The Genomic Basis of Adaptation in Bottlenose Dolphins (genus *Tursiops*)

BY

Eleanor Alice Louise Pratt

BSc (Mar Biol) (Hons)

Thesis submitted to Flinders University for the degree of

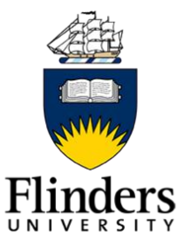
Doctor of Philosophy

Cetacean Ecology, Behaviour and Evolution Laboratory

Molecular Ecology Laboratory

College of Science and Engineering, Flinders University

Adelaide, South Australia



April 2020

Table of Contents

List of Figures 7
List of Tables 10
Summary 11
Declaration 14
Acknowledgments 15
Dedication 16
Chapter 1 : General Introduction 17
1.1 Evolutionary Theory and Natural Selection 18
1.1.1 Selection 18
1.1.2 Mutation 19
1.1.3 Genetic Drift 21
1.2 Macroevolution *vs* Microevolution 22
1.2.1 Macroevolution in Cetacea 23
1.2.1.1 The Eocene Epoch: Transition Back to the Oceans 23
1.2.1.2 The Late Eocene: The Rise of Modern Cetaceans 25
1.2.1.3 The Oligocene Epoch and Beyond: Radiation and Specialisation Driven by Oceanic Changes 27
1.2.2 Microevolution in Cetacea 29
1.2.2.1 Bottlenose Dolphins as a Case Study 31
1.3 Applications of Evolutionary Studies to Conservation Management 38
1.4 Thesis Aims 40
Chapter 2 : Genomic Divergence and Ecotype Formation in the genus *Tursiops* 41
2.1 Contributions 42
2.2 Permits and Ethics Approvals 43
2.3 Abstract 44
2.4 Introduction 45
2.5 Methods 52
2.5.1 Sample Collection 52
2.5.2 Genomic Laboratory Methods 53
2.5.2.1 DNA Extraction 53
2.5.2.2 ddRAD Library Preparation 54
2.5.3 Bioinformatics 55
2.5.4 Genomic Variation 56
2.5.5 Genomic Divergence 56
2.5.5.1 Phylogenomics 56
2.5.5.2 Population Genomic Structure 56

2.5.6 Genomic Basis of Ecotype Formation 57
2.5.6.1 Candidate Loci Detection 57
2.5.6.2 Functional Enrichment Analysis and Annotation 59
2.6 Results 60
2.6.1 Genomic Variation 60
2.6.2 Genomic Divergence 61
2.6.2.1 Phylogenomics 61
2.6.2.2 Population Genomic Structure 63
2.6.3 Genomic Basis of Ecotype Formation 66
2.6.3.1 Candidate Loci Detection 66
2.6.3.2 Functional Enrichment Analysis and Annotation 68
2.7 Discussion 70
2.7.1 Genomic Variation 70
2.7.2 Genomic Divergence 72
2.7.3 Genomic Basis of Ecotype Formation 74
2.7.3.1 Ecotype Evolution 75
2.7.3.2 Cardiovascular and Circulatory Systems 77
2.7.3.3 Gastrointestinal System and Mouth 79
2.7.3.4 Adipogenesis and Energy Production 80
2.7.3.5 Sensory Systems 82
2.7.3.6 Musculoskeletal and Integumentary Systems 84
2.7.3.7 Osmoregulation 86
2.7.3.8 Brain Development and Nervous System 87
2.7.4 Conservation Implications 89
2.7.5 Limitations and Future Directions 91
2.8 Conclusion 92
Chapter 3 : Seascape Genomics of Indo-Pacific Bottlenose Dolphins (*Tursiops aduncus*) in Eastern Australia 93
3.1 Contributions 94
3.2 Permits and Ethics Approvals 95
3.3 Abstract 96
3.4 Introduction 97
3.5 Methods 102
3.5.1 Sample Collection 102
3.5.2 Genomic Laboratory Methods 103
3.5.2.1 DNA Extraction 103
3.5.2.2 ddRAD Library Preparation 104
3.5.3 Bioinformatics 104

3.5.4 Neutral Population Genomic Diversity and Structure 104
3.5.5 Genotype-Environment Association 106
3.5.6 Adaptive Population Genetic Diversity and Structure 108
3.5.7 Functional Enrichment Analysis and Annotation 109
3.5.8 Additional Outlier Tests for Selection 109
3.6 Results 111
3.6.1 Neutral Population Genomic Diversity and Structure 111
3.6.2 Genotype-Environment Associations 116
3.6.3 Adaptive Population Genetic Diversity and Structure 118
3.6.4 Functional Enrichment Analysis and Annotation 119
3.7 Discussion 121
3.7.1 Genomic Variation 121
3.7.2 Population Structure and Adaptation 122
3.7.2.1 Neutral Differentiation 123
3.7.2.1.1 Spatial Influence of the East Australian Current 123
3.7.2.1.2 Broad-scale Habitat Type Change 125
3.7.2.2 Adaptive Differentiation 127
3.7.2.2.1 Open-coast Population Divergence 127
3.7.2.2.2 Port Stephens Embayment Divergence 130
3.7.3 Conservation Implications 132
3.8 Conclusion 135
Chapter 4 : Seascape Genomics of Southern Australian Bottlenose Dolphins (*Tursiops* cf. *australis*) 136
4.1 Contributions 137
4.2 Permits and Ethics Approvals 138
4.3 Abstract 139
4.4 Introduction 140
4.5 Methods 146
4.5.1 Sample Collection 146
4.5.2 Genomic Laboratory Methods 147
4.5.2.1 DNA Extraction 147
4.5.2.2 ddRAD Library Preparation 148
4.5.3 Bioinformatics 148
4.5.4 Neutral Population Genomic Diversity and Structure 148
4.5.5 Genotype-Environment Association 150
4.5.6 Adaptive Population Genetic Diversity and Structure 152
4.5.7 Functional Enrichment Analysis and Annotation 152
4.5.8 Additional Tests for Selection 153

4.6 Results ... 155
4.6.1 Neutral Population Genomic Diversity and Structure ... 155
4.6.2 Genotype-Environment Association ... 160
4.6.3 Adaptive Population Genetic Diversity and Structure ... 162
4.6.4 Functional Enrichment Analysis and Annotation ... 164
4.7 Discussion ... 167
4.7.1 Genomic Variation ... 168
4.7.2 Environmental Drivers of Genomic Differentiation ... 169
4.7.2.1 Coffin Bay and the Great Australian Bight ... 170
4.7.2.2 Gulf Localities ... 173
4.7.2.2.1 Spencer Gulf ... 175
4.7.2.2.2 Gulf St. Vincent ... 179
4.7.3 Conservation Implications ... 181
4.8 Conclusion ... 184
Chapter 5 : General Discussion ... 185
5.1 Introduction ... 186
5.2 *Tursiops* Genomic Divergence ... 187
5.3 Genomic Basis of Environmental Adaptation ... 190
5.3.1 Osmoregulation and Salinity Tolerance ... 195
5.3.2 Cardiovascular System ... 195
5.3.3 Energy Production and Thermoregulation ... 196
5.3.4 Musculoskeletal System ... 198
5.3.5 Sensory System ... 199
5.3.6 Brain and Nervous System ... 200
5.3.7 Differences and Parallelism in Adaptive Responses of Bottlenose Dolphin Inshore Lineages ... 201
5.4 Contribution to Science ... 202
5.5 Limitations and Future Directions ... 204
Conclusion ... 206
Reference List ... 208
Appendices ... 239
A. Abbreviations ... 239
B. Chapter 2 Appendix ... 241
B.i. Figures ... 241
B.ii. Tables ... 243
B.iii. Methods ... 269
C. Chapter 3 Appendix ... 292
C.i. Figures ... 293

C.ii. Tables ... 297
C.iii. Methods ... 306
D. Chapter 4 Appendix ... 308
D.i. Figures ... 308
D.ii. Tables ... 315
D.iii. Methods ... 335
E. Supplementary References ... 337

List of Figures

Figure 2.1 Sampling locations of *Tursiops* spp. across three ocean basins in the Southern Hemisphere. 52

Figure 2.2 *RAxML* maximum-likelihood tree with 1,000 RELL bootstraps based on 18,282 SNPs, displaying phylogenomic relationships among *Tursiops* spp. across the Southern Hemisphere and including nine common dolphins (*Delphinus delphis*) as outgroup. Population-level spatial differentiation is also shown, abbreviations are explained in Appendix B: Table B.ii.1 62

Figure 2.3 Population genomic structure of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere based on 18,060 SNPs as estimated by **A)** principal component analysis (PCA) and **B)** discriminant analysis of principal components (DAPC) with four clusters being the most supported number of 'populations'. Sampling locations are coloured as per putative lineage: *T. aduncus* (blue shades), SABD (green shades), *T. t. gephyreus* (purple shades), *T. t. truncatus* (red shades). Sampling location abbreviations are explained in Appendix B: Table B.ii.1. 65

Figure 2.4 Heat map of pairwise genomic differentiation (F_{ST}) between sampling sites of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere as estimated by *Arlequin* based on 18,060 SNPs. Values on the diagonal represent the average F_{ST} value for comparisons within each putative lineage, while those in the top half of the matrix represent the average value of pairwise comparisons between each lineage. Non-significant F_{ST} values at the B-Y corrected alpha value of 0.0076 are marked with a black square (▪). Transitions between putative lineages are marked by black lines. Average F_{ST} was 0.3604. Sampling location abbreviations are explained in Appendix B: Table B.ii.1. 66

Figure 2.5 Genotype distribution for the parallel evolution candidate genes with F_{ST} values in the top 10% when comparing each inshore bottlenose dolphin lineage (*Tursiops* spp.) to offshore *T. t. truncatus*. 67

Figure 2.6 Genotype distribution for the six candidate genes implicated in early-stage evolution between offshore *T. t. truncatus* and inshore *T. t. gephyreus* in the southwestern Atlantic Ocean 68

Figure 3.1 Indo-Pacific bottlenose dolphin (*T. aduncus*) sampling sites along the New South Wales (NSW) coast in eastern Australia. N = sample size before filtering. 103

Figure 3.2 Population genomic structure of Indo-Pacific bottlenose dolphins along the east coast of Australia based on 110 individuals and **A)** 14,143 SNPs in the neutral dataset and **B)** 281 outlier SNPs in the adaptive dataset. **i)** *Admixture* plots whereby each column represents an individual dolphin and the proportion of a given colour represents the percentage probability of that dolphin belonging to a given population. Vertical black lines mark the start of a new sampling location. Population structure is shown at **1)** *K* = 2 populations (neutral = highly supported; adaptive = low support); **2)** *K* = 3 (neutral = most supported; adaptive = highly supported); and **3)** *K* = 4 (neutral = highly supported; adaptive = most supported). **ii)** Principal

component analysis (PCA); and **iii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from north (left/top) to south (right/bottom) and colours correspond to neutral population membership. Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3)... 113

Figure 3.3 Heat map of pairwise genomic differentiation (F_{ST}) between sampling sites of Indo-Pacific bottlenose dolphins along the east coast of Australia (n=110 individuals) as estimated by *Arlequin*. Estimates based on the adaptive dataset (281 SNPs) are found in the top half of the matrix, while estimates based on the neutral dataset (14,143 SNPs) are in the bottom half. Comparisons non-significant at the B-Y corrected alpha value 0.0120 are marked with a black square (▪). Average F_{ST} for the neutral dataset was 0.051 and for the adaptive dataset was 0.099. Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3). 115

Figure 3.4 Mantel test for analysis of isolation by distance (IBD) in Indo-Pacific bottlenose dolphins from nine sampling localities along the east coast of Australia. Correlation between neutral (crosses) and adaptive (blue circles) genomic distance (linearised F_{ST}) and geographical distance along the coast (kms) among sampling locations is displayed. 116

Figure 3.5 Redundancy analysis (RDA) in genotype-environment association (GEA) analysis, displaying the influence of the four retained ecological variables on individual genomic variance coloured by sampling locality. Overall variance in the genomic dataset explained by the model was 215.54 (p = 0.001) (residual variance = 3,839.07), with 9.55% explained by space and 4.81% explained by the four ecological variables (see Appendix C: Table C.ii.5 for more information). Table inset details the variance explained, significance (p) and number of candidate loci most highly correlated with each of the four retained variables. Details of the ecological variables used are provided in Appendix C: Table C.ii.2 (SST = sea surface temperature, PP = primary productivity, ChloA = Chlorophyll A concentration). Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3) and are ordered from north (top) to south (bottom) in the legend. Colours correspond to neutral population membership... 117

Figure 3.6 Allele frequency change for candidate gene *ELOVL2* over the primary productivity (PP) range gradient across the New South Wales (NSW) seascape. Inset A shows a closer view of the gradient across the Hunter Region. Sampling locality abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3)... 120

Figure 4.1 Sampling sites of southern Australian coastal bottlenose dolphins. From west to east sampling locality abbreviations are: ALB (Albany), ESP (Esperance), SFI (St. Francis Island), CBO (outer Coffin Bay), CBI (inner Coffin Bay), PL (Port Lincoln), NSG (Northern Spencer Gulf), SESG (Southeast Spencer Gulf), SB (Stansbury), PW (Port Wakefield), ADE (Adelaide) and CJ (Cape Jervis). Note: N= sample size before filtering. 147

Figure 4.2 Population genomic structure of southern Australian coastal bottlenose dolphins based on 131 individuals and **A)** 7,817 SNPs in the neutral dataset and **B)** 241 outlier SNPs in the adaptive dataset. **i)** *Admixture* plots whereby each column represents an individual dolphin and the proportion of a given colour represents the percentage probability of that dolphin belonging to a given population. Vertical black lines mark the start of a new sampling

location, with the thick line representing the split between sites to the east and west of Eyre Peninsula and the dashed line marking the split between Spencer Gulf and Gulf St. Vincent. Population genomic structure is shown at **1)** *K*=2 populations (neutral = highly supported; adaptive = low support); **2)** *K=3* (neutral = highly supported; adaptive = low support); **3)** *K*=4 (neutral = most supported; adaptive = most supported); and **4)** *K*=5 (neutral = highly supported; adaptive = highly supported). **ii)** Principal component analysis (PCA); and **iii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from west (left/top) to east (right/bottom). Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3). 158

Figure 4.3 Heat map of pairwise genomic differentiation (F_{ST}) between sampling localities of southern Australian coastal bottlenose dolphins based on 131 individuals as estimated by *Arlequin*. F_{ST} estimates based on 241 adaptive SNPs can be found in the top half of the matrix, while estimates based on the 7,817 neutral SNPs are in the bottom half. The black square (▪) denotes the one estimate that was non-significant at the B-Y corrected alpha value 0.0105 across all pairwise comparisons for both datasets. Average F_{ST} for the neutral dataset was 0.072 and for the adaptive dataset was 0.147. Sampling location abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).. 159

Figure 4.4 Mantel test for isolation by distance (IBD) in southern Australian coastal bottlenose dolphins. Correlation between neutral (crosses) and adaptive (blue circles) genomic distance (linearised F_{ST}) and along-shore geographical distance (kms) among sampling localities is displayed.. 160

Figure 4.5 Genotype-environment association (GEA) redundancy analysis (RDA) testing for the association between the five retained ecological variables and individual genomic differentiation coloured by sampling locality. Overall variance in the genomic dataset explained by the model was 186.10 (p = 0.001) (1,702.40 residual variance), with 5.68% explained by space and 9.30% explained by the five ecological variables (see Appendix D: Table D.ii.5 for details). Table inset details the variance explained, significance (p) and number of candidate loci most highly correlated with each of the five retained variables. Additional information of the ecological variables used are provided in Appendix D: Table D.ii.2 (SST = sea surface temperature, CV = current velocity, ChloA = chlorophyll A concentration; also note min = minimum). Sampling location abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3) and are ordered from west (top) to east (bottom) in the legend. 162

Figure 4.6 Variation in candidate gene *CMKLR1*, showing **A)** allele frequency change over the strong minimum sea surface temperature (SST) gradient across the southern Australian seascape. Blue and red pie sections correspond to the frequency of occurrence at each sampling location of the major and minor alleles, respectively; and **B)** the percentage of southern Australian coastal bottlenose dolphins at each sampling locality found to be homozygotes at either allele. Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3). 166

List of Tables

Table 5.1 Candidate adaptive genes found to be under selection in multiple data chapters of this thesis and/or previously published literature (including members of the same gene family) 192

Summary

The application of ecological genomic techniques to marine biodiversity is becoming increasingly recognised, with a growing number of studies utilising genomics to address population diversification, structure and connectivity in a range of marine species. This has primarily focused on species of importance to recreational and commercial fishing, but is now being expanded to include megafauna, such as marine mammals. While genomic studies of cetacean (whale and dolphin) evolution are relatively prevalent, much of this research concerns the macroevolutionary transition of this lineage from land to the aquatic environment. Microevolutionary genomic differentiation among and within closely related species on the other hand, has only recently begun to be investigated. Bottlenose dolphins (genus *Tursiops*) exhibit repeated inshore and offshore ecotypes around the world, with fine-scale population genetic structure typically found within the inshore lineages. The drivers of ecotype formation and population differentiation have until now not been investigated using genomic techniques and formal testing of genotype-environment associations. This line of study provides an excellent opportunity to better understand the environmental drivers of speciation and differentiation in marine species. This is becoming increasingly important with ongoing anthropogenically-induced climate change rapidly impacting on marine species and altering their ecosystems worldwide.

