviated as in Table 5.1.
The ADMIXTURE analyses using the two outlier datasets did not noticeably differ. The best supported values of $K$ were eight and six for the pcadapt and bayenv datasets respectively (Figures 7.21 and 7.22). Both analyses identified JAM_S as a distinct population when $K > 5$ (Figure 5.4). Furthermore, the bayenv dataset suggested that the remaining shallow southerly sites (POR and FIS) were distinct from the northerly shallow sites (Figure 5.4B). This pattern was also observed using the pcadapt dataset (Figure 5.4A) and the DAPC plot using the bayenv dataset (Figure 5.3B), albeit the strength of the pattern was weaker in the latter.
Figure 5.4. ADMIXTURE plots of multiple $K$s using (A) the pcadapt dataset and (B) the bayenv dataset. The cross-validation method within ADMIXTURE found the best supported value of $K$ to be eight and six for (A) and (B) respectively. Sites are abbreviated as in Table 5.1.
Figure 5.5. F$_{ST}$ heatmaps for outlier loci from: (A) pcadapt and (B) bayenv software. Significant pairwise comparisons [after Benjamini-Yekutieli correction] are marked by asterisks. Sites are abbreviated as in Table 5.1.

Global F$_{ST}$ for the pcadapt dataset was 0.385 and 0.302 for the bayenv dataset, greater than observed for the neutral dataset (0.101). 93% of pairwise comparisons were significant in the pcadapt dataset, and 95% of comparisons were significant in the bayenv dataset. As expected, there was a stronger pattern of differentiation between shallow and deep sites using these outlier loci datasets (Figure 5.5) than neutral loci (Figure 7.27).

Figure 5.6. Mean posterior probability of assignment to hybrid classes for each empirical and simulated individual in parallelnewhybrid using (A) the pcadapt dataset and (B) the bayenv dataset.
The results of the *parallelnewhybrid* analysis found evidence for F2 and backcrossed individuals using both the pcadapt and bayenv datasets, however most individuals (ca. 91%) were assigned to the class expected based on their sampling location depth (Figure 5.6). Testing the efficacy of NEWHYBRIDS using simulated hybrid genotypes produced by HYBRIDLAB (F1, F2, and backcrossed individuals) showed that *parallelnewhybrid* could accurately detect hybrid classes in both datasets (96.7% and 99.1% accuracy in the pcadapt and bayenv datasets respectively).

### 5.4.4 Functional annotation of outlier loci

Despite the high correlation between environmental variables (Table 7.15), a large subset of loci correlated with only one of the environmental parameters (Figure 5.7). Of the 223 outlier loci identified by Bayenv2.0, only one locus was correlated to all three environmental variables (Figure 5.7).

106 loci in the bayenv dataset could be mapped to a genomic location using the *C. robusta* genome, with regions on chromosomes 1, 2, 3, 5, and 14 containing more than one locus within a

![Venn diagram](image)
39.5 Kb window (Figure 5.8). The overall clustering of outlier loci was more than expected by chance based on genome wide clustering of all loci (Fisher’s exact test \( P < 0.001 \)).

![Figure 5.8](image)

**Figure 5.8.** Location along each of the 14 *Ciona robusta* chromosomes of the 106 outlier loci that could be mapped to a genetic location. Numbered panels represent different chromosomes, black dots represent individual loci, red arrows represent regions with two or more loci within 39.5 Kb of each other. Note chromosomes 6 and 13 did not have any loci mapped to them, and therefore are not plotted.

We identified a total of 1,379 genes within the 79 kb windows surrounding our 106 mapped GBS outlier loci from the bayenv dataset, with on average 13 (± 3.7 SD) genes associated with each mapped GBS locus. Of these 1,379 genes, 231 had been named within the ensemble database and corresponded to a total of 733 GO biological processes. The subsequent GO enrichment analysis identified 28 biological processes that were significantly enriched after false discovery rate (FDR) correction (Table 5.3).
Table 5.3. Significantly enriched (P<0.05 after FDR correction) gene ontologies for *Ciona intestinalis* based on 231 genes. Indentations indicate parent-child GO term relationship.

<table>
<thead>
<tr>
<th>GO biological process complete</th>
<th>GO term ID</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma membrane phospholipid scrambling</td>
<td>GO:0017121</td>
<td>0.034</td>
</tr>
<tr>
<td>Plasma membrane organization</td>
<td>GO:0007009</td>
<td>0.018</td>
</tr>
<tr>
<td>Cellular process</td>
<td>GO:0009987</td>
<td>0.023</td>
</tr>
<tr>
<td>Phospholipid translocation</td>
<td>GO:0045332</td>
<td>0.032</td>
</tr>
<tr>
<td>Lipid translocation</td>
<td>GO:0034204</td>
<td>0.035</td>
</tr>
<tr>
<td>Regulation of membrane lipid distribution</td>
<td>GO:0097035</td>
<td>0.031</td>
</tr>
<tr>
<td>Regulation of developmental process</td>
<td>GO:0050793</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td>GO:0006412</td>
<td>0.011</td>
</tr>
<tr>
<td>Cellular protein metabolic process</td>
<td>GO:0044267</td>
<td>0.028</td>
</tr>
<tr>
<td>Protein metabolic process</td>
<td>GO:0019538</td>
<td>0.049</td>
</tr>
<tr>
<td>Macromolecule metabolic process</td>
<td>GO:0043170</td>
<td>0.049</td>
</tr>
<tr>
<td>Organic substance metabolic process</td>
<td>GO:0071704</td>
<td>0.049</td>
</tr>
<tr>
<td>Metabolic process</td>
<td>GO:0008152</td>
<td>0.019</td>
</tr>
<tr>
<td>Primary metabolic process</td>
<td>GO:0044238</td>
<td>0.024</td>
</tr>
<tr>
<td>Nitrogen compound metabolic process</td>
<td>GO:0006807</td>
<td>0.033</td>
</tr>
<tr>
<td>Cellular macromolecule metabolic process</td>
<td>GO:0044260</td>
<td>0.018</td>
</tr>
<tr>
<td>Cellular metabolic process</td>
<td>GO:0044237</td>
<td>0.008</td>
</tr>
<tr>
<td>Peptide biosynthetic process</td>
<td>GO:0043043</td>
<td>0.013</td>
</tr>
<tr>
<td>Amide biosynthetic process</td>
<td>GO:0043604</td>
<td>0.016</td>
</tr>
<tr>
<td>Cellular amide metabolic process</td>
<td>GO:0043603</td>
<td>0.031</td>
</tr>
<tr>
<td>Cellular nitrogen compound biosynthetic process</td>
<td>GO:0044271</td>
<td>0.049</td>
</tr>
<tr>
<td>Cellular biosynthetic process</td>
<td>GO:0044249</td>
<td>0.030</td>
</tr>
<tr>
<td>Biosynthetic process</td>
<td>GO:0009058</td>
<td>0.046</td>
</tr>
<tr>
<td>Organonitrogen compound biosynthetic process</td>
<td>GO:1901566</td>
<td>0.012</td>
</tr>
<tr>
<td>Organic substance biosynthetic process</td>
<td>GO:1901576</td>
<td>0.027</td>
</tr>
<tr>
<td>Peptide metabolic process</td>
<td>GO:0006518</td>
<td>0.020</td>
</tr>
<tr>
<td>Cellular macromolecule biosynthetic process</td>
<td>GO:0034645</td>
<td>0.017</td>
</tr>
<tr>
<td>Macromolecule biosynthetic process</td>
<td>GO:0009059</td>
<td>0.019</td>
</tr>
</tbody>
</table>
5.4.5 Laboratory crosses

All crosses produced settlers by 24 hours irrespective of treatment and cross (Figure 5.9A). There was no interaction effect between treatment and cross ($F_{3,39} = 0.340, P = 0.797$), effect of treatment ($F_{1,39} = 0.602, P = 0.442$), or effect of cross on settlement success. At 72 hours and 120 hours, the number of settlers decreased significantly for all crosses and treatments (Figures 5.9B and C).