This study uses genomic datasets to clarify the evolution of bottlenose dolphins at species and population levels. The relationship between divergence and ecological heterogeneity is explored and empirically tested, revealing the genomic basis of potential adaptations to selective pressures and environmental heterogeneity. Briefly, four separate bottlenose dolphin species or subspecies were supported for the Southern Hemisphere, each with unique evolutionary histories shaped by interactions with their respective habitats. This includes the common bottlenose dolphin (*T. t. truncatus*) widely distributed throughout offshore waters of

the Southern Hemisphere, and its recognised subspecies in inshore waters of the southwest Atlantic Ocean (*T. t. gephyreus*). Evidence was also found for genomic divergence between the Indo-Pacific bottlenose dolphin (*T. aduncus*) in eastern Australia and the proposed species (*T. australis*) in coastal southern Australia, suggested here to represent a subspecies of *T. aduncus* (southern Australian bottlenose dolphin, SABD). Genomic differentiation between the inshore and offshore ecotypes revealed adaptations that are potentially most important to early stages of inshore colonisation and provided evidence for parallel evolution in the inshore ecotype. Repeated selection on over one hundred candidate genes across the inshore lineages based on a genomic dataset of over 18,000 loci, revealed potential adaptations of several major bodily systems including the cardiovascular, sensory, musculoskeletal, gastrointestinal, energy production, nervous and osmoregulatory systems. This was hypothesised as a response to divergent selection pressures associated with environmental and ecological disparity, such as differences in depth and prey abundance and distribution between the inshore and offshore habitats.

At a population level, fine-scale neutral population genomic structure was found in bottlenose dolphins along the eastern and southern Australian coasts (*T. aduncus* and SABD, respectively). In both cases, this is likely associated with isolation by distance, strong social structure and natal philopatry. On the other hand, environmental gradients over small geographical distances were empirically shown to shape patterns of adaptive differentiation in these populations. Sea surface temperature and salinity gradients were highly correlated with SABD adaptive differentiation, while heterogeneity in productivity and habitat and oceanographic features were suggested to be most influential on *T. aduncus*. Several genes were found as candidates for driving adaptation of bottlenose dolphins to these particular environmental variables. This includes potential modification of the kidneys and ion transport pathways in response to hypersalinity in South Australia's Spencer Gulf and changes in digestion and metabolism systems to adapt to significant changes in productivity in areas of

the New South Wales coast. Substantial overlap in the bodily systems and specific genes under selection was found among datasets of the three data chapters, suggesting key pathways involved in parallel adaptation of different lineages of inshore bottlenose dolphins. Selection on many of these same pathways and genes have also been discovered in previous studies across several marine taxa, suggesting that they are not only important to environmental adaptation in bottlenose dolphins, but also to other marine species.

This study provides crucial information about drivers of species and population divergence and adaptive evolution in cetaceans. With climate change already causing major restructuring of ecological conditions and species distributions in the world's oceans, having a better understanding of the adaptive capacity of local populations will become increasingly important. Findings of this thesis can therefore, be incorporated into management plans to ensure well-informed and effective conservation strategies to support marine ecosystems into the future.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and

2. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Eleanor Pratt

Acknowledgments

I would first like to thank my supervisors, A/Prof. Luciana Möller and Prof. Luciano Beheregaray. Thank you for sharing your knowledge and experience with me, you have taught me a great deal more than just about dolphin genomics. Thank you also to all the members of CEBEL and MELFU, past and present, for your help and guidance, and also for your friendship.

I wish to acknowledge and thank the Sir Mark Mitchell Research Foundation and the Equity Trustees Holsworth Wildlife Research Endowment for providing monetary support for genomic sequencing for this work. I would also like to thank our international collaborators Gabriela de Tezanos-Pinto, Karen Stockin and Pedro Fruet for the supply of New Zealand and southwestern Atlantic Ocean samples for Chapter 2, without which this study would have been significantly diminished.

Thank you to my dear friends. You have never failed to take my mind off my troubles, allowing me the chance to switch off and relax. You'll never know how important that has been to me. And a special shout out to Emily and Cassie, I couldn't have done this without you.

Finally, I'd like to thank my family. Thank you to Joe and Annie for keeping me sane and giving us beautiful Henri, my little marine biologist protégé. Thank you to Will, who bore the brunt of my worst days, you're not just my brother but one of my best friends. And to my parents, I can't put into words how thankful I am for your love and encouragement. Without you I would not have been able to muster the determination and resilience required to complete this thesis, this is as much your PhD as it is mine.

Dedication

This thesis is dedicated to my late Gran, who is my biggest inspiration, and to my parents, David and Sue, for their unwavering support.

Chapter 1 : General Introduction

1.1 Evolutionary Theory and Natural Selection

Evolution is the process by which all living organisms have developed over many generations (Endler 1986). It is not simply a change in specific individuals, but a process of cumulative changes in an entire species or population toward a form with the most advantageous characteristics for the survival of that species or population as a whole (Wright 1931). Evolution shapes the history of each and every species, allowing populations to continuously adapt to changing conditions or otherwise decline toward extinction. The adaptive capacity of a species or population is largely dictated by intrinsic and extrinsic factors that influence the strength of evolutionary forces (Grummer et al. 2019). The main evolutionary forces are: mutation, genetic drift, selection and the homogenising force of gene flow (Charlesworth et al. 1982).

1.1.1 Selection

Selection is the process whereby particular alleles or genotypes are favourably selected due to being better suited to survival and reproduction in a given environment compared to others (Williams 1996). Selection pressures are constantly changing in the contemporary environment. In the case of a short-term pulse disturbance to the environment, such as a drought event, phenotypic plasticity may allow the population to acclimatise, reducing the risk of local extinction (Bernatchez 2016). In highly variable environments or under long-term climate change however, adapting to the altered environment through genetic changes is more likely to ensure the ongoing survival of a population (Gienapp et al. 2008; Bernatchez 2016). Genetic variation among individuals is likely to significantly increase the adaptive potential of a given population when exposed to a new selective pressure (Lande and Shannon 1996). The presence of genetic variation within a population allows natural selection to preferentially favour individuals with adaptively beneficial genes, making them more likely to pass on the favourable characteristic to the next generation (Endler 1986). This is crucial in order for

populations to adapt to their local environment (Lacy 1987). Adaptation of populations to local environments can have both positive and negative impacts on the likelihood of species persistence. While increased adaptation to a new environment means that the population is more likely to persist and increase in abundance in the short term, the reduced genetic variability within a population as a result of strong directional selection means that there is less standing genetic variation for which natural selection to act upon if the environment changes (Stinchcombe and Hoekstra 2008). Local adaptation of particular populations also results in increased genetic divergence among populations within a species, potentially creating fragmented populations (Stinchcombe and Hoekstra 2008). This however, can lead to an increase in the genetic diversity across a species, which may be beneficial to the persistence of the species as a whole if the populations maintain a degree of gene flow between one another (i.e. do not move to complete reproductive isolation) (Hastings and Harrison 1994). With scientists predicting that ongoing climate change and other forms of human disturbances will significantly alter habitats worldwide (Meehl et al. 2007), this may become increasingly crucial to species persistence (Hoffmann and Sgro 2011).

1.1.2 Mutation

The second evolutionary force is that of random gene mutations that can alter a population's fitness. Mutations are the random occurrence of new alleles as a result of incorrect DNA duplication (Wright 1931). Point mutations, as well as small and large structural variations (e.g. duplications, copy-number variants, insertions, inversions and translocations) are the origin of all genetic variation without which evolution could not occur (Lacy 1987; Whitlock 2000). Most mutations are of a lethal or deleterious nature, but if they are of small effect they can become fixed in a population leading to a slow decline in the fitness of the population over time (Peck 1994). In this way, mutations have the potential to spontaneously change the trajectory of a population. Beneficial mutations do however, occur in populations at a low rate (Whitlock 2000) and can be very important for adapting to a habitat after an environmental change (Peck

1994). In a rapidly changing environment, standing genetic variation (i.e. variation already present in the population), ultimately originating from a long history of mutations and structural variations, becomes particularly important (Lande and Shannon 1996). Adaptation from pre-existing genetic variation is likely to allow the population to adapt to the new conditions much faster than if selection was acting only on beneficial mutations arising by chance (Barrett and Schluter 2008). This is due to alleles that may be beneficial for adaptation already being found and potentially 'pre-tested' in the population, and being typically present at a higher frequency than a new mutation (reviewed in Barrett & Schluter 2007). The process of adaptation from standing variation typically occurs after a change in the environment or the colonisation of a new habitat, whereby an allele already present at a low frequency in the population may become the most adaptively favourable and increase the fitness of individuals and populations with that allele (Whitlock 2000). In this case, the frequency of the beneficial allele in the population will likely increase, potentially to the point of fixation (i.e. all individuals in the population are homozygous for the allele) (Barrett and Schluter 2008). This was demonstrated in threespine sticklebacks (*Gasterosteus aculeatus*), whereby the gene associated with repeated loss of armour in each freshwater population was found to have been present in the population before the initial colonisation of the habitat (Colosimo et al. 2005). The development of similar traits and characteristics in separate populations derived from a relatively recent common ancestor, as exemplified here, is called parallel evolution (Wood et al. 2005). When populations migrate into similar habitats in different regions, they are likely to be exposed to comparable selective pressures and respond with similar adaptations. While phenotypic parallelism does not always reflect genomic parallelism (Christin et al. 2010; Stern 2013), several cases have been reported, including that of threespine sticklebacks (Jones et al. 2012). Genomic parallelism has also been documented in repeated, independent ecotype formation in lake whitefish (*Coregonus* sp.; Rogers and Bernatchez 2007) and European anchovies (*Engraulis encrasicolus*; Le Moan et al. 2016). While this could potentially occur from identical independent mutations, it more commonly occurs through selection on low frequency alleles already present in the population (Stern 2013). This can be particularly

important for species in quickly colonising vacant or underutilised niche space in the wake of major environmental changes (e.g. sea level rise after the last glacial maximum (LGM)). Subsequent mutations in the small founding populations can cause them to rapidly diverge and adapt to their new habitats.

1.1.3 Genetic Drift

Mutations are often lost from a population through the third evolutionary force - genetic drift (Huang et al. 2016). Genetic drift is the process whereby allele frequencies are randomly altered due to gametes being transmitted from one generation to the next containing only half of the alleles present in the parental generation (Ellstrand and Elam 1993). Alleles that are at a low frequency in a population are more likely to be lost than those at a higher frequency and this is amplified for species with a small effective population size (Whitlock 2000; Conner and Hartl 2004). Genetic drift causes genetic variation to decline, particularly in small populations, as certain alleles are lost and others are pushed toward fixation (Luque 2016). This creates an increase in homozygosity (loci for which only one allele is present) in the population (Lacy 1987) and can lead to issues with the development, survival and growth rate of individuals (see Falconer et al. 1996; Allendorf 1986). Loss of genetic variability can also reduce the adaptive potential of populations, particularly in a highly variable environment, and thus reduce the likelihood of long-term persistence and evolution (Lande 1995; but see Bernatchez 2016). Increased gene flow among populations can however, help to reduce the impact of genetic drift and replenish genetic variability (Slatkin 1987). While gene flow may prevent populations from becoming adapted to local conditions, this can be important for introducing new genes that can be beneficial for a population’s survival in a changing environment (Slatkin 1987).

1.2 Macroevolution *vs* Microevolution

Genetic variation not only provides populations with a greater chance of adapting to changing conditions, but also enables species and populations to explore new niche spaces (Reznick and Ricklefs 2009; Agashe and Bolnick 2010). In this way, variability among individuals, populations and species is a key driver of evolution. Major niche changes of taxonomic groups over evolutionary time are examples of macroevolution, defined as the origin of new species and taxonomic divisions above the species level, as well as the evolutionary development of complex adaptations (Reznick and Ricklefs 2009). Macroevolution is concerned with the differential survival of species within clades (Erwin 2000; Jablonski 2000). Microevolutionary adaptations on the other hand, are those that occur at the population or species level (Erwin 2010); for example, the formation of ecotype differences in freshwater lake populations of threespine sticklebacks (see Taylor and McPhail 1999). Microevolution appears to be largely synonymous with the overall process of natural selection (Reznick and Ricklefs 2009), referring to the differential survival and sorting of individuals within populations due to mutation, random genetic drift and selection (Hansen and Martins 1996).

While it can be argued that all taxonomic groups have undergone some form of macroevolutionary transition, this is not clearly reflected in the literature. Prominent examples of macroevolution include the development of flight in bats and birds (e.g. Shen et al. 2012b; Zhang et al. 2014), echolocation in some bat species (e.g. Parker et al. 2013) and the transition from land to sea in cetaceans (whales and dolphins) and other marine mammals (e.g. Thewissen and Bajpai 2001; Rybczynski et al. 2009; McGowen et al. 2014). These major evolutionary transitions often result from large-scale climatic changes causing previously filled niches to become vacant or opening up entirely novel niche space. For example, it is believed that the ability to fly and echolocate in bats evolved as a response to vacant niche space for flying insectivores following the extinction of many insectivorous bird species in the late Cretaceous period (Speakman 2001; also see Simmons 2005). Microevolutionary variation in echolocation and flight strategies then arose as bats evolved to exploit different fine-scale

niche spaces (Simmons 2005). The use of genomics in studying macroevolution allows scientists to identify changes to particular genes that may be responsible for major adaptations in these lineages. For example, the development of echolocation has been implicated with the gene *SLC26A5,* underlying the protein Prestin, in both bats and cetaceans (Li et al. 2010; Liu et al. 2010). Genomic studies of this nature are particularly important for understanding how observable morphological adaptations are associated with the evolution of the underlying genetic make-up. Additional information regarding concurrent environmental changes can also provide useful insight into the types of physical processes that have driven the macroevolution of these lineages.

1.2.1 Macroevolution in Cetacea

1.2.1.1 The Eocene Epoch: Transition Back to the Oceans

The transition of cetaceans from their terrestrial origin to a fully aquatic lifestyle is perhaps the most well documented example of macroevolution (Thewissen and Bajpai 2001; Steeman et al. 2009; Thewissen et al. 2009; Nery et al. 2013b; McGowen et al. 2014). This is for good reason, with a large number and variety of major adaptations making them a particularly interesting model for investigating both macroevolution and microevolution. It is estimated that cetaceans split from their closest living terrestrial relative, the hippopotamus (*Hippopotamus amphibious*), approximately 53 million years ago (MYA), undergoing major morphological and physiological changes to move from a semi- to fully aquatic form (McGowen et al. 2014). While artiodactyls (e.g. pigs, sheep and hippopotamus) and cetaceans have little in common to the naked eye, a number of ancestral features are retained in both groups that indicate shared ancestry. This includes the presence of air-breathing lungs, beginnings of vestigial hindlimbs in cetacean embryos, and retention of facial hairs and/or whiskers during foetal development of some cetaceans despite losing the overall coverage of hair or fur seen in terrestrial mammals (Thewissen et al. 2009). Furthermore, cetaceans and hippopotamus both lack sebaceous glands commonly present in terrestrial mammals and possess the ability to give

birth and nurse underwater, indicating a shared semi-aquatic ancestry (see Meredith et al. 2013). The relationship between artiodactyls and cetaceans is further supported by the discovery of fossils around India and Pakistan that documented the intermediate stages of this transition (Thewissen et al. 2009). The cetacean fossil archive is now one of the most extensive records of macroevolution (Thewissen et al. 2009) and by pairing this with paleoceanographic records and emerging genomic techniques, scientists now have a good understanding of how this evolution has occurred (McGowen et al. 2014).

Despite retaining some ancestral features, the evolution of modern-day cetaceans included extreme morphological, physiological and behavioural changes (Thewissen et al. 2009; McGowen et al. 2014). The most significant part of this transition occurred throughout the Eocene epoch, with the initial split of basal cetaceans from artiodactyls around 53 MYA (McGowen et al. 2014). Early amphibious cetaceans, Pakicetidae, slowly evolved into a more aquatic lifestyle over this period, allowing them to expand their geographical range (Thewissen and Bajpai 2001). Fossils from this basal cetacean group suggest that they lived primarily around India and Pakistan, while more derived forms such as the remingtonocetids (49-43 MYA) can be found throughout coastal south Asia. Fossils of this group show a stark reduction in the limbs and a likely reliance on hearing for underwater hunting (Thewissen and Bajpai 2001). The Protocetidae lineage marked the beginning of the cetacean colonisation of the world's oceans, with huge geographical dispersal moving from the Tethys into the southwestern Pacific and south Atlantic Oceans, approximately 46-39 MYA (Fordyce 1980; Thewissen and Bajpai 2001). Protocetids had limbs that were not weight bearing (Thewissen and Bajpai 2001) and were probably the last major cetacean clade to return to land to breed, with a fully aquatic lifestyle thought to have developed by around 40 MYA (Fordyce 2009; Gingerich et al. 2009).

1.2.1.2 The Late Eocene: The Rise of Modern Cetaceans

The Late Eocene marked an important time in the evolution of cetaceans, with the emergence of Neoceti (crown cetaceans: odontocetes (modern toothed whales and dolphins) and mysticetes (modern baleen whales)) from the archaeocetes (ancient whales) around 35 MYA (Steeman et al. 2009). It is at this time that the basic body plan of modern cetaceans is first fully represented in the fossil record of the neocetes and the extinct sister groups Basilosauridae and Dorudontidae (Thewissen and Bajpai 2001). Coinciding with a period of major ocean cooling in the Late Eocene (Fordyce 1980; Steeman et al. 2009), the Mysticeti and Odontoceti lineages diverged approximately 36 MYA and became increasingly specialised to their respective niches (Fordyce 2009; Steeman et al. 2009). At this time, there was an inactivation of genes associated with the aquatic eye, and tooth and enamel development in the Mysticeti lineage, suggesting that the evolution of baleen in this clade occurred around 36-35 MYA (McGowen et al. 2014). Concurrent inactivation of genes implicated with the aquatic eye also occurred in the odontocetes and was coupled with the loss of the olfactory bulb and positive selection for genes leading to the development of high frequency hearing and echolocation (McGowen et al. 2014; also see Steeman et al. 2009). The progressive evolution of echolocation for hunting in Odontoceti over the next few million years involved a complex combination of genetic and structural changes and likely allowed the toothed whales and dolphins to exploit new niches and food sources (see Fordyce 1980; Steeman et al. 2009). In addition to the mutations in the Prestin protein, a further two genes associated with hearing in mammals were found to be under strong positive selection in dolphins and some echolocating bats (Davies et al. 2012). Coinciding with the development of sophisticated hearing systems in Odontoceti in the Late Eocene, were genetic changes leading to deterioration of the olfactory system, suggesting a decreased reliance on the sense of smell as echolocation developed (Kishida et al. 2015). While modern-day odontocetes have lost their sense of smell (Oelschläger et al. 2010), mysticetes have retained the required structures in a reduced state and may use olfaction to some degree to locate their prey (Thewissen et al. 2011). These highly specialised sensory system adaptations have evolved

gradually over evolutionary time. Studies investigating these adaptations and their genomic basis allow us to gain a better understanding of the complex processes required for them to develop.

While the ability to echolocate was evolving in the odontocetes, deep diving was likely being concurrently developed in both Neoceti clades as they colonised new habitats in the relatively homogenous Late Eocene oceans (Fordyce 1980). This is evidenced by the net surface charge of the protein myoglobin steadily increasing to be significantly higher in cetaceans 36 MYA compared to that recorded 54 MYA, before their colonisation of the marine environment (Mirceta et al. 2013). Myoglobin is a protein that carries and stores oxygen in the muscle cells and can be found in concentrations over 30 times higher in specialised marine mammal divers than in terrestrial mammals (Mirceta et al. 2013). Further changes in myoglobin net surface charge were subsequently recorded within each of the Odontoceti lineages after the divergence of Physeteroidea from Ziphiidae and Delphinoidea, approximately 34 MYA (McGowen et al. 2014). The development of improved deep diving capabilities allowed individuals to take advantage of new prey items, as well as avoid surface predators (see Mirceta et al. 2013). There are however, large energetic and physiological costs involved with deep diving that have driven the evolution of major functional and morphological changes in cetaceans. Hypoxia, caused by the lack of oxygen being delivered around the body, is one of the major issues faced by deep-diving marine mammals. To cope with this, species have also evolved significantly increased blood volume and haemoglobin concentrations (Kooyman and Ponganis 1998; Kooyman 2009). Studies investigating the molecular basis of hypoxia tolerance in cetaceans found positively selected genes associated with myoglobin production, vasoconstriction and oxygen transportation and storage in the blood and muscles (Nery et al. 2013a; Tian et al. 2016), as well as with prevention of cellular damage under hypoxic conditions (Yim et al. 2014). Vasoconstriction is the process in marine mammals whereby blood and oxygen is preferentially delivered to the brain, heart and exercising muscles during

deep dives, leaving nonessential organs to rely on stored oxygen (Mirceta et al. 2013). During this period of worldwide and deeper water colonisation by cetaceans, the respiratory system also underwent substantial structural changes to aid in deep diving, with several positively selected genes related to lung development found in dolphins (Nery et al. 2013b). When subjected to extreme pressure at depth the lungs collapse and cease to exchange gas (Kooyman 2009). This likely drove the development of increased cartilaginous support in cetacean lungs (Kooyman 2009) and the transition into storing oxygen in the blood and muscles in cetaceans, as opposed to the lungs as is typical of terrestrial mammals (Cherniack and Longobardo 1970; Kooyman and Ponganis 1998). These adaptations undoubtedly aided the cetacean lineage's colonisation of oceans worldwide, as well as exploration of relatively unexploited niches at depth.

1.2.1.3 The Oligocene Epoch and Beyond: Radiation and Specialisation Driven by Oceanic Changes

The Eocene-Oligocene boundary saw dramatic changes in the world's oceans, with increased water activity and productivity (Fordyce 1980; Pyenson et al. 2014). With the Australian-Antarctic Tasman seaway already somewhat established, the gradual opening of the Drake Passage between South America and Antarctica sparked major changes in oceanic conditions in the Southern Hemisphere (Fordyce 1980; Steeman et al. 2009). Development of bottom-water circulation in the Southern Ocean had a significant impact on ocean current systems and led to the eventual establishment of the Antarctic Circumpolar Current (ACC) (reviewed in Steeman et al. 2009). The development of the ACC led to a surge in primary productivity throughout the Southern Hemisphere (Berger 2007; Marx and Uhen 2010). This phenomenon increased the amount of food available to cetaceans, enabling them to forage more efficiently and therefore, grow larger and become more abundant and diverse (Fordyce 1980; Berger 2007; Marx and Uhen 2010). The ACC is believed to have driven the final stages of baleen development and subsequent micro-specialisation of the mysticetes, with significant diversification of this lineage throughout the Oligocene (Fordyce 1980; Steeman et al. 2009).

The radiation of odontocetes saw the split of Ziphiidae and Delphinoidea approximately 32 MYA (McGowen et al. 2014), with the four extant Odontoceti clades, Physeteroidea, Ziphiidae, Delphinoidea and Platanistidae, well established by 30 MYA (Steeman et al. 2009). During this period of odontocete diversification, a number of additional adaptations occurred, including further changes to myoglobin net surface charges and echolocation and other sensory abilities (McGowen et al. 2014). This began with a shift away from ancestral dichromatic vision, with inactivating mutations leading to the subsequent loss of colour vision in many cetacean species in the Late Eocene and Early Oligocene (Meredith et al. 2013). These changes are likely an adaptation to the low light conditions in the ocean, in particular to improve vision at depth and at night (Meredith et al. 2013). Additionally, all extant cetaceans lack functional copies of the genes associated with the ability to taste sweet, bitter and umami flavours, thought to be lost through inactivating mutations around 32 MYA (McGowen et al. 2014; Kishida et al. 2015). These dramatic changes to olfaction, gustation and vision in cetaceans highlight the significance of audition for these animals, particularly the development of echolocation to navigate and hunt in Odontoceti.

Compared to the relatively homogenous oceans of the Late Eocene, changes to ocean temperature and circulation patterns led to vast ocean heterogeneity in the Late Oligocene (Fordyce 1980). Through specialisation to emerging niches, cetacean lineages were able to further diversify (Fordyce 1980). Indeed, it was around this time that many extant cetacean lineages came to exist and differentiate, including the modern dolphins (Delphinidae), deep-diving suction-feeding Ziphiidae and major Balaenopteridae groups (Fordyce 1980; Steeman et al. 2009). Further paleoceanic restructuring occurred over the next few million years into the Miocene Epoch, including the closing of a number of seaways, such as the Tethys, Paratethys and Central America Passage (Steeman et al. 2009). The combination of tectonically-driven oceanographic changes and increased adaptation to local resources saw the Delphinoidea superfamily, and in particular delphinids and porpoises, undergo substantial

diversification around 7-6 MYA (LeDuc 2009; Steeman et al. 2009). The Platanistidae clade (river dolphins) is thought to have reinvaded riverine systems during this period of delphinid diversification (Cassens et al. 2000; Pyenson et al. 2014). Several small Odontoceti species went extinct during this time, perhaps due to their inability to compete with better-adapted emerging delphinid species and/or lack of adaptation to changing climatic conditions (Steeman et al. 2009). The aforementioned Platanistidae species on the other hand, survived due to their adaptation to riverine habitats, unexplored by Delphinidae at the time and less affected by climatic changes than the marine system (Cassens et al. 2000). It is believed that all extant genera within Cetacea had appeared by the Early Pleistocene (Fordyce 2009), but further microevolutionary transitions throughout the marine and freshwater environments followed over the next 2 million years and continue to occur today.

1.2.2 Microevolution in Cetacea

Microevolution is widely evident in the cetacean lineage, with the presence of ecotypic division within many species (e.g. bottlenose dolphins, Hoelzel et al. 1998; tucuxi dolphins, Cunha et al. 2005; killer whales, Ford 2009; finless porpoises, Ruan et al. 2015). Ecotypes are populations within a species that have evolved heritable variation in physiology, morphology, behaviour and/or life history due to environmental differences (see Le Moan et al. 2016). Studies investigating the molecular basis of adaptations of this nature within the cetacean lineage are largely absent from the literature, with the exception of killer whales (*Orcinus orca*). Killer whales exhibit complex population structure, both socially and genetically (Ford 2009). Within each geographical region occupied by killer whales a number of ecotypes exist, often in sympatry. These ecotypes differ most prominently in their prey choice and hunting techniques, but also show variability in morphology, pigmentation patterns and social behaviours (Morin et al. 2010). Genetic differentiation also exists between each of the different ecotypes, as well as among geographical locations within each ecotype (Moura et al. 2014). These populations are thought to have evolved in parapatry and/or sympatry through strong

disruptive selection as a result of differential resource use (Moura et al. 2014; Moura et al. 2015; Foote and Morin 2016). Behavioural and cultural differences have however, been suggested to drive the initial divergence of killer whale ecotypes when colonising novel niche space (Foote et al. 2016). Genetic differentiation is then likely reinforced by strong social structure, teaching of learned behaviours and genomic adaptations to dietary preferences and local climatic conditions (Moura et al. 2014; Moura et al. 2015; Foote et al. 2016).

While the population structure, demographic history, sociality and behaviour of killer whales have been extensively investigated, only two studies so far have been published on the genomic basis of the microevolution of the divergent ecotypes (Moura et al. 2014; Foote et al. 2016). Moura et al. (2014) revealed positive selection and/or fixed differences in loci potentially associated with digestion, growth, metabolism, reproduction and development and function of the heart and muscles. This study suggested that within each ecotype, the loci that were under selection were similar, as may be expected when selection pressures are alike. Foote et al. (2016) later found several positively selected genes potentially associated with ecological specialisation, local adaptation and reproductive isolation between ecotypes. Specifically, this included genes putatively involved in adaptation to cold water climates, such as in the development of adipose tissue and skin, and genes associated with differences in diet and digestion (Foote et al. 2016). Despite some disagreement in the use and interpretation of certain phylogenomic analyses, the two studies found several of the same genes under differential selection that are associated with dietary variation among ecotypes. Recently, Ruan et al. (2015) and Zhou et al. (2018b) investigated the genomic basis of differential osmoregulation in freshwater and marine subspecies of the finless porpoise (*Neophocaena* spp.). These studies found differentially expressed genes related to urine formation and the regulation of water and electrolyte balance in the kidney (Ruan et al. 2015) and over-expression in gene ontology (GO) terms associated with kidney function (Zhou et al. 2018b).