Regardless of treatment, no pre-metamorphs developed by 24 hours (Figure 5.9D). The maximum number of pre-metamorphs for all crosses was observed at 72 hours (Figure 5.9E), where there was a significant interaction effect between cross and treatment ($F_{3,39} = 3.886, P = 0.016$). At 120 hours, the number of pre-metamorphs had decreased, with no statistical difference between any cross or treatment (Figure 5.9F). For all crosses the number of post-metamorphs was greater at 120 hours than 72 hours (Figures 5.9H and I), suggesting that pre-metamorphs were still undergoing metamorphosis after 72 hours.
Figure 5.9. The percentage of larvae that have undergone development to (A) settler stage, (B) pre-metamorph stage, and (C) post-metamorph stage over 24, 72, and 120 hours. Error bars represent standard error. The legend symbol ♀ represents the egg donor and ♂ represents the sperm donor. Letters on top of the bars indicate significant differences ascertained by Tukey’s post-hoc test.

Our GLMs showed an interaction effect between cross and treatment on post-metamorph development after 72 hours ($F_{3,39} = 3.344$, $P = 0.024$). There was no interaction effect between cross and treatment on post-metamorph development after 120 hours ($F_{3,39} = 1.578$, $P = 0.210$). However there was both an effect of treatment ($F_{1,39} = 5.750$, $P = 0.021$) and cross ($F_{3,39} = 7.659$, $P < 0.001$) individually. Indeed, crosses involving shallow eggs were significantly lower in post-metamorph success under shallow conditions. Interestingly, under shallow conditions, the deep x deep cross was significantly more successful in developing to the post-metamorph stage than any of the crosses involving shallow eggs (Figure 5.9I). Overall, we found that crosses involving deep
eggs were similarly successful at all developmental stages regardless of treatment, whereas crosses involving shallow eggs developed poorly under deep conditions.

5.5 Discussion

Areas of steep environmental gradients over small spatial scales provide unique opportunities to study the role of local adaptation on population divergence. Whilst one would expect to find signatures of local adaptation to features of the horizontal-plane, such as adaptation to temperature associated with latitude (Kuo & Sanford, 2009) or variation of salinity among seas (DeFaveri & Merilä, 2014), the vertical plane represents an additional important source of environmental variation that should be considered (e.g. Morales et al., 2019). The capacity to adapt to steep environmental gradients is especially important for sessile animal species in which dispersal depends on propagules that remain in the water column for limited periods of time. As such, vertical transportation over 10s of metres may introduce propagules to vastly differing environments (Rasmussen & Richardson, 1989; Andersen & Nielsen, 2002; Cowles, 2016). It may therefore be expected that sessile organisms with limited dispersal capabilities exhibit strong signatures of local adaptation when faced with intense selective forces at a small scale (Westram et al., 2018; Morales et al., 2019). There is a dearth of genomic studies assessing the presence of depth-associated local adaptation in the sea, as most studies looking at the exceptionally steep gradients in intertidal systems use small numbers of loci or involve transplant experiments (Hays, 2007; Johnson & Black, 2008). By using an extensive dataset of genomic markers and identifying outlier loci correlated with contrasting environmental conditions, we provide evidence of local adaptation across different depths, suggesting that local adaptation across depths is a key mechanism promoting genomic divergence in aquatic ecosystems.

5.5.1 Population genomic data and differences between depths

For both outlier loci datasets (i.e. the pcadapt and bayenv datasets), we identified higher global $F_{ST}$ than in the neutral dataset and, as expected, highly significant pairwise $F_{ST}$ comparisons. The use of outlier SNPs compared to the neutral dataset however did not result in major differences in the groupings of populations along the eastern North Sea. For both neutral and outlier data, the main source of genetic divergence was between shallow and deep sites, and we also observed more genetic differentiation among deep sites than among shallow sites. Whilst outlier loci have unravelled previously cryptic divergence within populations (Moura et al., 2014; Yang et al., 2016; Segovia et al., 2017; Sandoval-Castillo et al., 2018), we found few differences in population structure inferred by neutral and outlier loci, congruent with previous research (Batista et al., 2016; Bongaerts et al., 2017). The observed similarity in results between neutral and outlier loci
may be seen as evidence for “genome-wide resistance” to gene flow between populations as a result of depth-associated selection (Bongaerts et al., 2017), or the result of the combination of neutral forces together with adaptive forces. An additional explanation for the similarity between our datasets could be that our identified outlier loci represent ‘false positives’ and that the strong genomic differentiation between shallow and deep individuals is due to neutral processes such as genetic drift, the signal of which would be amplified in the outlier loci dataset. However, as the proportion of outlier loci compared to total GBS loci (8.5%) is comparable to other studies (Forsström et al., 2017; Ruiz Daniels et al., 2019), it is unlikely that we sampled an unusually high proportion of false positives.

Other evolutionary forces, such as gene flow, may act against selection, and thus levels of local adaptation may depend on the balance between gene flow and selection (Savolainen et al., 2013). In contrast, gene flow may favour local adaptation by augmenting standing genetic variation on which selection can act on within populations (Tigano & Friesen, 2016). Although we found evidence of limited contemporary gene flow between shallow and deep individuals (Figure 5.6), our hybrid assignment tests indicated ongoing gene flow in sites where shallow and deep individuals are found nearby (BUH and JAM_S). Shallow and deep populations of C. intestinalis have previously undergone long periods of isolation and are now in secondary contact (Hudson et al., 2020). Only a few sites in our study contained shallow and deep populations in close proximity, where one would expect gene flow. Future studies should aim to sample more of these potential “hybrid zones”. A similar result was reported between depth-defined populations of the coral Agaricia fragilis, for which despite evidence of introgression between the studied populations, limited migration was observed (Bongaerts et al., 2017). The lack of migration over “ecological time scales” in A. fragilis was proposed as evidence for local adaptation selecting against genotypes from different populations, leading to assortative mating (Bongaerts et al., 2017). An important assumption when suggesting the existence of local adaptation is that abiotic or biotic factors impose selection pressure on migrants within the new habitat, but that these factors do not preclude migration between populations and that successful reproduction between populations is possible under field conditions (see section below).

We recognise, however, that the extent of dispersal across the pycnocline remains unknown. Larvae of colonial ascidians are able to cross haloclines of similar strength to that generally found off the coast of Sweden, but some larvae exhibit passive sinking after crossing (Vázquez & Young, 1996). Other factors may also influence exchange between populations. The larvae of C. intestinalis are known to exhibit both negative geotaxis and negative phototaxis (Rius et al., 2010a), which may promote differential movement of larvae from different depths. Previous genomic work on these populations of C. intestinalis has proposed historic secondary contact.
between shallow and deep lineages (Hudson et al., 2020), suggesting that the pycnocline may not cause complete reproductive isolation. It therefore appears that whilst the pycnocline acts as a general barrier to gene flow, it is not completely impermeable. Indeed, evidence of F2 and backcrossed individuals, combined with the knowledge that there are 1-2 generations of C. intestinalis per year with individuals surviving for 1-2 years, suggests that infrequent oceanographic events (i.e. breakdown of the pycnocline during storms) may promote cases of gene flow between shallow and deep populations if such a breakdown of the pycnocline overlaps with spawning of C. intestinalis. Our hybrid assignment results may therefore be a signature of such an event that occurred two or more generations ago.