It was also proposed that for some species differential gene expression may play an important role in the preliminary colonisation of new environments (Ruan et al. 2015). Ecotype differences within species have often caused contention among scientists debating the validity of separate species and/or subspecies (e.g. bottlenose dolphins, Ross and Cockcroft 1990; Mead and Potter 1995; killer whales, Morin et al. 2010; tucuxi dolphins, Cunha et al. 2005). In the case of the finless porpoise, two subspecies were named *N. asiaeorientalis asiaeorientalis* and *N. a. sunameri*, corresponding to the freshwater and marine populations, respectively (Ruan et al. 2015; Zhou et al. 2018b). Killer whales on the other hand, are currently classified as one species with a large number of divergent ecotypes, despite past proposals of separate species status for particular populations (see LeDuc et al. 2008; Morin et al. 2010). A similar case exists for bottlenose dolphins (genus *Tursiops*), with distinct offshore/pelagic and inshore/nearshore/coastal/estuarine (hereafter referred to as offshore and inshore, respectively) ecotypes causing debate over the potential presence of separate species and/or subspecies. A subspecies is defined here as “a population, or collection of populations, that appears to be a separately evolving lineage with discontinuities resulting from geography, ecological specialisation, or other forces that restrict gene flow to the point that the population, or collection of populations, is diagnosably distinct” as per the definition provided by Taylor et al. (2017) for cetacean-specific contexts.

1.2.2.1 Bottlenose Dolphins as a Case Study

As many as 20 species of bottlenose dolphins have been previously described in the genus *Tursiops* (Hershkovitz 1966). Currently, there are only two species formally recognised by the Committee on Taxonomy for the Society of Marine Mammalogy (2019): the Indo-Pacific bottlenose dolphin (*T. aduncus*) and the common bottlenose dolphin (*T. truncatus*). The latter has a worldwide distribution, being found in tropical and temperate waters in both offshore and inshore environments. As such, great morphological, physiological and genetic variation is present within the species. This has led to the formal classification of three subspecies within

T. truncatus, the Black Sea bottlenose dolphin (*T. t. ponticus*), the common bottlenose dolphin (*T. t. truncatus*) and the Lahille's bottlenose dolphin (*T. t. gephyreus*) (Committee on Taxonomy of the Society for Marine Mammalogy 2019). Lahille's bottlenose dolphins are found in inshore waters of Brazil, Argentina and Uruguay, and often in sympatry with the darker-bodied, smaller *T. t. truncatus* who typically favour offshore waters (Fruet et al. 2017). In southern Australia on the other hand, differences between the inshore and offshore forms have led to the recent proposal of a separate species, the Burrunan dolphin (*T. australis*) (Charlton-Robb et al. 2011). Along the eastern coast of Australia, the inshore ecotype is recognised as *T. aduncus* and the larger-bodied offshore ecotype as *T. truncatus* (Möller and Beheregaray 2001). The presence of ventral spotting in the northeastern inshore *T. aduncus* populations has previously been used to distinguish the species from *T. truncatus* (Ross and Cockcroft 1990; Hale et al. 2000), although *T. aduncus* individuals further south (New South Wales) have since been found to not possess this feature (Möller and Beheregaray 2001). Along the southern coastline of Australia offshore populations are also recognised as *T. truncatus*, while the classification of the smaller, lighter inshore ecotype as either *T. aduncus* or *T. australis* is still contentious (Charlton-Robb et al. 2011). The Burrunan dolphin was described by Charlton-Robb et al. (2011) based on a series of previously identified morphological, physiological and genetic dissimilarities (Charlton et al. 2006; Möller et al. 2008; Charlton-Robb et al. 2011), but has recently been suggested based on phylogenomic data to probably represent a subspecies of *T. aduncus* (Moura et al. 2020). Due to the currently contentious classification, the southern Australian lineage will hereafter be referred to as the southern Australian bottlenose dolphin (SABD).

Historically, it was thought that offshore bottlenose dolphin populations repeatedly colonised inshore habitats worldwide as new coastal environments were released during interglacial periods (Natoli et al. 2004). This idea is supported by findings of significantly lower genetic diversity in inshore dolphins compared to offshore populations, potentially as a result of

founder events (Hoelzel et al. 1998; Natoli et al. 2004; Louis et al. 2014b; Lowther-Thieleking et al. 2015). More recent clarification of this pattern however, suggests a coastal Australasian origin for the genus, with subsequent transition into the pelagic environment and then repeated colonisation of newly released coastal habitats as *Tursiops* radiated throughout the world's oceans (Moura et al. 2013). Divergence between the offshore and inshore ecotypes may be driven by adaptation to local prey resources and environmental conditions and reinforced by subsequent natal site philopatry, as has been suggested to influence population structuring and divergence in modern-day inshore bottlenose dolphin populations over relatively small spatial scales (Möller and Beheregaray 2004; Tezanos-Pinto et al. 2009; Wiszniewski et al. 2010; Fruet et al. 2014a). To the best of my knowledge, the underlying genomic basis of ecotype formation and adaptation in bottlenose dolphins has however, not been studied. This warrants further investigation due to the presence of both overlapping and divergent phenotypic traits of the inshore ecotype around the world, as compared to the offshore ecotype.

While the inshore and offshore bottlenose dolphin ecotypes are typically reproductively isolated and differ in body size, diet, pigmentation, social structure and behaviour (e.g. Wang et al. 1999; Perrin et al. 2011; Costa et al. 2016), differences between them are not congruent around the world. For example, in southeastern Australia and the northwestern Atlantic Ocean the offshore ecotype is larger in body size than the inshore form (Mead and Potter 1995; Charlton-Robb et al. 2011), while in the southwestern Atlantic and northeastern Pacific Oceans the inshore type is usually larger (Walker 1981; Costa et al. 2016). Where the smaller inshore ecotype has been recorded, particularly in the northwestern Atlantic, this has been coupled with relatively larger pectoral flippers than the offshore type (Wells et al. 1999). Hersh and Duffield (1990) suggested that differences in body and flipper size are likely adaptations for manoeuvrability and thermoregulation. Despite inconsistencies in body size among inshore populations relative to those offshore, inshore dolphins have repeatedly been found to have

fewer vertebrae regardless of total body size (Hale et al. 2000; Kemper 2004; Costa et al. 2016). Several other features are considered characteristic of the inshore ecotype worldwide. In many parts of the world, inshore bottlenose dolphins are found feeding mostly on sciaenid fish and have consequently developed larger teeth than their offshore counterparts, which are more likely to feed on cephalopods (Walker 1981; Mead and Potter 1995; Costa et al. 2016). Studies of differences in skull morphology between ecotypes have also repeatedly reported differences in the shape of the pterygoid hamuli; a structure associated with echolocation in dolphins that might reflect differences in prey and feeding habitats leading to differences in echolocation requirements (Kemper 2004; Perrin et al. 2011; Wickert et al. 2016). Additionally, offshore dolphins have larger internal nares and narrower external nares than the inshore ecotype (Mead and Potter 1995; Wells et al. 1999; Perrin et al. 2011). This is thought to be an adaptation to improve the efficiency of air exchange in the offshore animals while deep-diving (Perrin et al. 2011). Offshore dolphins also have higher haemoglobin concentrations and a higher ratio of red blood cells to total blood volume than the inshore dolphins (Duffield et al. 1983), likely to increase oxygen storage while deep-diving (Hersh and Duffield 1990). It is currently unknown if these adaptations have evolved through changes to the same (i.e. parallel adaptive evolution) or different genes in each inshore population. Unlike killer whales, which have been studied in some detail, current research of bottlenose dolphin ecotypes is only scratching the surface. Studies of this genus are largely restricted to the social, morphological and skeletal differences as described above. While genetic differentiation has been reported between inshore and offshore populations, the genomic basis of the formation of these ecotypes and how this has potentially led to speciation in parapatric and sometimes sympatric habitats has not been studied. Similar to studies done on killer whales and finless porpoises, we now have the ability to deduce which genes are putatively associated with local adaptation and subsequent genomic divergence of bottlenose dolphin ecotypes. This gives us the opportunity to investigate at a genome level the potential for parallel evolution in the inshore ecotype as they repeatedly adapted to similar selective pressures.

In addition to investigating the basis of genomic divergence between ecotypes and species, it is now also possible to study within-species differentiation (i.e. at the population level). Coastal seascapes often span over several divergent habitats and stark environmental and oceanographic gradients. As such, marine species that inhabit these regions have adapted to a range of conditions. Local adaptation and resource specialisation have been suggested to be major drivers in the fine-scale population genetic structure of inshore bottlenose dolphins (e.g. Wiszniewski et al. 2010; Mirimin et al. 2011; Fruet et al. 2014b). Variables such as sea surface temperature (SST), salinity, topography, primary productivity and patterns of ocean circulation are thought to impact on bottlenose dolphins indirectly by causing discontinuities in prey abundance and distribution between regions (Bilgmann et al. 2007b; Wiszniewski et al. 2010). In the eastern Mediterranean, for example, decline in bottlenose dolphin numbers during the LGM corresponded with high salinity levels that possibly caused severe reductions in prey abundance (Gaspari et al. 2015). While the relationship between environment and bottlenose dolphin population structure has long been suggested, this has not been empirically tested. This can now be done using a landscape genomics framework. This approach combines environmental and genome-wide data of individuals sampled across a heterogeneous habitat to deduce patterns of variation at both putatively neutral and adaptive loci (Joost et al. 2007). Neutral loci are DNA regions that do not influence the fitness of an organism and therefore, are putatively not under selection. Adaptive loci on the other hand, are those that potentially impact on fitness and alleles at these loci can be selected for or against by natural selection (Holderegger et al. 2006). The landscape genomic method is becoming increasingly popular but has only recently began to be implemented in marine systems (i.e. seascape genomics) (e.g. Sandoval-Castillo et al. 2018; Teske et al. 2019). This approach tests for relationships between genomic and environmental variation and can identify genomic regions putatively under selection. Seascape genomics allows investigation of both the indirect impacts the environment has on bottlenose dolphins by influencing their

prey, as well as the direct impacts the environment has on the physiological adaptation of dolphins to local conditions and resources. This is a concept heavily understudied and potentially underestimated in cetaceans and other marine megafauna. Indeed, no seascape genomics studies have been published so far on cetaceans (but see Mendez et al. (2010) and Amaral et al. (2012a) for examples of seascape *genetic* studies). These kinds of studies allow greater insight than was previously possible into the forces driving population genomic differentiation and local adaptation in cetaceans.

In Australia inshore bottlenose dolphins are found in a wide range of habitats and environmental conditions. This region is therefore, an excellent system for investigating microevolutionary adaptations of cetaceans to variable environmental and oceanographic features. Divergent lineages of inshore bottlenose dolphin on the southern (SABD) and eastern (*T. aduncus*) coastlines allow exploration of how the two taxa have adapted to their local habitats and how ecological features are potentially driving genomic differentiation within each lineage. Along the southern coastline six genetically distinct populations have been previously suggested (Bilgmann et al. 2007b; Charlton-Robb et al. 2015; Pratt et al. 2018), with particularly fine-scale population structure within South Australia (Bilgmann et al. 2007b; Pratt et al. 2018). These subpopulations span over several bioregions characterised by differences in oceanographic, environmental and biological features (see DEH 2006). This is particularly evident in the two South Australian gulfs, with three bioregions classified in Spencer Gulf (SG) due to the influence of discontinuous north-south flushing systems and the resulting strong salinity gradient (Bullock 1975; Kämpf et al. 2010). Gulf St. Vincent (GSV) however, has a more homogenous environment and as such only one bioregion (IMCRA Technical Group 1998). Environmental discontinuities, as well as differences in geographical history and marine transgression, in particular between the two South Australian gulfs (see Belperio et al. 2002; Harvey 2006), have been previously hypothesised to be influencing population genetic structure (Bilgmann et al. 2007b; Pratt et al. 2018). Fine-scale genetic

structure can also be seen along the east Australian coast, with four genetic populations previously suggested along the New South Wales coast (Möller et al. 2007; Wiszniewski et al. 2010). This may be influenced by the East Australian Current (EAC) creating strong latitudinal patterns in environmental gradients and population structure of local species, including common dolphins (*Delphinus delphis*) (Möller et al. 2011; also see Hoskin 2000; Banks et al. 2007; Piggott et al. 2008; Shaddick et al. 2011). This is also seemingly associated with differing habitat types, particularly the presence of the large Port Stephens embayment where strong *T. aduncus* social structure has been reported (Möller et al. 2001; Möller et al. 2006). For both inshore Australian bottlenose dolphin lineages and indeed all marine mammal species, scientists have previously only been able to speculate as to what was driving population divergences based on the positions of genetic breaks. The seascape genomics approach however, can be used to not only provide greater resolution to studies of neutral population structure, but to identify gene regions that are putatively under differential selection between populations in different regions and explicitly test for associations with environmental variables. When paired with annotation of candidate genes, a much greater understanding of the processes that are potentially driving adaptive divergence and local adaptation of these marine populations can be achieved.

1.3 Applications of Evolutionary Studies to Conservation Management

With rapid climatic and habitat changes occurring worldwide, it is becoming increasingly important to further our understanding of the evolutionary history and adaptive potential of species. Emerging genomic techniques are now being applied to a wide range of non-model species and are quickly being implemented to marine organisms. Genomic studies not only encompass the diversity of techniques previously utilised in genetic research, but now allow increased power, resolution and depth for this work (Luikart et al. 2003). In particular, genomic techniques allow scientists to investigate how populations and species are influenced by, and have adapted to, certain environmental conditions and habitats. Furthermore, these techniques allow exploration of population genomic structure on a very fine-scale and could potentially reveal the early stages of genomic divergence associated with local adaptation to specific ecological features. This is crucial in developing our understanding of how species are likely to respond to ongoing climate change and habitat modifications. The marine ecosystem is threatened by a wide range of anthropogenic factors, including pollution (noise, light, chemical and physical), habitat degradation and destruction, ocean acidification, ocean warming and overfishing, and is predicted to undergo dramatic changes in coming years (Halpern et al. 2008; Doney et al. 2011; Scheffers et al. 2016). Evolutionary studies are critically important in assessing the adaptive potential of species to ensure conservation efforts are targeted to those that are most vulnerable to population declines. Populations with limited adaptive potential include those that reside in regions that are, or are predicted to be, heavily disturbed and those that exhibit low genetic diversity, restricted geographical ranges, small effective population sizes and/or long generation times. These are all common characteristics of cetaceans, especially for populations of inshore delphinids. Genomic studies can therefore, be useful in identifying key pathways by which dolphins could potentially adapt to changing environmental conditions in the future and highlight conservation strategies that can be used to help facilitate this process. Directed and proactive strategies, including extensive marine

park networks and/or specific mitigation of anthropogenic impacts, are needed to ensure the persistence of healthy marine populations, communities and ecosystems.