Interestingly, our DAPC analysis consistently found the deep site GUL to be separate from all other sites. Johannesson et al. (2018) sampled a shallow site within the same fjord as GUL, and found similar results, with this site being genetically different from the other coastal shallow site sampled. Fjords provide unique conditions compared to open coast environments, with limited water exchange between the two (Howe et al., 2010). Indeed, local adaptation to fjord environments has been reported in cod (Barth et al., 2017). It is therefore unsurprising that we observed genomic differentiation between our deep fjord site and deep coastal sites. Interestingly, our DAPC analysis using the neutral and pcadapt datasets found the deep sites VAT and GUL (Table 5.1) to be separate from each other, and from JAM_D and KAV, two sites in close geographic proximity, whereas the bayenv dataset found less genomic differentiation amongst deep coastal sites than the neutral markers. This suggests that, whilst gene flow between the northern deep site (VAT) and more southerly deep sites (JAM_D and KAV) is limited (as inferred by neutral loci), ‘deep’ environmental conditions affect similar regions of the genome regardless of geography.

5.5.2 Laboratory crosses

Our study showed successful fertilisation and development of early life history stages in crosses between deep and shallow individuals. We identified asymmetrical performance between depths, with crosses involving shallow eggs performing poorer under deep conditions than shallow conditions (Figure 5.9). This is in contrast with our expectations that shallow individuals would survive a broader range of environmental conditions due to the seasonal variation experienced in the study site, and deep individuals would perform poorly in shallow conditions. In contrast, crosses involving deep eggs developed at similar success under both treatments (Figure 5.9). These general patterns were observed across all early-life history stages (ELHS) assessed. This is in contrast to other studies measuring performance of ascidian ELHS under different conditions, which suggest a developmental bottleneck during ELHS (Hudson et al., in press). However, care
should be taken when interpreting these results: interindividual variability in egg number rendered our experiment heavily unbalanced, reducing the number of replications available for certain crosses and limiting our statistical power (Figure 7.28). In addition, our experimental design did not allow testing the possible inference of phenotypic plasticity, which may shape the tolerance range of *C. intestinalis* (Renborg *et al.*, 2014).

### 5.5.3 Gene annotation and conclusions

We found heterogenous positioning of outlier loci across the genome of *C. intestinalis*, with no outliers present on chromosomes 6 or 13. On chromosomes where outlier loci were found, we identified regions where multiple ‘clustered’ loci were found together. This clustering may be evidence for ‘islands of genomic differentiation’ that may drive divergence between the two lineages. Such regions of heterogenous differentiation across the genome have been attributed to chromosomal rearrangements, inversions or differing recombination rates (Hemmer-Hansen *et al.*, 2013; Wellenreuther & Bernatchez, 2018). It is important to appreciate that the outlier loci we identified are unlikely themselves to be a direct target of natural selection, but rather linked to loci that confer an adaptive advantage (Silliman, 2019). It is for this reason that we searched for genes within a 79kB region surrounding each outlier loci.

The GO terms enriched in the regions flanking our outlier loci mainly corresponded to metabolic processes and plasma membrane function (Table 5.3). The plasma membrane is also plays an important role in salinity tolerance in plants (Janicka-Russak & Kabala, 2015), with short term stress of salinity and temperature known to impair plasma membrane function (Pareek *et al.*, 1997). In addition we found 13 genes associated with translation. Previous work has suggested an effect of temperature and UV damage on nucleic acid integrity (Manova & Gruszka, 2015). It is important to note that the genome used to map our outlier loci was for the sister species *C. robusta*, which was for many years misidentified as *C. intestinalis* (Brunetti *et al.*, 2015). Indeed, this may have downstream effects on our ability to accurately map outlier loci to a genomic position (only 106 of 223 loci identified by Bayenv could be mapped to a chromosome), infer correct genes within the 79 kB window of each mapped outlier, and finally accurately ascertain functional processes for each gene within this window. Furthermore, our value of 39.5 Kb as a 1 cM value was also based on the *C. robusta* genome (Caputi *et al.*, 2008), and it is unknown how reliable this value is for the *C. intestinalis* genome.

By utilising both neutral and outlier genomic loci, we found evidence that local adaptation may be playing a role in maintaining the depth-defined divergence identified in *C. intestinalis* populations. Genomic divergence builds up over periods of isolation, and once populations undergo gene flow due to secondary contact in contact zones, a combination of divergent selection and species-
specific restricted dispersal contribute to continued genetic separation. Our study emphasises the need for incorporating depth in genomic studies of local adaption and provides key insights into how future directions should incorporate these.

5.6 Acknowledgements

We would like to thank the divers at Tjärnö Marine Laboratory and Kristineberg Marine Station, and Dr Ellen Schagerström, for their help in collecting samples. We thank Per Jonsson for his assistance in extracting environmental data. This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation (grant number 64801). We are grateful to the KVA fund for internationalisation and scientific renewal at the Sven Lovén Centre for Marine Sciences for awarding a grant to MR that supported JH stay at the Tjärnö Marine Laboratory.
Chapter 6  Conclusions and synthesis

A major consequence of global change is the redistribution of biodiversity. Whilst historic movements of species have led to range expansions and contractions over geological time scales, contemporary drivers associated to human activities such as artificial transport and CCC are responsible for the relocation of species at an unprecedented rate (Carlton, 1999; Ricciardi, 2007). This thesis aimed to advance knowledge of the drivers determining the distribution of marine biodiversity under such global change, and outlined five key objectives which I briefly address below.

1. To elucidate the genomic patterns underlying the distribution of widespread ascidian species throughout their native and introduced ranges

The widespread ascidian species studied in this thesis show similar genomic diversity within both the native and introduced ranges. Additionally, the native ranges of these species are highly structured with multiple lineages present.

2. To identify evolutionary mechanisms that may have led to the genomic patterns identified in objective 1.

Secondary contacts leading to admixture between divergent populations within the native range can not only produce novel genotypes bringing about additional population divergence, but also potentially be a driver for transoceanic distribution with the assistance of human-mediated transport. In contrast, other widespread species do not display a signature of admixture amongst divergent native populations. In these cases, high propagule pressure owing to multiple introductions from closely related source populations can explain the observed high genomic diversity with the native diversity.

3. To examine the role that environmental-matching can play in determining successful and failed introduction events.

Comparisons between environmental conditions of native and introduced ranges can both offer an explanation for failed introductions, and also suggest future potential pathways of further species spread from localised introduced areas. When combined with high resolution genomic data, they offer a powerful tool in elucidating the drivers behind species ranges.

4. To determine whether hybridisation may promote range expansion under conditions relevant to climate change.
For species whose ranges overlap and can hybridise, interspecific offspring display successful development under a narrower set of temperatures than parental crosses. Furthermore, hybrids do not show superior developmental success than parental crosses under temperatures predicted with CCC. Together, CCC may temper the ability for hybrids to alter their range to areas of novel environmental conditions.

5. To use further our understanding of the evolutionary forces in play during periods of secondary contacts.

Local adaptation to steep environmental clines plays an important role in maintaining population divergence that has built up during periods of isolation, despite evidence of gene flow through controlled crosses and genomic analyses.