1.4 Thesis Aims

Evolutionary genomic studies are becoming increasingly accessible, especially with non-model species, such as cetaceans. These studies have a great diversity of applications and are particularly useful in addressing questions surrounding the impacts of environmental change both temporally and spatially. However, research surrounding the microevolutionary genomic adaptations of cetaceans to environmental heterogeneity is still in its infancy. These studies are important to understand drivers of population structure and adaptive differentiation, particularly in the face of ongoing climate change. With this in mind the overall aim of this thesis is to investigate the genomic basis of environmental adaptation in bottlenose dolphins (genus *Tursiops*). This will be addressed using genomic double-digest restriction-site associated DNA sequencing (ddRADseq) datasets generated from biopsy samples of wild bottlenose dolphins from localities spanning three ocean basins in the Southern Hemisphere. First, the genomic basis of bottlenose dolphin ecotype formation will be investigated. Specifically, this study will explore genomic divergence in the genus *Tursiops* and question how this is associated with the repeated emergence and environmental adaptation of inshore and offshore ecotypes. Secondly, a seascape genomics framework will be implemented to study the relationship between environmental variation and population genomic structure in inshore bottlenose dolphins of Australia. The putatively divergent lineages of *T. aduncus* along coastal eastern Australian and SABD along coastal southern Australia will be investigated separately. These studies will question how and why bottlenose dolphin neutral and adaptive population genomic structure differ, and which ecological variables and specific genes may be involved in this adaptive differentiation. This work will be used to identify drivers of ecotypes and population divergence, as well as pathways potentially involved in the adaptation of bottlenose dolphins to heterogeneous environments. This will aid in assessing the adaptive potential of these vulnerable species, subspecies and populations and inform conservation strategies to support them.

Chapter 2 : Genomic Divergence and Ecotype Formation in the genus *Tursiops*

2.1 Contributions

Eleanor Pratt – conception of study and design of methods, collection of Gulf St. Vincent samples, DNA extraction and ddRAD library preparation, bioinformatics and analysis, writing of thesis.

Luciana Möller – primary supervisor – conception of study design, guidance in interpretation, collection of eastern and southern Australian samples, drafting and revision of thesis.

Luciano Beheregaray – associate supervisor – guidance in study design and interpretation, drafting and revision of thesis.

Jonathan Sandoval-Castillo – guidance for laboratory methods and bioinformatics, and assistance with analysis.

Kerstin Bilgmann – collection of Spencer Gulf, St. Francis Island and Western Australian samples.

Nikki Zanardo – collection of Adelaide samples and assistance in collecting other Gulf St. Vincent samples.

Fernando Diaz-Aguirre – collection of Coffin Bay samples.

Gabriela de Tezanos-Pinto – collection of New Zealand samples.

Pedro Fruet – collection of southwestern Atlantic Ocean samples.

2.2 Permits and Ethics Approvals

Southern/Western Australia

Biopsy samples were collected with Ministerial Exemption from Primary Industries Resources South Australia (PIRSA), exemptions #9902404, #9902648, #9902714 and #9902601, with permits #K25761-6, #E25889 and #E26171 from the Department of Environment, Water and Natural Resources (DEWNR), South Australia, #SF008961 from the Department of Environment and Conservation, Western Australia and #2008-0001 from the Department of Environment, Water, Heritage and the Arts (for sampling in Commonwealth waters). Animal ethics approvals were acquired from the Flinders University Animal Welfare Committee, projects #E310, #E375 and #E326.

Eastern Australia

Biopsy samples were obtained under licences from the Department of Environment and Climate Change (Licence Number: S10763) and Marine Parks Authority (Permit Number: PSGLMP 2008 ⁄ 003) and under approval by the Macquarie University Animal Ethics Committee (AEC Reference Number: 2007⁄013) as per Wiszniewski et al. (2010 and 2012).

New Zealand

Samples were collected under Massey University, NZ permits and imported into Australia under Australian Quarantine and Inspection Service permit 0001172530 and the relevant CITES Appendix II permit (NZ013 to AU089) in March 2017.

Southwestern Atlantic Ocean

Samples were collected under regional permits (Brazil: SISBIO 24429-1 issued to PAC Flores, SISBIO 24407-2 issued to PF Fruet) and transferred to Australia under CITES permits 11BR007432/DF and 2011-AU-647980.

2.3 Abstract

Climatic changes over time have caused major environmental restructuring throughout the world's oceans. Marine organisms have responded in a number of ways, including through genomic adaptation to the new conditions. Growing accessibility of genomic methods to study non-model species now allows genomic changes underlying environmental adaptation to be investigated. This study addresses the genomic basis of ecotype formation in bottlenose dolphins (genus *Tursiops*) in the Southern Hemisphere, utilising a double-digest restriction site-associated DNA (ddRAD) dataset of over 18,000 single nucleotide polymorphisms (SNPs). Subspecies-level genomic divergence was confirmed between the offshore common bottlenose dolphin *T. t. truncatus* and the inshore *T. t. gephyreus* in the southwestern Atlantic Ocean (SWAO). Similarly, subspecies-level divergence is suggested between the inshore Indo-Pacific bottlenose dolphin *T. aduncus* and the Burrunan dolphin (*T. australis*) in Australia. Inshore bottlenose dolphin lineages generally had lower genomic diversity than offshore lineages, a pattern particularly noticeable for *T. t. gephyreus* which showed exceptionally low diversity. Several major bodily systems, including the cardiovascular, musculoskeletal and energy production systems, appear to be implicated in the repeated evolution of the inshore lineages across the Southern Hemisphere. It is hypothesised that comparable selective pressures in the inshore environment, including changes in water depth, habitat complexity, prey choice and several other environmental variables, drove similar adaptive responses in each lineage, providing support for parallel evolution in inshore bottlenose dolphins. With climate change altering environmental conditions worldwide, particularly in coastal marine ecosystems, it is important to gain information about the adaptive capacity of marine species and populations. This study highlights several major physiological systems that are under differential selection between the bottlenose dolphin ecotypes, providing insights into key adaptive pathways that are potentially crucial for the long-term survival of cetaceans and other organisms in a changing marine environment.

2.4 Introduction

Environmental change and the opening of new niche spaces have been important drivers in the evolution of species (Wellborn and Langerhans 2015; Stroud and Losos 2016). Vacant or underutilised niche spaces after mass extinctions, in particular, has led to the macroevolution of several groups, such as flight in bats and birds (Simmons 2005; Zhang et al. 2014) and the radiation of terrestrial mammals (Meredith et al. 2011). Macroevolution is the process by which new clades are formed, typically in conjunction with major morphological and physiological adaptations (e.g. the evolution of the wing for vertebrate flight) (Reznick and Ricklefs 2009). These adaptations are driven by a change in selective pressures. Natural selection acts in this way by favouring organisms with particular traits that make them more likely to survive and reproduce in the new niche space, thereby potentially passing these traits on to the next generation (Endler 1986). After the initial colonisation of a new environment, subsequent radiation of species will typically occur as organisms become increasingly adapted to their local conditions. This level of adaptation is commonly referred to as microevolution (Jablonski 2000). New species are formed as populations within species become reproductively isolated, either through the formation of physical barriers or as a result of extensive genomic differentiation. In a heterogeneous environment, different selective pressures will act on each local population. This often drives them to become locally adapted and paired with neutral processes, such as mutation and genetic drift, can cause genomic differentiation and divergence among populations (Nosil and Feder 2012; Luque 2016). This is particularly enhanced in small populations which exhibit limited dispersal ranges and low gene flow to other populations (see Raeymaekers et al. 2017). Over time, if gene flow is not re-established, this can lead to the formation of new species and lineages.

Colonisation of similar niche spaces in different regions can in some cases result in parallel evolution. Similar environmental conditions typically have comparable selective pressures and can lead to the repeated formation of particular adaptations and traits in separate populations

and species (Christin et al. 2010; Stern 2013). Parallel evolution differs from convergent evolution in that the latter refers to the independent formation of similar traits in distantly related or unrelated species, for example, the evolution of echolocation in both bats and cetaceans (Au 1997; Li et al. 2010; Davies et al. 2012). Parallel evolution on the other hand, deals with this phenomenon when it occurs in lineages derived from a recent common ancestor (Wood et al. 2005). The independent rise of similar traits can result from a number of different processes: through identical, independent mutations in different populations or species, through selection on a polymorphic allele present in both populations or species from shared ancestral history and/or through the introduction of an allele into a population via introgression (Stern 2013). A prominent and well documented example of parallel evolution is the case of ecotype formation in the threespine stickleback (*Gasterosteus aculeatus)*. From the ancestral marine ecotype, these fish have repeatedly colonised newly opened freshwater systems, forming several marine-freshwater ecotype pairs across their range (McKinnon and Rundle 2002). In each case of freshwater colonisation, similar adaptations associated with body shape, armour, pigmentation, trophic position and salt handling have been recorded (McKinnon and Rundle 2002; Jones et al. 2012). Phenotypic parallelism does not however, necessarily stem from changes in the same genomic loci. On the contrary, there may be many ways to produce similar phenotypic traits (Christin et al. 2010; Stern 2013). It is therefore, important to investigate the genomic underpinnings of these adaptations to establish the extent and causes of parallel evolution. In the case of threespine sticklebacks, several loci were found to be repeatedly implicated in the evolution of the freshwater ecotype (Jones et al. 2012). Parallel genomic evolution was also revealed in recurrent ecotype formation in other fish species, such as in lake whitefish (*Coregonus* sp.; Rogers and Bernatchez 2007) and European anchovies (*Engraulis encrasicolus*; Le Moan et al. 2016). In contrast, phenotypic parallelism in the evolution of planktivorous and piscivorous lake trout (*Salvelinus namaycush*) ecotypes was not reflected at the genomic level, indicating different genomic changes had led to the evolution of a similar phenotype across populations (Perreault-Payette et al. 2017). Many studies of parallel evolution utilise quantitative trait loci (QTL) mapping techniques,

involving selective breeding to track particular traits of interest (e.g. Rogers and Bernatchez 2005; Jones et al. 2012). For many species however, QTL mapping is not possible and thus, cases of parallel evolution are more difficult to track. Recent advances in genomic techniques now allow sampling of thousands of loci across the genome of non-model species, enabling traditional tests for selection to establish regions of the genome that are putatively under selection within populations or species. Comparison can then be made across lineages to establish if parallel evolution has possibly occurred. This framework can be particularly useful in studying the radiation and adaptation of non-model species, including that of cetaceans.

Cetaceans provide an excellent opportunity to study both macro- and microevolutionary adaptations. These animals underwent major morphological, physiological and behavioural changes as they transitioned from a terrestrial to a fully aquatic lifestyle approximately 53 MYA (Thewissen et al. 2007; McGowen et al. 2014). They subsequently radiated throughout the world's oceans and into freshwater ecosystems, leading to the vast diversity of cetacean families and species seen today. Further subdivision exists within many of these species as a result of adaptation and specialisation to particular niches. While adaptations associated with the macroevolution of cetaceans have been well documented (Thewissen and Bajpai 2001; Thewissen et al. 2009; Sun et al. 2012; Zhou et al. 2018a), microevolutionary adaptations are only now beginning to be investigated in detail, especially at the genomic level. The most well studied example is that of the killer whale (*Orcinus orca*), where distinctive differences among sympatric and allopatric ecotypes are evident (Foote et al. 2009; Pitman et al. 2011; Moura et al. 2014). Ecotypes are defined here as populations within a species that have evolved heritable variation in physiology, morphology, behaviour and/or life history due to environmental differences (see Le Moan et al. 2016). Killer whale ecotypes are known to differ in their morphology, prey choices and hunting techniques (e.g. piscivorous versus mammal-eating) (Heimlich-Boran 1988; Ford et al. 1998; Foote et al. 2009; Morin et al. 2010). This is reinforced by strong social behaviours and matrilineal relationships within groups and groups

and has led to genomic divergence and adaptation among different populations (Moura et al. 2014; Foote et al. 2016). Although only one species of killer whale is currently recognised worldwide, ecotype differentiation in other cetaceans has led to the formal classification of separate species or subspecies. This is the case for the marine and freshwater ecotypes of the narrow-ridged finless porpoise (genus *Neophocaena*; Jefferson and Wang 2011) and tucuxi dolphins (genus *Sotalia*; Cunha et al. 2005; Caballero et al. 2007; Committee on Taxonomy of the Society for Marine Mammalogy 2019). Despite the presence of ecotype differentiation in several cetacean species, the repeated formation of inshore and offshore ecotype pairs in bottlenose dolphins (genus *Tursiops*) around the world provides a unique opportunity to study parallel evolution in marine mammals.

Inshore (i.e. all nearshore, coastal, estuarine and brackish environments) and offshore forms of bottlenose dolphins typically differ in a number of traits. This includes body size (Ross and Cockcroft 1990; Costa et al. 2016), fin size and shape (Félix et al. 2018), diet (Wang et al. 2000), colouration (Diaz-Gamboa et al. 2018), parasite load (Walker 1981), level of population genetic diversity (Lowther-Thieleking et al. 2015; Fruet et al. 2017) and social behaviours (e.g. group size and home ranging patterns) (Costa et al. 2015; Diaz-Gamboa et al. 2018). The typical characteristics of inshore and offshore dolphins are however, not consistent on a worldwide scale. For example, the inshore type is smaller in body size than the offshore type in the northwestern Atlantic Ocean (Hersh and Duffield 1990; Mead and Potter 1995) and in southern Australia (Charlton-Robb et al. 2011), while the opposite pattern has been reported in the southeastern Pacific Ocean (Diaz-Gamboa et al. 2018) and the SWAO (Costa et al. 2016). The inshore form has however, been repeatedly shown to have fewer vertebrae than the offshore form, irrespective of total body size (Hale et al. 2000; Kemper 2004; Costa et al. 2016; Wickert et al. 2016). Similar changes in the nares and pterygoid bones have also been shown among inshore populations compared to offshore, potentially related to differences in diving behaviour and echolocation requirements in the two habitats (Kemper 2004; Costa et

al. 2016; Wickert et al. 2016). The inshore environments inhabited by bottlenose dolphins share numerous characteristics that differ from the offshore ecosystem, including increased structural complexity, shallow depth, generally warmer SST and greater freshwater input. Despite some differences, similarity in these features may create comparable selective pressures across the inshore habitats, resulting in cases of phenotypic parallelism in the inshore bottlenose dolphin ecotype. By investigating the underlying genomic basis of ecotype formation, it is possible to determine the extent to which phenotypic parallelism is underpinned by genotypic parallelism and potentially reveal additional adaptive differences. With several marine species showing inshore and offshore differentiation, such as European anchovies (Le Moan et al. 2016; Montes et al. 2016), Icelandic cod (*Gadus morhua*; Bardarson et al. 2018) and Australian sea lions (*Neophoca cinerea*; Lowther and Goldsworthy 2011), this framework may be useful in deducing how the inshore environment drives genomic divergence and adaptation across cetaceans and other marine vertebrate species.

While inshore-offshore bottlenose dolphin ecotype pairs have been recorded worldwide, the extent of divergence differs depending on the region. This is likely a reflection of the relative age of divergence between lineages. Inshore and offshore bottlenose dolphins in Australian waters have been classified as separate species, with the offshore ecotype recognised as the common bottlenose dolphin (*T. truncatus*) and the inshore ecotype as the Indo-Pacific bottlenose dolphin (*T. aduncus*; Hale et al. 2000; Möller and Beheregaray 2001; Kemper 2004; Moura et al. 2020). However, in southern Australia Charlton-Robb et al. (2011) described a separate inshore species, the Burrunan dolphin (*T. australis*) based on previous morphological and genetic evidence (also see Charlton et al. 2006; Möller et al. 2008). In SWAO the offshore ecotype is recognised as *T. t. truncatus*, while the inshore ecotype is classified as the Lahille's bottlenose dolphin, *T. t. gephyreus* (Fruet et al. 2014a; Costa et al. 2016; Wickert et al. 2016; Fruet et al. 2017; Costa et al. 2019). On the other hand, in the northwestern Atlantic and northeastern Pacific Oceans, the inshore and offshore ecotypes are divergent genetically,

morphologically and physiologically, but are both currently classified as *T. t. truncatus* (Perrin et al. 2011; Lowther-Thieleking et al. 2015). The classification of these ecotypes has however, been controversial, with the genus *Tursiops* previously divided into as many as 20 different species (Hershkovitz 1966). Incomplete lineage sorting, inconsistent patterns in morphology and potential hybridisation with other delphinid species has resulted in extensive confusion in the taxonomy of the Delphinidae family (Amaral et al. 2012b; Moura et al. 2013). As such, the potential presence of several other subspecies and highly divergent evolutionary lineages are still being considered. This includes the Burrunan dolphin (*T. australis*) in southern Australia (Charlton-Robb et al. 2011), which will hereafter be referred to as the SABD lineage due to its controversial taxonomy. It has been suggested that SABD split from the rest of the genus *Tursiops* approximately 992 thousand years ago (Gray et al. 2018). A more recent divergence has been proposed between *T. aduncus* and *T. truncatus* approximately 714 thousand years ago (Gray et al. 2018), contrary to earlier predictions of a split around 2.5 MYA (Vilstrup et al. 2011). A recent comprehensive study of phylogenomic relationships in this genus based on over 25,000 genetic markers has proposed a subspecies level classification for SABD under *T. aduncus* (Moura et al. 2020). A subspecies in the context of cetaceans is defined here as per Taylor et al. (2017) and refers to "a population, or collection of populations, that appears to be a separately evolving lineage with discontinuities resulting from geography, ecological specialisation, or other forces that restrict gene flow to the point that the population, or collection of populations, is diagnosably distinct." Further clarification of phylogenomic relationships in the *Tursiops* genus is needed, particularly to provide additional evidence for the separation of SABD from *T. aduncus*, either at species or subspecies level. With the divergence of species and subspecies in this genus seemingly associated with adaptation to new habitats and niche spaces (i.e. the inshore environment), deeper investigation of particular ecological features that may be driving adaptation and evolution in this genus is warranted.

This study aims to investigate the genomic basis of ecotype microevolution in bottlenose dolphins (genus *Tursiops*) in the Southern Hemisphere. A ddRADseq dataset of 18,060 SNPs was used to first establish the phylogenomic relationships between *Tursiops* lineages. The sampling includes several of the recognised and proposed lineages from the Southern Hemisphere but is not exhaustive. Potentially divergent lineages not included here are those inhabiting southern Africa (*T. aduncus;* Natoli et al. 2004) and the southeastern Pacific Ocean (*T. truncatus*; Félix et al. 2018). It is hypothesised that high genomic differentiation will be detected between the inshore and offshore ecotypes, and among inshore populations from different ocean basins, across the Southern Hemisphere. Adaptation to opposing environments is expected to be driving genomic differentiation between ecotypes, while response to similar selective pressures in the inshore environments may be reflected in parallel evolution of their populations. A number of the sampled populations and lineages inhabit waters in close proximity to urbanised areas and are therefore, subject to human-related stressors, such as pollution, bycatch, overfishing, tourism, boat strikes and habitat degradation (see Daura-Jorge and Simões-Lopes 2011; Tezanos-Pinto and Constantine 2013; Charlton-Robb et al. 2015). Well-informed management and conservation strategies are therefore, needed to ensure that these populations are not negatively affected by human interactions to an irreversible extent. A crucial step is to clarify species and subspecies levels of genomic structure and patterns of gene flow among regions, as well as to identify populations of high conservation importance. Studying how these dolphins have evolved in response to different selective pressures allows a better understanding of how they may continue to diverge and adapt to environmental changes and in particular to rapid anthropogenic climate change.

2.5 Methods

2.5.1 Sample Collection

Biopsy samples from free-ranging bottlenose dolphins (*Tursiops* spp.) were collected from 29 locations in four countries of the Southern Hemisphere between 1998 and 2016 (Figure 2.1; Appendix B: Table B.ii.1). Skin and blubber samples were collected from individuals using either a hand-held biopsy pole for bow-riding dolphins (Bilgmann et al. 2007a) or a remote biopsy gun system for individuals surfacing five to ten metres away from the boat (Krützen et al. 2002). Resampling of individuals was minimised by visually checking for biopsy wound marks on the animal's body and through identification of recognisable dorsal fin characteristics. No samples were obtained from dependent calves. Biopsy samples were preserved in either 100% ethanol or a salt-saturated solution of 20% dimethyl sulphoxide (DMSO) and stored at -80ºC upon return to the laboratory.

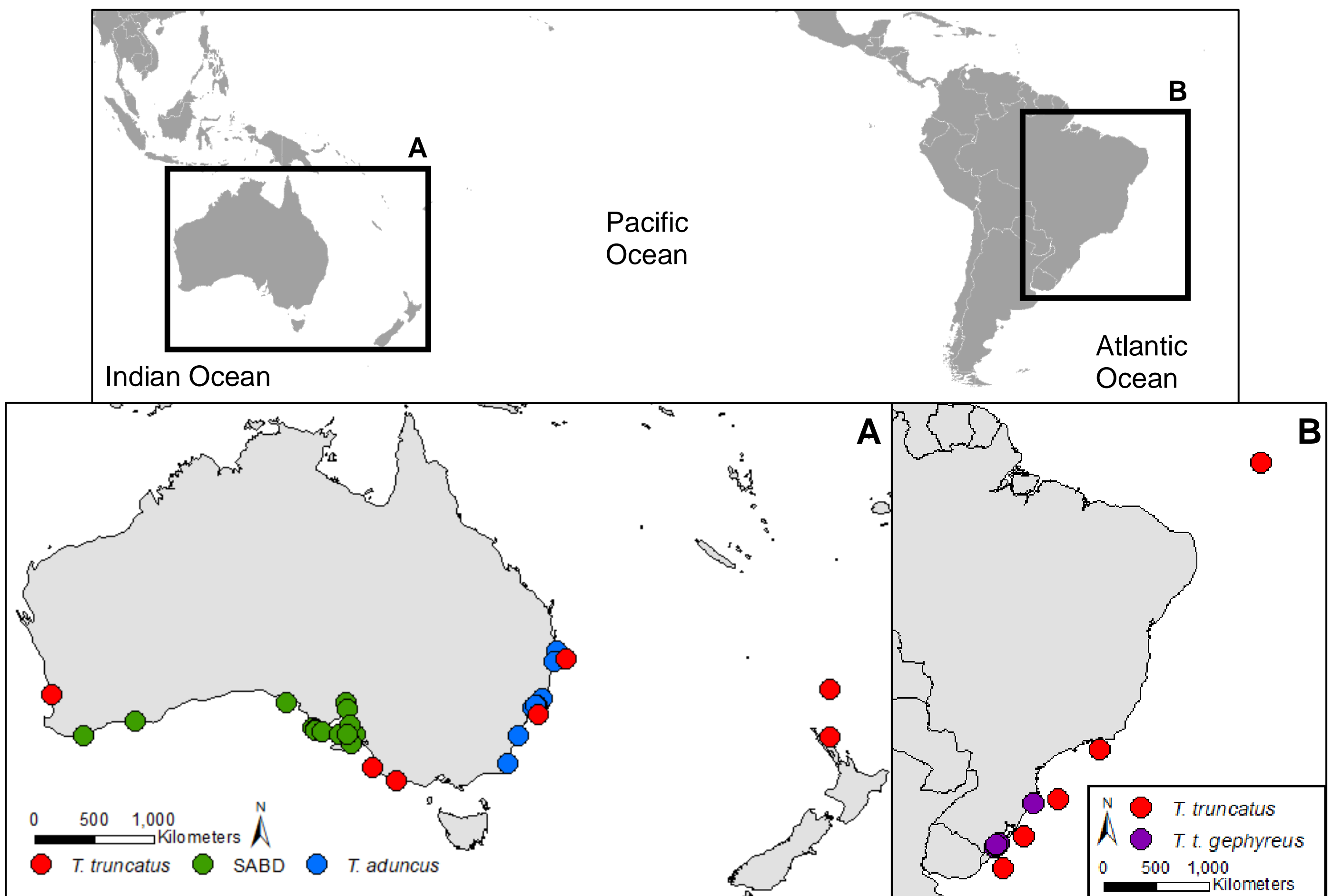


Figure 2.1 Sampling locations of *Tursiops* spp. across three ocean basins in the Southern Hemisphere.

2.5.2 *Genomic Laboratory Methods*

2.5.2.1 DNA Extraction

DNA was extracted from biopsy samples using a salting-out protocol (Sunnucks and Hales 1996) with modifications. Briefly, this involves cutting a small section of skin tissue (~2 x 2 mm) and rinsing with H_2O to remove the preservation buffer. The dried skin fragments were then placed in a solution of TNES buffer, Proteinase K and RNase and incubated overnight at $37^{O}C$ and 800 rpm. After removing from the incubator, 5 mol NaCl was added to the solution. Samples were then centrifuged at 14,000 rpm for five minutes and the liquid was carefully transferred to a new tube. This step was then repeated to obtain clean DNA. To extract the DNA from the solution, washes with 100-75% cold ethanol were used, centrifuging at maximum speed for seven minutes in between washes. The DNA pellet was retained each time and after the last wash was dried under a heat lamp until all remaining ethanol had evaporated. DNA was then rehydrated with distilled H_2O. DNA integrity was assessed by gel electrophoresis and purity was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific). Microsatellite data were used to remove closely related animals by selecting only one sample from any pair that had a relatedness estimate of ≥0.5 (theoretical value for first-order relatives, $r = 0.5$). This was calculated in *GenAlEx* (Peakall and Smouse 2006, 2012) using the Queller and Goodnight (1989) estimator. Microsatellite datasets were already available for many of the sampled locations (Wiszniewski et al. 2010; Fruet et al. 2014b; Fruet et al. 2017; Pratt et al. 2018), including unpublished data from the St. Peter and St. Paul Archipelago (ASPSP) samples, off Brazil. Samples from Robe and Cape Nelson in southern Australia did not have an existing microsatellite dataset and therefore, seven loci (Tur80, Tur87, Tur105, TurE12, Tur142, Tur91, Tur141) were amplified using the polymerase chain reaction (PCR) and genotyped as per conditions specified in Pratt et al. (2018) to remove closely related individuals.

2.5.2.2 ddRAD Library Preparation

Libraries were prepared following a ddRADseq protocol modified from Peterson et al. (2012), as per Brauer et al. (2016). Briefly, 300 ng of DNA from each individual was digested with the restriction enzymes SbfI and MseI (New England Biolabs) and then individual barcodes and RAD adaptor sequences were ligated. Samples were then combined into multiplex pools consisting of 12 individuals each. Pooled samples were then purified with Agencourt AMPure XP (Beckman Coulter Genomics) and washed twice with 80% ethanol. Using a 1.5% Pippin prep electrophoresis gel (Sage Science) the libraries were size selected for fragments of between 250 and 800 base pairs (bp). DNA concentration was measured using a Qubit 2.0 fluorometer (Life Technologies). Fragments were then PCR amplified in two 25 uL reactions per pool which were recombined after amplification to reduce PCR bias and were again purified as per the initial purification process. Fragment sizes were then quantified using the Bioanalyzer 2100 (Agilent Technologies) with a DNA 7500 assay kit. Next, a quantitative PCR (qPCR) was performed to measure DNA quantity. Due to the high sensitivity of qPCR DNA was first diluted to 1 uL in 2000 uL of H_2O. DNA was combined with SensiFAST SYBR Hi-ROX (BIO-92002) enzyme/buffer and two qPCR adapter primers (5 uM). For each individual sample, three replicates were used to obtain accurate DNA quantity estimations. Three blank replicates (H_2O) and three replicates of each of four control samples of known DNA concentration (0.1, 1.0, 10 and 100 pM) were also used to assess the accuracy of the qPCR estimates. The PCR was run with an initial stage at 94°C for 2 minutes, before progressing to 40 cycles of 94°C for 15 seconds, 62°C for 15 seconds and 72°C for 32 seconds. The qPCR final cycle was 95°C for 15 seconds, 62°C for 1 minute and 95°C for 15 seconds. DNA quantity estimates were averaged between the three sample replicates and the qPCR was repeated for samples with large discrepancy between replicates. Informed by these results, the volume of each pool to be added to the final library was calculated so as to achieve equal quantities of DNA from the four pools. Four of the final multiplexed libraries consisted of 48 individually barcoded samples, while the other five libraries consisted of 96 samples. All libraries were

sequenced at the South Australian Health and Medical Research Institute (SAHMRI) on an Illumina HiSeq2000 platform as single-end, 100 bp reads.

2.5.3 Bioinformatics

The *dDocent* v.2.2.19 (Puritz et al. 2014) pipeline was used to demultiplex and process the raw data files. This was done by first using *Trimmomatic* v.0.36 (Bolger et al. 2014) to remove adapter sequences and low quality bases. Two steps were then used to assemble a catalogue of reference contigs *de novo*. First, *Rainbow* v.2.0.3 (Chong et al. 2012) was used to cluster reads into contigs based on similarity and then assemble the clusters into longer reference contigs (maximum number of mismatches = 6). Second, *CD-HIIT* v.4.6 (Li et al. 2001) was used to cluster the contigs based on 90% similarity. Only the longest contig from each cluster was retained for the final assembled reference contig catalogue. Using default settings in *BWA* v.0.7 (Li and Durbin 2010) the MEM algorithm (Li 2013) was used to map reads to the reference contigs for each individual. *FreeBayes* v.0.9 (Garrison and Marth 2012) was then used with default settings to detect variants, including SNPs, insertions and deletions (INDELs) and multi-nucleotide polymorphisms. Scripts for the above *dDocent* processes are in Appendix B.iii.1 and 2. Finally, *VCFtools* was used to filter the resulting variant call file (VCF) using custom BASH scripts (Appendix B.iii.3) for the filtering steps outlined in Appendix B: Table B.ii.2 (modified from Brauer et al. 2016). Retained loci were then mapped against the *T. aduncus* genome, downloaded from the National Center for Biotechnology Information (NCBI) (GCA_003227395.1 ASM322739v1). Only loci that aligned to the genome were retained for analysis. This process was then repeated, starting after the demultiplexing stage, with the inclusion of nine common dolphins (*Delphinus delphis*), to be used as outgroup for phylogenomic analyses. Common dolphin sequences were available from another project underway at the Molecular Ecology Lab at Flinders University (Barcelo-Celis et al., unpublished data).

2.5.4 Genomic Variation

Molecular diversity indices for dolphins at each sampling location, including the percentage of polymorphic loci (%PL) and expected and observed heterozygosity (H_E and H_O, respectively) were calculated at the SNP level in *Arlequin* v. 3.5.2.2 (Excoffier and Lischer 2010). Wright's inbreeding coefficients (F_{IS}) for each sampling location were calculated as (H_E - H_O)/H_E (Wright 1922). The R package *PopGenKit* (Paquette 2011) was then used to determine the number of private alleles (PA) in each putative lineage (i.e. loci with a particular variant detected in only one lineage).