As outlined throughout this thesis, one consequence of the movement of species is an increased opportunity for divergent genotypes to come into contact. This “secondary contact” can lead to hybridisation when pre- and post-zygotic reproductive barriers are overcome. Hybridisation has been shown to facilitate further range expansions and invasions, through contemporary gene flow or admixture (Rius & Darling, 2014). Until now, the role that historic secondary contacts and admixture plays in determining current patterns of biodiversity and population structure of marine species had remained unclear, with studies calling for genome-wide sequencing strategies (Pérez-Portela et al., 2017). I have shown that within native ranges, species can leave behind footprints of complex demographic and evolutionary histories that genomic analyses can detect. Natural events, such as glaciation, can lead to allopatry where pockets of isolated populations become genetically divergent. The removal of geographic barriers (glacial retreat) can promote natural range expansion and facilitate the colonisation of previously uninhabited, or newly formed, areas. I have shown that such expansion can lead to secondary contact amongst isolated lineages, leading to admixture, promoting the formation of novel genotypes and populations. Furthermore, contemporary shipping patterns offer unprecedented opportunity for the reshuffling of genotypes, and further human-mediated secondary contact can promote widespread distributions of species. Previous studies that have assessed a large portion of the native range of widespread invasive species have often been in an attempt to identify a source population through comparing allele frequencies (e.g., Simon-Bouhet et al., 2006; Darling et al., 2008). Whilst my detailed sampling strategy also enabled this, I was also able to further knowledge in invasion biology by providing evidence that extensive sampling of the native range is fundamental in fully clarifying the influence that native genotypes have on the introduced range through historic events.
During this thesis I have presented a multidisciplinary approach to further understand the processes shaping a unique biological invasion. Highly invasive species normally exhibit further spread from their initial point of introduction (Blackburn et al., 2011), however multiple biotic and abiotic factors may limit such further spread in localised invasive species (Kinziger et al., 2011). The strength of these factors, and how they interact remains understudied though. When the native range of species contain high genomic differentiation across populations, multiple introductions from divergent lineages may promote invasion through admixture (Rius & Darling, 2014). I have shown that this is not always the case, with high genomic diversity in the native range relating to the introduction of a single source lineage. Furthermore, I present data suggesting that adaptation to thermal ranges experienced within the native range may explain the failed establishment of one native lineage, whilst also suggesting the potential for secondary expansion of another other lineage. Future research should include controlled crosses between and within populations (Chapter Four) to further elucidate how thermal tolerance may explain current distribution of invasive species. By using a combination of genomic and environmental data, I provide further evidence that unique characteristics of recipient locations of introduced species can buffer further spread (Castilla et al., 2004). This chapter demonstrates the importance for future studies to include both environmental matching and population genomics to enhance our predictions of distributions of invasive species.

There are multiple lines of evidence that CCC has the potential to promote range shifts of species (e.g., Wilson et al., 2016b) and also promote hybridisation (Canestrelli et al., 2017). However, how hybridisation may facilitate range expansions under CCC remains unclear [but see Sánchez-Guillén et al., (2013) for an example in insects]. I have addressed this knowledge gap by performing controlled intra- and interspecific crosses. I was able to assess how temperature affects early life history stages of hybrids and progenitors. Studying multiple early life history stages enables a nuanced approach to species’ tolerances to environmental pressures. For example, one can assess whether a particular developmental stage may preclude survival (Byrne, 2011), or make broader inferences on ELHS tolerance compared to adults (Pineda et al., 2012). To this extent, I found that early stages of development can be more tolerant of environmental stress than later stages, resulting in an “environmental bottleneck”. I concluded that offspring performance under temperature treatments may be a good predictor of range expansions, and that ongoing CCC may inhibit, rather than promote, range expansions by hybrids. This can have important implications for management strategies requiring predictions of which species will expand their range, and to where, in the future. Future studies assessing individual performance to environmental stressors would benefit from selecting individuals from throughout a species range, as individuals
occupying fringe populations, or the range centre, may vary in their physiological performance and response to warming (Donelson et al., 2019).

Understanding the mechanisms maintaining or adjusting genomic divergence is fundamental for clarifying the role secondary contacts can play in reshaping evolutionary trajectories. Whilst I have offered a broad approach to further understand how past and contemporary factors can promote secondary contacts, in turn facilitating the redistribution of species, I have also presented a snapshot of the evolutionary process occurring across the genome during secondary contact. Due to the large population size of marine invertebrates, genetic drift is expected to have a small effect on the genome (Allendorf et al., 2010) and therefore other evolutionary forces may be expected to play a role in divergence (Ravinet et al., 2017). In areas where populations experience differing environmental conditions, divergent selection is expected to act (Abebe et al., 2015) and may result in local adaptation. Previous work on local adaptation has generally concentrated on the horizontal plane, especially in the marine environment, with studies focussing on the distribution of species across a two-dimensional space (e.g. latitude and longitude). However the ‘vertical plane’ of altitude, or depth, should also be an important consideration, as gradients of environmental change are often much steeper vertically than horizontally. In aquatic ecosystems, thermoclines and pycnoclines in the water column influence species distributions and are likely to influence patterns of local adaptation and potential hybridisation. I have shown that when secondary contact occurs in areas of environmental heterogeneity, local adaptation can occur over very short distances. Local adaptation at different depths and associated divergent selection provide mechanisms for maintaining genomic differentiation, despite evidence of gene flow and successful reproduction between individuals in laboratory conditions. This chapter emphasises the need for incorporating the vertical plane in genomic studies of local adaptation.

Together, this thesis contributes to the ever growing literature of the impacts global change is having on the marine biota. This thesis has explored how CCC, range expansions, biological invasions, and hybridisation can affect the distribution of marine biodiversity. Historical and current secondary contacts amongst divergent genotypes within the native range can explain current widespread species distributions (Chapter Two), with evolutionary processes occurring over steep environmental clines maintaining population divergence (Chapter Five). However, this is not always the case, as I have shown dissimilar environmental niches occupied by species may preclude introduction to distant locations, limiting the potential for secondary contacts and hybridisation (Chapter Three). Additionally, the consequences of such secondary contacts may be altered by CCC, as hybrids may perform poorly under conditions predicted with CCC (Chapter Four).
Finally, I propose that future studies should use this thesis as a guide to what is possible to uncover in ecological and evolutionary research. By using knowledge of contemporary demographic patterns of widespread species via population genomics, one can elucidate historic or current evolutionary processes that may contribute to such patterns. In addition, by having a better appreciation of the current range of species, including the distribution of divergent lineages and occurrence of hybrid zones, researchers should (study system dependent) perform controlled crosses to assess environmental stressor, or biotic factors, on individual development or other forms of performance. With the increasing applicability for genomic techniques to study non-model organisms (da Fonseca et al., 2016; Beichman et al., 2018), functional gene annotation can provide a way of identifying genes that play an important role in determining the distribution of marine species, which in combination with data from controlled crosses will enable clearer understanding of biodiversity patterns.
Chapter 7  Appendices

Appendix A  Appendix to Chapter Two

Table 7.1. Pairwise population comparisons of $F_{ST}$ values and associated $P$ values.

|       | VAT   | GUL   | JAM_D  | KAV    | BUH    | KOS    | BRA    | LIN    | JAM_S  | FIS    | DEN    | HPL    | TNQ    | JER    | STM    | CAN_1  | CAN_2  |
|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| VAT   | 0.000 | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| GUL   | 0.166 | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| JAM_D | 0.136 | 0.103 | 0.841  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| KAV   | 0.130 | 0.089 | -0.002 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| BUH   | 0.150 | 0.135 | 0.106  | 0.092  | 0.007  | 0.005  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| KOS   | 0.183 | 0.167 | 0.135  | 0.126  | 0.028  | 0.904  | 0.642  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| BRA   | 0.177 | 0.167 | 0.123  | 0.126  | 0.030  | -0.004 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| LIN   | 0.194 | 0.200 | 0.166  | 0.150  | 0.031  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| POR   | 0.255 | 0.259 | 0.227  | 0.205  | 0.100  | 0.055  | 0.073  | 0.072  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| JAM_S | 0.166 | 0.159 | 0.123  | 0.115  | 0.061  | 0.078  | 0.073  | 0.094  | 0.092  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| FIS   | 0.214 | 0.198 | 0.177  | 0.160  | 0.082  | 0.077  | 0.070  | 0.085  | 0.063  | 0.063  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| DEN   | 0.207 | 0.204 | 0.173  | 0.166  | 0.138  | 0.137  | 0.120  | 0.152  | 0.160  | 0.124  | 0.136  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| HPL   | 0.175 | 0.198 | 0.168  | 0.171  | 0.160  | 0.165  | 0.158  | 0.175  | 0.209  | 0.158  | 0.194  | 0.154  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| TNQ   | 0.149 | 0.183 | 0.147  | 0.144  | 0.129  | 0.147  | 0.142  | 0.159  | 0.172  | 0.116  | 0.158  | 0.130  | 0.060  | 0.654  | 0.031  | 0.000  | 0.000  |
| JER   | 0.158 | 0.197 | 0.165  | 0.160  | 0.149  | 0.156  | 0.150  | 0.170  | 0.189  | 0.138  | 0.178  | 0.140  | 0.059  | 0.001  | 0.000  | 0.000  | 0.000  |
| STM   | 0.145 | 0.192 | 0.137  | 0.141  | 0.136  | 0.153  | 0.144  | 0.163  | 0.185  | 0.133  | 0.170  | 0.139  | 0.049  | 0.016  | 0.025  | 0.000  | 0.000  |
| CAN_1 | 0.147 | 0.148 | 0.108  | 0.105  | 0.063  | 0.066  | 0.064  | 0.074  | 0.127  | 0.090  | 0.110  | 0.095  | 0.073  | 0.081  | 0.067  | 0.000  | 0.000  |
| CAN_2 | 0.162 | 0.192 | 0.138  | 0.140  | 0.082  | 0.055  | 0.064  | 0.067  | 0.126  | 0.107  | 0.124  | 0.119  | 0.088  | 0.091  | 0.090  | 0.033  | 0.000  |