2.5.5 Genomic Divergence

2.5.5.1 Phylogenomics

A phylogenomic tree was generated in *RAxML* v.1.5 (Stamatakis et al. 2004) to investigate phylogenetic relationships within the genus *Tursiops*. This was run with nine common dolphins (*D. delphis*) as outgroup, selected based on a recent study which suggested monophyly of the genus *Tursiops* (Moura et al. 2020). Fourteen potential hybrids (individuals with >20% estimated membership to a second lineage) were identified in *Admixture* analysis and were subsequently removed from this analysis. *RAxML* was run using the GTRGAMMA model of evolution and 1,000 resampling estimated log-likelihood (RELL) bootstraps. The output was visualised in *FigTree* v.1.4.3 (Rambaut 2014), rooted with the outgroup.

2.5.5.2 Population Genomic Structure

To assess genomic divergence among and within the four lineages (i.e. *T. t. truncatus, T. t. gephyreus, T. aduncus* and SABD) a number of traditional population genomic structure methods were implemented. The term "population" used here thus refers to a putative lineage. *Arlequin* was used to estimate pairwise genomic differentiation (F_{ST}) and corresponding significance levels between sampling locations based on 10,000 permutations. To account for

biases potentially created by multiple testing, the significance levels were corrected using Benjamini and Yekutieli's (2001) method (B-Y correction) (see Narum 2006). This resulted in an alpha (α) level of 0.0076. F_{ST} values among and within the four putative lineages were then averaged across sites. To establish the most statistically supported number of populations in the dataset, the model-based maximum-likelihood method in *Admixture* v.3.5.2.2 (Alexander et al. 2009) was run testing for population values from one to 25 (based on the number of sampling localities and putative populations). The lowest cross validation error value was used to determine the most likely number of populations (*K*) present in the dataset. Non-model methods included using R to run a principal component analysis (PCA) and a discriminant analysis of principal components (DAPC) using the packages *FactoMineR* and *adegenet* (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011; Francois et al. 2015). The number of clusters with the lowest Bayesian information criterion (BIC) value in DAPC was considered the most statistically supported number of populations in the dataset. Both PCA and DAPC were then re-run with just *T. t. truncatus* and *T. t. gephyreus* individuals to further investigate the subspecies level division between these lineages. *Arlequin* was then used to carry out an analysis of molecular variance (AMOVA) testing the level of genomic variance explained by lineage division compared to sampling location.

2.5.6 Genomic Basis of Ecotype Formation

2.5.6.1 Candidate Loci Detection

Two outlier loci detection methods (FDIST and RandomForest) were used to investigate the genomic basis of ecotype formation in bottlenose dolphins. The coalescent-based FDIST method (Beaumont and Nichols 1996) in *Arlequin* was run under the hierarchical island model with 100,000 simulations and 100 demes. The number of groups was set to the number of sampling locations, plus one. Using the *p.adjust* function in the R package *plyr* (Wickham 2011), p-values were false discovery rate (FDR) corrected to avoid biases due to multiple testing. Loci with a FDR <10% (i.e. a q-value of <0.1) were classified as candidates for being

under selection. RandomForest was implemented in R using the *rfPermute* and *randomForest* packages (v.4.6-14) (Breiman 2001; Archer 2016). The *na.roughfix* function was used to impute missing data before beginning the analysis. The RandomForest method builds classification and regression trees whereby each split in the tree corresponds to the SNP that best divides the data into the defined groups. An importance value is assigned to each individual SNP, which is then averaged across all trees to determine which loci best explain the overall variation between the populations. The permutation method was used to calculate significance values for each SNP to statistically assess the likelihood of that SNP being a candidate for selection (see Brieuc et al. 2018). RandomForest was implemented with 125,000 trees and default settings for the proximity and importance parameters. The number of randomly chosen SNPs tested for each split of the tree (mtry) was set to the value that minimised the out-of-bag error rate and computational time (as suggested by Brieuc et al. 2018). Candidate loci were selected by plotting importance value distributions and selecting those SNPs above the upper elbow of the distribution curve as candidates as per Batley et al. (2018).

Both *Arlequin* and RandomForest were first run on pairwise comparisons of offshore *T. t. truncatus* with each of the three inshore lineages (i.e. *T. t. gephyreus*, *T. aduncus* and SABD). The combined lists of candidate loci for each of these comparisons were then compared and loci identified as outliers in all three lists were selected for further analysis. These candidates are putatively under selection in each instance of inshore ecotype evolution in the Southern Hemisphere and are therefore, potentially implicated in parallel genomic evolution of the inshore bottlenose dolphin ecotype. They will hereafter be referred to as the "parallel evolution candidates". The two outlier detection methods were then run to separately compare SWAO *T. t. truncatus* and *T. t. gephyreus.* This is likely the most recent ecotypic divergence in the study region, as informed by the results of this study and that of Moura et al. (2013). Candidate loci identified between SWAO *T. t. truncatus* and *T. t. gephyreus* will therefore, potentially reveal adaptations key to the early stages of colonisation of the inshore environment. These

candidates will hereafter be referred to as the "early-stage evolution candidates". Genotype frequencies were then calculated and plotted for all early-stage evolution candidates and for parallel evolution candidates with an F_{ST} value in the top 10%.

2.5.6.2 Functional Enrichment Analysis and Annotation

Flanking sequences for each SNP (300 bp either side) were extracted from the *T. aduncus* genome to carry out a functional enrichment analysis. A basic local alignment search tool (BLAST) was performed using blastn (Altschul et al. 1990; Sayers et al. 2019) from the nucleotide database available through NCBI on the 601 bp sequences of all 18,060 loci, using an expectation (e) value of 1E-6. All "blasted" loci were then mapped and annotated in Blast2GO with an e-value of 1E-3 (Conesa et al. 2005). A functional enrichment analysis using a Fisher's exact test to look for over- or under-representation of particular GO annotation terms in the parallel evolution candidate loci was then conducted in Blast2GO using an alpha value of 0.05. To further investigate the putative functions of the candidate loci and their associated genes, locus sequences were run in the NCBI web BLAST search against the *T. truncatus* genome assembly (NIST Tur_tru v1 Reference Annotation Release 101) (Altschul et al. 1990; Sayers et al. 2019). A threshold of an e-value of 1E-3 and an identity of >90% were used to select the most reliable candidates. Candidate genes were identified within 20 kilobases of the query sequence (as previously used for SABD; Batley et al. 2019). Putative gene functions were investigated in UniProtKB using the Swiss-Prot database (Boutet et al. 2007; The UniProt Consortium 2018).

2.6 Results

A total of 375 biopsy samples were ddRAD sequenced, encompassing three ocean basins and all currently recognised species and subspecies of bottlenose dolphin in the Southern Hemisphere, as well as the proposed *T. australis* (hereafter referred to as SABD) in inshore waters of southern Australia. For the dataset consisting of *Tursiops*, samples were sequenced across nine lanes with samples from Chapters 3 and 4, resulting in over 1.1 billion raw sequence reads. To remove issues created by low number of reads, individuals with <500,000 reads were removed, leaving an average of 3,274,483 reads per individual (standard deviation ± 2,758,909). The raw *Tursiops* dataset consisted of 196,751 SNPs. After a series of rigorous filtering steps, 18,112 SNPs and 353 individuals were retained (Appendix B: Table B.ii.2). These loci were then mapped to the *T. aduncus* reference genome, with a 99.71% alignment rate. The final *Tursiops* dataset available for analysis therefore, consisted of 18,060 SNPs (Appendix B: Table B.ii.2). The final individuals had an average of 6.6% missing data (standard deviation ± 5.6%).

The dataset including nine common dolphins used for phylogenomics analysis consisted of 386 individuals and 223,408 SNPs, with an average of 3,121,368 reads per individual (standard deviation ± 2,741,764) (Appendix B: Table B.ii.2). After filtering, 362 individuals were retained with an average of 7.0% missing data (standard deviation ± 6.2%) per individual. No common dolphins were removed during the filtering process. The 18,338 SNPs retained after filtering were then aligned to the *T. aduncus* reference genome at a rate of 99.69% and thus 18,282 SNPs were retained for phylogenomic analysis (Appendix B: Table B.ii.2).

2.6.1 Genomic Variation

Genomic diversity was estimated for each individual sampling site and then averaged across each of the four lineages. This was done to minimise the effect of small sample sizes in some localities and to better understand overall trends in diversity. *T. t. gephyreus* had substantially

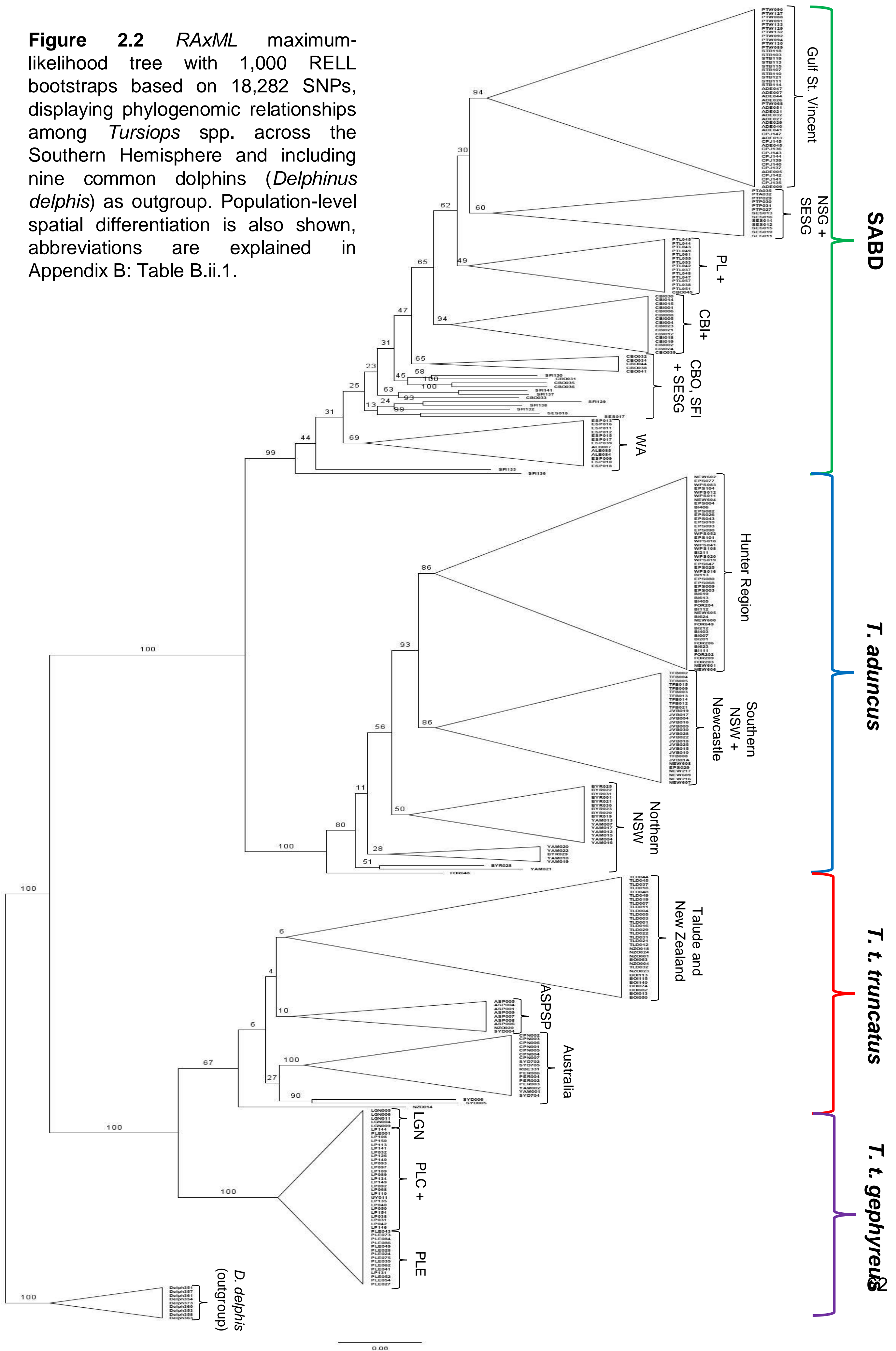
lower genomic diversity than the other taxa across all measures (e.g. H_E of 0.232 compared to an average of 0.325 across the three other lineages). Nonetheless, this lineage does not appear to have high levels of inbreeding (F_{IS} -0.024 compared to an average of 0.029 across the three other lineages) (Appendix B: Table B.ii.3). Relatively high genomic diversity was estimated for the other inshore lineages (*T. aduncus,* H_E of 0.315; SABD, H_E of 0.318), but offshore *T. t. truncatus* from across the Southern Hemisphere recorded slightly higher genomic diversity on average (H_E of 0.342). The number of private alleles was also lowest for *T. t. gephyreus*, with just 53 variants found uniquely in this taxon. *T. t. truncatus* on the other hand, had the highest number of private alleles at 642, while *T. aduncus* had substantially more than SABD (512 and 370, respectively) (Appendix B: Table B.ii.3).

2.6.2 Genomic Divergence

2.6.2.1 Phylogenomics

After 14 potential hybrids were removed the dataset contained 339 *Tursiops* spp. and nine *D. delphis*, and was based on 18,282 SNPs. A clear initial split between *T. aduncus/*SABD and *T. t. truncatus/T. t. gephyreus* was evident and supported by bootstrap values of 100% (Figure 2.2). This is consistent with current species classifications (Committee on Taxonomy of the Society for Marine Mammalogy 2019). There was subsequent strong genomic separation within each of these clades, with a similar level of divergence between SABD and *T. aduncus* and between *T. t. truncatus* and *T. t. gephyreus*. These divergences, including the recognised subspecies of *T. truncatus*, were also supported by bootstrap values of 100%. Lower levels of sub-population divergence corresponding to geographical regions were also evident within each lineage. Branch lengths were considerably shorter within the *T. t. gephyreus* lineage than for the other three lineages, suggestive of a more recent evolution (Figure 2.2).

Figure 2.2 *RAxML* maximum-likelihood tree with 1,000 RELL bootstraps based on 18,282 SNPs, displaying phylogenomic relationships among *Tursiops* spp. across the Southern Hemisphere and including nine common dolphins (*Delphinus delphis*) as outgroup. Population-level spatial differentiation is also shown, abbreviations are explained in Appendix B: Table B.ii.1.