Note: $F_{ST}$ values are below the diagonal and $P$ values above the diagonal. Values in *italics* are significant after Bonferroni correction. Site abbreviations are as in Table 2.1.

Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapt removed (14 loci removed).
Appendix A

Table 7.2. Analysis of Molecular Variance under the assumption of four clusters as identified by ADMIXTURE. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcdadapt removed (14 loci removed).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>3</td>
<td>744.83</td>
<td>1.55 Va</td>
<td>8.14</td>
<td>$F_{CT} = 0.081^{***}$</td>
</tr>
<tr>
<td>Among sites within groups</td>
<td>14</td>
<td>748.36</td>
<td>1.29 Vb</td>
<td>6.78</td>
<td>$F_{SC} = 0.074^{***}$</td>
</tr>
<tr>
<td>Within sites</td>
<td>512</td>
<td>8279.13</td>
<td>16.17 Vc</td>
<td>85.08</td>
<td>$F_{ST} = 0.149^{***}$</td>
</tr>
</tbody>
</table>

Note: Clusters are 1) Deep Sweden, 2) Shallow Sweden (except JAM_S) + CAN_1 + CAN_2, 3) JAM_S + DEN, 4) HPL + TNQ + JER + STM. *** $P < 0.001$.

Table 7.3. Analysis of Molecular Variance under the assumption of three clusters as identified by DAPC. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcdadapt removed (14 loci removed).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>2</td>
<td>628.68</td>
<td>1.66 Va</td>
<td>8.61</td>
<td>$F_{CT} = 0.086^{***}$</td>
</tr>
<tr>
<td>Among sites within groups</td>
<td>15</td>
<td>864.51</td>
<td>1.41 Vb</td>
<td>7.35</td>
<td>$F_{SC} = 0.080^{***}$</td>
</tr>
<tr>
<td>Within sites</td>
<td>512</td>
<td>8279.13</td>
<td>16.17 Vc</td>
<td>84.04</td>
<td>$F_{ST} = 0.160^{***}$</td>
</tr>
</tbody>
</table>

Note: Clusters are 1) Deep Sweden, 2) Shallow Sweden + CAN_1 + CAN_2 + DEN, 3) HPL + TNQ + JER + STM. *** $P < 0.001$. 

120
Table 7.4. Analysis of Molecular Variance under the assumption of two clusters (shallow artificial v shallow natural sites). Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapt removed (14 loci removed).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>1</td>
<td>31.40</td>
<td>-0.16 Va</td>
<td>-0.98</td>
<td>$F_{CT} = -0.010$</td>
</tr>
<tr>
<td>Among sites within groups</td>
<td>5</td>
<td>240.17</td>
<td>1.08 Vb</td>
<td>6.61</td>
<td>$F_{SC} = 0.066^{***}$</td>
</tr>
<tr>
<td>Within populations</td>
<td>410</td>
<td>3150.70</td>
<td>15.37 Vc</td>
<td>94.36</td>
<td>$F_{ST} = 0.056^{***}$</td>
</tr>
</tbody>
</table>
Table 7.5. Prior distribution parameters describing the set of scenarios investigated using DIYABC.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective population size</td>
<td>Uniform</td>
<td>10000000</td>
<td>10000000</td>
</tr>
<tr>
<td>Time of events</td>
<td>Uniform</td>
<td>10</td>
<td>10000000</td>
</tr>
<tr>
<td>Admixture rate</td>
<td>Uniform</td>
<td>0.001</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Note: The time priors were constrained (t1< t2< t3) and included split or admixed events,
N1, N2, N3, N4: effective population size of the DIYABC groupings (see methods and Table 2.1).
Table 7.6. Posterior probabilities and 95% confidence intervals of the sets of DIYABC scenarios. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapt removed (14 loci removed).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Posterior probability</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0000,0.3467</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.0000,0.3467</td>
</tr>
<tr>
<td>3</td>
<td>0.9961</td>
<td>0.8007,0.8990</td>
</tr>
<tr>
<td>4</td>
<td>0.0039</td>
<td>0.0000,0.3496</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0000,0.3467</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0000,0.3467</td>
</tr>
<tr>
<td>7</td>
<td>0.0000</td>
<td>0.0000,0.3477</td>
</tr>
<tr>
<td>Scenario set 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0000,0.0010</td>
</tr>
<tr>
<td>2</td>
<td>0.9984</td>
<td>0.9980,0.9988</td>
</tr>
<tr>
<td>3</td>
<td>0.0008</td>
<td>0.0000,0.2422</td>
</tr>
</tbody>
</table>

| Scenario set 2 | | |
| 1        | 0.0000                | 0.0000,0.0011       |
| 2        | 0.1041                | 0.0953,0.1130       |
| 3        | 0.0531                | 0.0491,0.0571       |
| 4        | 0.0014                | 0.0004,0.0024       |
| 5        | 0.8413                | 0.8319,0.8506       |

Note: Scenarios and scenario sets outlined in Figure 7.3. Scenarios in *italics* had the highest posterior probability without overlapping confidence intervals.
Table 7.7. Model checking in DIYABC for scenario 5. Summary statistic abbreviations include variance of non-zero values for genic diversities (HV1_1), FST distances (FV1_1), Nei’s distances (NV1_1) and admixture estimates (AV1_1). Asterisks represent proportions lower than 1% or greater than 99% (**), or lower than 0.1% or greater than 99.9% (**). Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapT removed (14 loci removed).

<table>
<thead>
<tr>
<th>Summary statistic</th>
<th>Observed value</th>
<th>Proportion (simulated &gt; observed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV1_1_1</td>
<td>0.022</td>
<td>0 (***</td>
</tr>
<tr>
<td>HV1_1_2</td>
<td>0.023</td>
<td>0 (***</td>
</tr>
<tr>
<td>HV1_1_3</td>
<td>0.024</td>
<td>0.002 (**)</td>
</tr>
<tr>
<td>HV1_1_4</td>
<td>0.023</td>
<td>0.170</td>
</tr>
<tr>
<td>FV1_1_1&amp;2</td>
<td>0.008</td>
<td>1 (***</td>
</tr>
<tr>
<td>FV1_1_1&amp;3</td>
<td>0.015</td>
<td>0.662</td>
</tr>
<tr>
<td>FV1_1_1&amp;4</td>
<td>0.015</td>
<td>0.449</td>
</tr>
<tr>
<td>FV1_1_2&amp;3</td>
<td>0.018</td>
<td>0.703</td>
</tr>
<tr>
<td>FV1_1_2&amp;4</td>
<td>0.014</td>
<td>0.333</td>
</tr>
<tr>
<td>FV1_1_3&amp;4</td>
<td>0.016</td>
<td>0.421</td>
</tr>
<tr>
<td>NV1_1_1&amp;2</td>
<td>0.003</td>
<td>0.994 (**</td>
</tr>
<tr>
<td>NV1_1_1&amp;3</td>
<td>0.002</td>
<td>0.500</td>
</tr>
<tr>
<td>NV1_1_1&amp;4</td>
<td>0.005</td>
<td>0.428</td>
</tr>
<tr>
<td>NV1_1_2&amp;3</td>
<td>0.007</td>
<td>0.849</td>
</tr>
<tr>
<td>NV1_1_2&amp;4</td>
<td>0.006</td>
<td>0.333</td>
</tr>
<tr>
<td>NV1_1_3&amp;4</td>
<td>0.006</td>
<td>0.406</td>
</tr>
<tr>
<td>AV1_1_4&amp;1&amp;2</td>
<td>0.107</td>
<td>1 (***</td>
</tr>
<tr>
<td>AV1_1_4&amp;1&amp;3</td>
<td>0.086</td>
<td>0.533</td>
</tr>
<tr>
<td>AV1_1_4&amp;2&amp;3</td>
<td>0.097</td>
<td>0.605</td>
</tr>
</tbody>
</table>
Figure 7.1. Cross validation error estimate from the ADMIXTURE analysis showing $K = 4$ as an appropriate modelling choice. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapt removed (14 loci removed).
Figure 7.2. Prior checking for each DIYABC run. A principal component analysis was performed in the space of summary statistics on 100,000 simulated data sets and the observed data was added on each plane. A: Scenario set 1, B: Scenario set 2, and C: Scenario set 3. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcatadapt removed (14 loci removed).
Appendix A

Figure 7.3. DIYABC scenarios used to assess colonisation history within the northeast Atlantic (A and B) and across the north Atlantic (C). Scenarios in A (scenario set 1) assessed the ancestral populations within the northeast Atlantic and included all possible combinations (six) of an initial divergence between two groups, followed by divergence from the third group, as well as one scenario where the three groups diverged at the same time. As scenario 3 was deemed the most likely (see Results), this was used as the basis for B (scenario set 2), where scenarios included: 1) Initial divergence between EJF and DS, followed by divergence from EJF to form SS; 2) Initial divergence between EJF and DS, followed by secondary contact and admixture between EJF and DS which lead to SS; and 3) Initial divergence between DS and an unsampled population, which them diverged to become EJF, with SS diverging from DS. As scenario 2 was deemed the most likely (see Results), this was used as the basis for C (scenario set 3) which assessed the colonisation history of Canada. Scenario set 3 included CAN diverging from 1) DS; 2) SS; 3) EJF; 4) an admixture event between DS and SS; 5) and admixture event between DS and EJF. Abbreviations are shallow Sweden sites (SS), Canada sites (CAN), the Denmark site (DEN), England, Jersey, and France sites (EJF), and deep Sweden sites (DS).

Figure 7.4. The first two axes of the PCA as part of the model checking function of DIYABC. Small yellow circles represent datasets simulated from priors, large filled yellow circles represent datasets simulated from posteriors and the large purple circle represents the observed dataset. Note that the dataset used is with $F_{ST}$ outlier loci recovered from both BayeScan and pcadapt removed (14 loci removed).
A.1. Genomic analyses excluding all outlier loci

In order to consider all possible $F_{ST}$ outlier loci, we ran all of the genomic analyses described in the main text but this time using a dataset with all $F_{ST}$ outlier loci (77 in total). These are the outlier loci identified by either BayeScan and / or pcdadapt. This leads to a final dataset of 1,590 loci from the 265 individuals sequenced.

We found that values of $F_{IS}$ ranged from 0.033 to 0.140 (Table 7.8). Whilst $F_{IS}$ values are similar to the analyses included in the main text, only four sites showed no signs of deviation from Hardy-Weinberg equilibrium, compared to nine sites when only 14 outlier loci were excluded. Expected heterozygosity ranged from 0.193 to 0.279 (Table 7.8).

When we ran the ADMIXTURE and DAPC analyses, we found very similar patterns using both datasets (Figures 7.5 and 7.6A). However, when we performed a DAPC with a priori population information on the dataset with 77 outlier loci, we found the site in Denmark (DEN) to be separated from all other samples (Figure 7.6B). When this single sampling site was excluded, the pattern results then matched what is reported in Fig. 3B (Figure 7.6C).

Pairwise site comparisons of $F_{ST}$ followed the exact same pattern in both datasets, with 143 out of 153 pairwise site comparisons for $F_{ST}$ being significant (93% of comparisons). In addition, there were also no differences in the results of the AMOVAs.

When we ran the Approximate Bayesian Method in the dataset of 77 outlier loci, the first two scenario sets recovered the same scenarios with the highest support as in the main text (i.e. the analysis including the northeast Atlantic sample sites, original divergence was between deep Sweden and England, Jersey, and France [logistic estimate of posterior probability $P=0.9795$, CI=$0.9720, 0.9870$]; the origin of the shallow Sweden population involved admixture between Deep Sweden and England, Jersey, and France [$P=0.9961$, CI=$0.9953, 0.9969$]). Regarding scenario set 3, we found again evidence for a recent admixture between EJF and SS ($P=0.4487$, CI=$0.4285, 0.4689$) but confidence intervals of this scenario overlapped with those of the scenario presenting Canadian individuals being introduced purely by genotypes from shallow Sweden (i.e. Scenario 2 in Figure 7.3; $P=0.4775$, CI=$0.4587, 0.4964$).
Appendix A

Table 7.8. $F_{IS}$ values (values in italics are statistically significant [P<0.05]) and population mean expected heterozygosity ($H_e$) for all sites using the dataset with all putative $F_{ST}$ outlier loci excluded. Note that the dataset used is with $F_{ST}$ outlier loci recovered from either BayeScan and pcadapt removed (77 loci removed).

<table>
<thead>
<tr>
<th>Country</th>
<th>Site Name</th>
<th>Code</th>
<th>$F_{IS}$</th>
<th>$H_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>Vattenholmen</td>
<td>VAT</td>
<td>0.078</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>Gåseklåvan</td>
<td>GUL</td>
<td>0.116</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>Jämmingarna</td>
<td>JAM_D</td>
<td>0.109</td>
<td>0.279</td>
</tr>
<tr>
<td></td>
<td>Kåvra</td>
<td>KAV</td>
<td>0.127</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td>Burholmen</td>
<td>BUH</td>
<td>0.140</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>South Koster</td>
<td>KOS</td>
<td>0.081</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>Brattholmen</td>
<td>BRA</td>
<td>0.098</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>Lindholmen</td>
<td>LIN</td>
<td>0.083</td>
<td>0.241</td>
</tr>
<tr>
<td></td>
<td>Porsholmen</td>
<td>POR</td>
<td>0.058</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>Jämmingarna</td>
<td>JAM_S</td>
<td>0.033</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>Fiskebäckskil</td>
<td>FIS</td>
<td>0.070</td>
<td>0.236</td>
</tr>
<tr>
<td>Denmark</td>
<td>Limfjord</td>
<td>DEN</td>
<td>0.100</td>
<td>0.215</td>
</tr>
<tr>
<td>England</td>
<td>Hartlepool</td>
<td>HPL</td>
<td>0.085</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>Town Quay</td>
<td>TNQ</td>
<td>0.123</td>
<td>0.225</td>
</tr>
<tr>
<td>Jersey</td>
<td>St. Helier</td>
<td>JER</td>
<td>0.092</td>
<td>0.214</td>
</tr>
<tr>
<td>France</td>
<td>St. Malo</td>
<td>STM</td>
<td>0.120</td>
<td>0.226</td>
</tr>
<tr>
<td>Canada</td>
<td>Yarmouth</td>
<td>YAM</td>
<td>0.105*</td>
<td>0.220*</td>
</tr>
<tr>
<td></td>
<td>Shelburne</td>
<td>SB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brudenell River</td>
<td>BR</td>
<td>0.048§</td>
<td>0.255§</td>
</tr>
<tr>
<td></td>
<td>Sydney</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * refers to samples merged and known as CAN_1 (Yarmouth and Shelburne), § refers to merged samples known as CAN_2 (Brudenell River and Sydney).
Figure 7.5. ADMIXTURE plots representing all sampled populations of *Ciona intestinalis*, using the dataset with all F$_{ST}$ outlier recovered from either BayeScan and pcadapt removed (77 loci removed). The main regions are highlighted above. The different colours represent putative genetic clusters with $K$ ranging from 2 to 7, with $K=4$ being found to be the most optimal value.
Figure 7.6. Discriminant Analysis of Principal Components with all putative F_{ST} outlier loci (77) excluded. (A) Discriminant Analysis of Principal Components using unlinked loci with no a priori population information. (B) Discriminant Analysis of Principal Components using unlinked loci with a priori population information. (C) Discriminant Analysis of Principal Components using unlinked loci with a priori population information excluding site DEN. Site abbreviation as in Table 1. Sites in (A) are assigned to clusters as follows; Cluster 1: FIS, KOS, BRA, LIN, POR, BUH, JAM_S, DEN, CAN_1, CAN_2, and eight individuals from BUH; Cluster 2; JER, TNQ, HPL, STM ;Cluster 3: VAT, JAM_D, GUL, KAV, and eight individuals from BUH.
Appendix B  Appendix to Chapter Three

Figure 7.7. DIYABC scenarios used to assess the introduction of *Pyura praeputialis* from Australia to Chile.

Figure 7.8. DIYABC scenarios used to assess the introduction of *Pyura praeputialis* from Eastern Australia to Chile.
Figure 7.9. Heatmap of $F_{ST}$ values with significant pairwise comparisons (after Bonferroni correction for multiple comparisons) symbolised with an asterisk. Sampling sites are abbreviated as in Table 3.1.
Figure 7.10. Discriminant Analysis of Principal Components with population information included as priors performed on the (A) bayenv and (B) pcadapt datasets, excluding individuals from Antofagasta. Sites in red represent individuals collected from eastern Australia, and sites in green represent individuals collected from south-eastern Australia. Site codes as in Table 3.1.
Figure 7.11. Cross-validation plots produced from ADMIXTURE for the (A) neutral, (B) bayenv, and (C) pcadapt dataset.
Figure 7.12. Admixture plots depicting the inferred ancestry of each *Pyura praeputialis* individual to pre-defined genetic clusters from the (A) neutral dataset, (B) bayenv dataset, and (C) pcadapt dataset. Each individual is presented by a vertical bar. Abbreviations as in Table 3.1.
Figure 7.13. Discriminant Analysis of Principal Components with population information included as priors. Plots (A-C) represent all sample sites, (D-F) represent just eastern Australia and Antofagasta, and (G-I) represent just Antofagasta. Plots (A,D,G) are from the neutral dataset, plots B,E,H are from the bayenv dataset, and plots C,F,I are from the pcadapt dataset. Sites in orange represent individuals collected from Antofagasta Bay, sites in purple represent individuals collected from eastern Australia, and sites in green represent individuals collected from south-eastern Australia. Site codes as in Table 3.1.
Figure 7.14. Prior checking for each DIYABC scenario. A principal component analysis was performed in the space of summary statistics on 100,000 simulated data sets and the observed data was added on each plane.

Figure 7.15. Prior checking for each DIYABC scenario. A principal component analysis was performed in the space of summary statistics on 100,000 simulated data sets and the observed data was added on each plane.
Figure 7.16. Comparison of the posterior probabilities of five scenarios simulated in DIYABC pertaining to the number of individuals originating from eastern Australia that were introduced to Antofagasta, Chile. Error bars are 95% confidence intervals.
Figure 7.17. Average yearly sea surface temperature (SST) over a fifteen-year period (2003-2017) of different sites within Australia and Chile. The south-eastern Australia line represents the average SST of the sites A4 – A7.
Figure 7.18. Daily differences in sea surface temperature (SST) obtained from the satellite derived GHRSSST dataset (A Group for High Resolution Sea Surface Temperature; JPL MUR MEaSUREs Project, 2015) between the most northerly (C1) and southerly (C6) sites within Antofagasta Bay. The date with the largest difference in SST (2.48°C) is marked by a red point.

Table 7.9. Number of loci recovered for the different datasets under different ipyrad clustering parameters.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Clustering parameter</th>
<th>Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.95</td>
<td>1146</td>
</tr>
<tr>
<td>All sites</td>
<td>0.9</td>
<td>1229</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>1006</td>
</tr>
</tbody>
</table>
Table 7.10. Locations where sea surface temperature (SST) data were collected along the coast of Chile. Note sites C13 and C14 are within Antofagasta.

<table>
<thead>
<tr>
<th>Code</th>
<th>Site name</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>C7</td>
<td>Port of Pisagua</td>
<td>-19.597</td>
<td>-70.214</td>
</tr>
<tr>
<td>C8</td>
<td>Port of Junin</td>
<td>-19.670</td>
<td>-70.173</td>
</tr>
<tr>
<td>C9</td>
<td>Caleta Buena</td>
<td>-19.888</td>
<td>-70.135</td>
</tr>
<tr>
<td>C10</td>
<td>Port of Iquique</td>
<td>-20.212</td>
<td>-70.157</td>
</tr>
<tr>
<td>C11</td>
<td>Port of Tocopilla</td>
<td>-22.094</td>
<td>-70.210</td>
</tr>
<tr>
<td>C12</td>
<td>Port of Mejillones</td>
<td>-70.416</td>
<td>-23.054</td>
</tr>
<tr>
<td>C13</td>
<td>North Antofagasta</td>
<td>-70.440</td>
<td>-23.536</td>
</tr>
<tr>
<td>C14</td>
<td>South Antofagasta</td>
<td>-70.446</td>
<td>-23.743</td>
</tr>
<tr>
<td>C15</td>
<td>Port of Taltal</td>
<td>-70.493</td>
<td>-25.403</td>
</tr>
</tbody>
</table>
Table 7.11. Analysis of Molecular Variance under the assumption of three geographic clusters.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among groups</td>
<td>2</td>
<td>1655.74</td>
<td>5.39 Va</td>
<td>2.97</td>
<td>$F_{CT} = 0.030^{***}$</td>
</tr>
<tr>
<td>Among sites within groups</td>
<td>10</td>
<td>2218.25</td>
<td>1.91 Vb</td>
<td>1.05</td>
<td>$F_{SC} = 0.011^{***}$</td>
</tr>
<tr>
<td>Within sites</td>
<td>321</td>
<td>55899.37</td>
<td>174.14 Vc</td>
<td>95.97</td>
<td>$F_{ST} = 0.040^{***}$</td>
</tr>
</tbody>
</table>
Table 7.12. Posterior probabilities and 95% confidence intervals of the first set of DIYABC scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Posterior probability</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>1.0000, 1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.0000, 0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.0000, 0.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000, 0.0000</td>
</tr>
</tbody>
</table>

Table 7.13. Posterior probabilities and 95% confidence intervals of the second set of DIYABC scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Posterior probability</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1462</td>
<td>1.0000, 0.5298</td>
</tr>
<tr>
<td>2</td>
<td>0.8538</td>
<td>0.7880, 0.9195</td>
</tr>
<tr>
<td>3</td>
<td>0.0001</td>
<td>0.0000, 0.0764</td>
</tr>
</tbody>
</table>
### Appendix C  Appendix to Chapter Four

Table 7.14. Number of replicates produced for each reciprocal cross (one replicate is a Petri dish containing 20 larvae).

<table>
<thead>
<tr>
<th>Egg donor</th>
<th>Sperm donor</th>
<th>12°C</th>
<th>16°C</th>
<th>20°C</th>
<th>24°C</th>
<th>28°C</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. herdmani</em></td>
<td><em>P. herdmani</em></td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>P. stolonifera</em></td>
<td><em>P. stolonifera</em></td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>P. herdmani</em></td>
<td><em>P. stolonifera</em></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><em>P. stolonifera</em></td>
<td><em>P. herdmani</em></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 7.19. Microscope images of early life history stages that were counted. A) Hatched larvae with unfertilised eggs (ul: unfertilised larvae, vl: viable larvae), B) Settler without tail reabsorption, C) Settler with tail reabsorption, D) Pre-metamorph, E) Post-metamorph.
Figure 7.20. Sea surface temperature at Port Elizabeth between 2003 – 2017. The black line represents the mean annual SST for the region, the red line represents the mean summer SST (January to March), and the blue line represents the mean winter SST (July to September). Data collected from JPL MUR MEaSUREs Project (2015). Error bars denote standard deviation.
## Appendix D  Appendix to Chapter Five

Table 7.15. Matrix of Pearson’s R (below horizontal) and associated P values (above the horizontal) between environmental variables. Emboldened cells represent significant correlation after Bonferroni correction for multiple comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Temp</th>
<th></th>
<th></th>
<th>Salinity</th>
<th></th>
<th></th>
<th>Chlorophyll a</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>summer</td>
<td>winter</td>
<td>annual</td>
<td>summer</td>
<td>winter</td>
<td>annual</td>
<td>summer</td>
<td>winter</td>
<td>annual</td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td><strong>0.0004</strong></td>
<td><strong>0.0001</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0548</td>
<td>0.0014</td>
<td>0.0038</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>-0.88</td>
<td>0.0015</td>
<td></td>
<td><strong>0.0012</strong></td>
<td>0.0016</td>
<td><strong>0.0007</strong></td>
<td>0.0012</td>
<td>0.0022</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>annual</td>
<td>0.91</td>
<td>-0.83</td>
<td></td>
<td>0.0091</td>
<td>0.0075</td>
<td>0.0067</td>
<td>0.0125</td>
<td>0.0737</td>
<td>0.0351</td>
<td></td>
</tr>
<tr>
<td>Salinity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>-0.95</td>
<td>0.84</td>
<td>-0.74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1321</td>
<td>0</td>
<td><strong>0.0009</strong></td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>-0.93</td>
<td>0.83</td>
<td>-0.75</td>
<td><strong>0.95</strong></td>
<td>0</td>
<td>0</td>
<td>0.2109</td>
<td><strong>0.0006</strong></td>
<td>0.0078</td>
<td></td>
</tr>
<tr>
<td>annual</td>
<td>-0.96</td>
<td>0.86</td>
<td>-0.76</td>
<td><strong>0.99</strong></td>
<td><strong>0.98</strong></td>
<td></td>
<td>0.1311</td>
<td>0</td>
<td><strong>0.0013</strong></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>0.59</td>
<td><strong>-0.84</strong></td>
<td>0.72</td>
<td>-0.48</td>
<td>-0.41</td>
<td>-0.48</td>
<td>0.1305</td>
<td>0.0048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>0.83</td>
<td>-0.82</td>
<td>0.56</td>
<td><strong>-0.95</strong></td>
<td><strong>-0.86</strong></td>
<td><strong>-0.93</strong></td>
<td>0.48</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>annual</td>
<td>0.79</td>
<td><strong>-0.93</strong></td>
<td>0.64</td>
<td><strong>-0.85</strong></td>
<td>-0.75</td>
<td><strong>-0.84</strong></td>
<td>0.78</td>
<td></td>
<td><strong>0.92</strong></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.21. Cross-Validation values as inferred by ADMIXTURE at different $K$s using the pcadapt dataset. The lowest CV value is found at $K = 8$. 
Figure 7.22. Cross-Validation values as inferred by ADMIXTURE at different $K$s using the bayenv dataset. The lowest CV value is found at $K = 8$.

**D.1. Fertilisation success at different sperm concentrations**

To assess whether there was an optimum concentration of sperm to use during our fertilisation experiments, we pooled together the sperm from four shallow individuals and performed ten-fold sperm dilutions starting from dry sperm (ca. $10^5$ cells $\mu$L$^{-1}$) down to ca. 1 cell $\mu$L$^{-1}$ (as counted using a haemocytometer). We extracted eggs from four mothers, and aliquoted each set of eggs to one of six Petri dishes (corresponding to the sperm dilution) so that each individual mother was a replicate. We added 40 $\mu$L of sperm to each Petri dish and allowed the eggs to fertilise at 16°C. After 45 minutes we continuously took images of each replicate using an Olympus E5 DSLR camera with a 50 mm F2 macro lens until two hours after sperm addition. This enabled each Petri dish to be photographed five times during this time period.

We found no obvious change in fertilisation success under different sperm concentrations (Figure 7.23), and therefore we used a concentration ca. $10^4$ sperm $\mu$L$^{-1}$ in our laboratory crosses.
Figure 7.23. Fertilisation curve for different sperm concentrations at shallow water salinity and 16°C. Error bars represent standard error. No error bars are available for maximum sperm concentration as only enough dry sperm was available for one replicate.

D.2. Genomic structure inferred by neutral loci

The DAPC analysis using the “neutral dataset” found clear differentiation between shallow and deep sites (Figure 7.24). Interestingly, the shallow site JAM_S clustered closely with the deep fjordal site GUL, despite JAM_S being separated only by ca. 15m to the deep site JAM_D. The northern coastal deep site VAT was also distinct from the more southerly coastal deep sites JAM_D and KAV. The cross-validation method in ADMIXTURE identified the best supported value of K to be 2 (Figure 7.25). This found clear differentiation between the deep and most shallow sites, with half of the individuals from BUH and the individuals from JAM_S exhibiting a mixed genetic background (Figure 7.26). Global $F_{ST}$ for the neutral dataset was 0.101. All pairwise $F_{ST}$ population comparisons using neutral loci were significant, however a clear pattern still emerged with comparisons between shallow and deep having larger $F_{ST}$ values (Figure 7.27). The lowest pairwise $F_{ST}$ values were between shallow sites in the north of Sweden, and also one comparison between two deep-water sites (KAV vs JAM_D).
Figure 7.24. Discriminant analysis of principal components on the neutral dataset. Orange colours represent deep populations and magenta colours represent shallow populations. Sites are abbreviated as in Table 5.1.
Figure 7.25. Cross-Validation values as inferred by ADMIXTURE at different $K$s using the neutral dataset. The lowest CV value is found at $K = 2$.

Figure 7.26. ADMIXTURE plots of $K = 2$-4 using the neutral dataset showing depth-defined divergence. The cross-validation method within ADMIXTURE found the best supported value of $K$ to be 2. Sites are abbreviated as in Table 5.1.
Figure 7.27. $F_{ST}$ heatmap for the neutral dataset. Significant pairwise comparisons [after Benjamini-Yekutieli correction (Yekutieli & Benjamini, 1999)] are symbolised by asterisks. Sites are abbreviated as in Table 5.1.
Figure 7.28. Replicate of Figure 5.9 with raw data points rather than bar plots to demonstrate the unbalanced nature of our experimental crosses. Note that D x D and D x S crosses (yellow and light green) only have 2 replicate values, whereas S x D and S x S (blue and dark blue) have 10 replicate values. D = Deep and S = Shallow.
List of References


List of References


List of References


IPCC. 2013. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.


List of References


177


List of References


List of References


List of References


List of References