2.6.2.2 Population Genomic Structure

Substantial genomic division among the taxa was revealed by PCA, with PC1 explaining 22.76% of the variance and splitting SABD/*T. aduncus* from *T. t. truncatus*/*T. t. gephyreus* (Figure 2.3 A). PC2 then showed the division between the two inshore Australian bottlenose dolphin lineages, SABD and *T. aduncus*, and a subtler divergence between *T. t. truncatus* and *T. t. gephyreus* (9.86% of variance explained). Fourteen individuals were detected to be closely associated with a taxon in a pattern inconsistent with the sampling location and/or observed morphology (see *Admixture* and PCA results). This suggests the potential presence of hybrid individuals in the dataset. DAPC strongly supported the divergence into the proposed four lineages in the Southern Hemisphere, with four being the most statistically supported number of genomic clusters (Figure 2.3B). When these methods were run again with only *T. t. truncatus* and *T. t. gephyreus* individuals, both PCA and DAPC showed clear separation of the two taxa (Appendix B: Figure B.i.1A and B). The distinction of the four lineages was further supported by AMOVA with 42.81% of the variance (p <0.001) being explained by among lineage divergence, compared to just 4.20% (p <0.001) of variance explained among populations within the four putative taxa (Appendix B: Table B.ii.4). Additional evidence for fine-scale subpopulation division within each of the lineages was provided by *Admixture* analysis (Appendix B: Figure B.i.2A-J). This highlights the power of this ddRADseq dataset to detect species, subspecies and population-level genomic differentiation. Substructure within *T. aduncus* and SABD will be addressed in subsequent chapters. The pattern of genomic divergence found here closely resembles that found by the phylogenomic analysis. The presence of 14 potential hybrid individuals was also further supported by *Admixture* analysis (Appendix B: Figure B.i.2A-J). In particular, two individuals from outer Coffin Bay in southern Australia and one individual from Eden in eastern Australia, were sampled within a group of SABD and *T. aduncus*, respectively, but show high estimated membership to the offshore *T. t. truncatus* ecotype. Estimates of F_{ST} were in general moderate to high among sampling localities, with a global average of 0.3604 (Figure 2.4). This particularly highlighted the divergence of *T. t. gephyreus* from all other taxa, with the highest value of 0.8370 estimated

between the coastal Patos Lagoon (PLC) *T. t. gephyreus* community in SWAO and the SABD population in northern Spencer Gulf (NSG), southern Australia. When averaged among lineages, the mean of estimates between *T. t. gephyreus* and *T. t. truncatus* was substantially higher than those between *T. aduncus* and SABD, at 0.5398 and 0.2857, respectively (Figure 2.4).

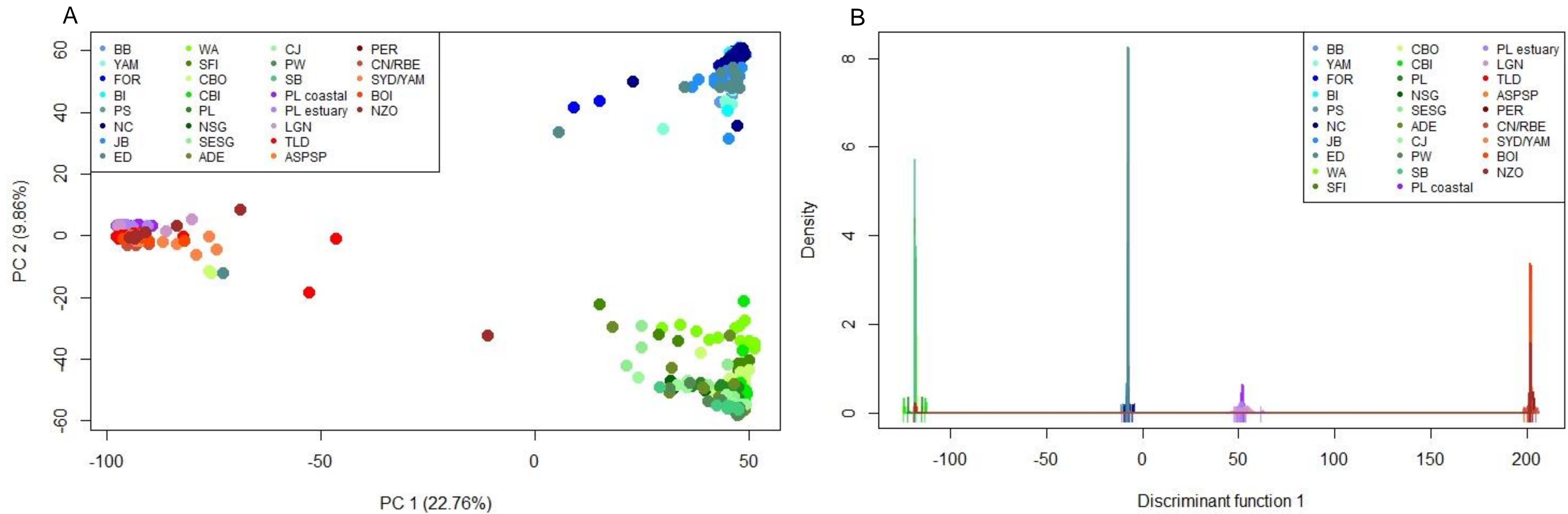


Figure 2.3 Population genomic structure of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere based on 18,060 SNPs as estimated by **A)** principal component analysis (PCA) and **B)** discriminant analysis of principal components (DAPC) with four clusters being the most supported number of 'populations'. Sampling locations are coloured as per putative lineage: *T. aduncus* (blue shades), SABD (green shades), *T. t. gephyreus* (purple shades), *T. t. truncatus* (red shades). Sampling location abbreviations are explained in Appendix B: Table B.ii.1.

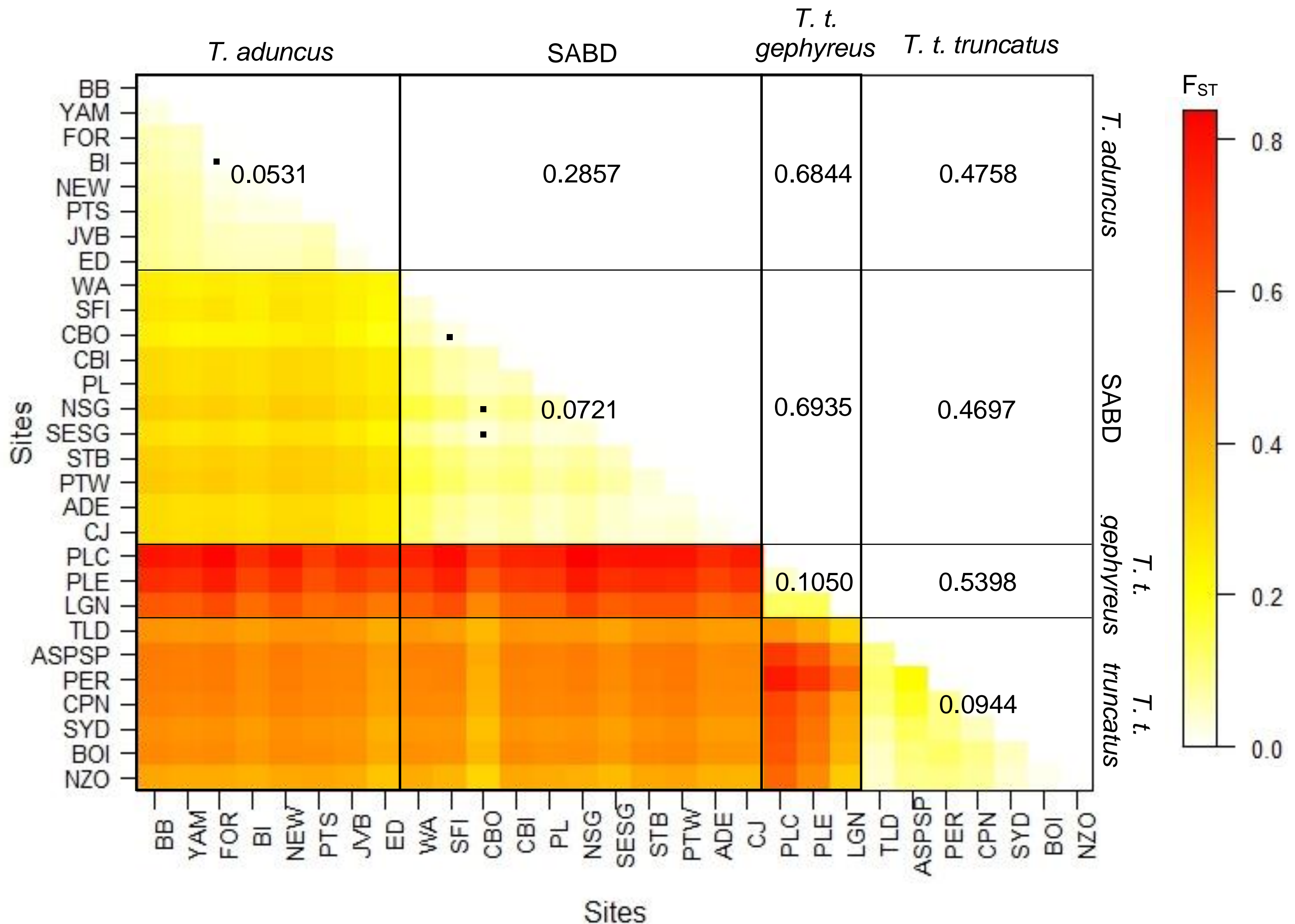


Figure 2.4 Heat map of pairwise genomic differentiation (F_{ST}) between sampling sites of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere as estimated by *Arlequin* based on 18,060 SNPs. Values on the diagonal represent the average F_{ST} value for comparisons within each putative lineage, while those in the top half of the matrix represent the average value of pairwise comparisons between each lineage. Non-significant F_{ST} values at the B-Y corrected alpha value of 0.0076 are marked with a black square (▪). Transitions between putative lineages are marked by black lines. Average F_{ST} was 0.3604. Sampling location abbreviations are explained in Appendix B: Table B.ii.1.

2.6.3 *Genomic Basis of Ecotype Formation*

2.6.3.1 Candidate Loci Detection

The two outlier methods identified a total of 325 outliers as candidates for selection between *T. t. truncatus* and *T. t. gephyreus*, 1,126 outliers between *T. t. truncatus* and *T. aduncus* and 842 outliers between *T. t. truncatus* and SABD. The lists of candidate loci were then compared to identify those SNPs that were present in all three, being potentially implicated in parallel

genomic evolution of the inshore ecotype across the Southern Hemisphere. This resulted in a total of 142 candidates for parallel evolution. Fourteen candidate genes were highlighted as having an F_{ST} value in the top 10%. Genotype frequencies for these candidates were then plotted for each putative taxon, with stark differences in distributions revealed between the inshore and offshore lineages. Across the three inshore lineages each top candidate locus had an average of 98.63% homozygote individuals (Figure 2.5). For 11 out of the 14 top candidates, this reflected near-fixation of the major allele in each of the inshore putative taxa. For the other three candidates, it led to fixation of the minor allele in *T. t. gephyreus* while the major allele dominates in *T. aduncus* and SABD. In the offshore dolphins on the other hand, heterozygosity was much higher with each top candidate locus having an average of 76.62% homozygote individuals and a much higher representation of the minor allele (Figure 2.5).

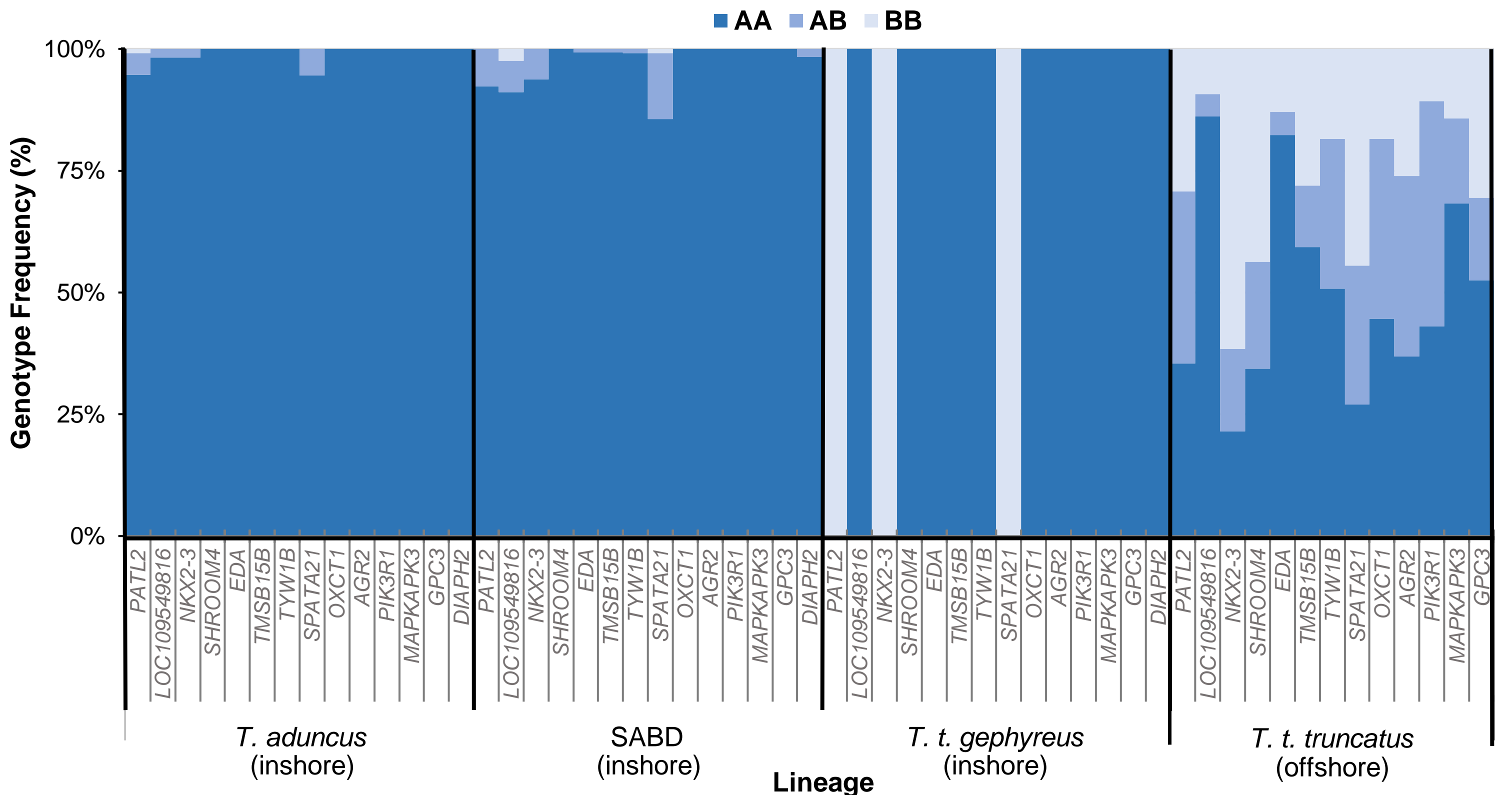


Figure 2.5 Genotype distribution for the parallel evolution candidate genes with F_{ST} values in the top 10% when comparing each inshore bottlenose dolphin lineage (*Tursiops* spp.) to offshore *T. t. truncatus*.

Arlequin and RandomForest identified 12 loci as candidates in the most recent divergence of SWAO *T. t. truncatus* and *T. t. gephyreus* (i.e. early-stage evolution candidates). Genotype frequencies were then plotted for each of the six annotated candidate genes, revealing high homozygosity in the inshore ecotype (Figure 2.6). In *T. t. gephyreus*, across each of these genes, an average of 99.6% of individuals were homozygous at the major allele, with almost complete absence of the minor allele. The offshore SWAO dolphins on the other hand, had an average of 82.7% homozygous individuals, primarily representing the frequency of the minor allele (Figure 2.6).

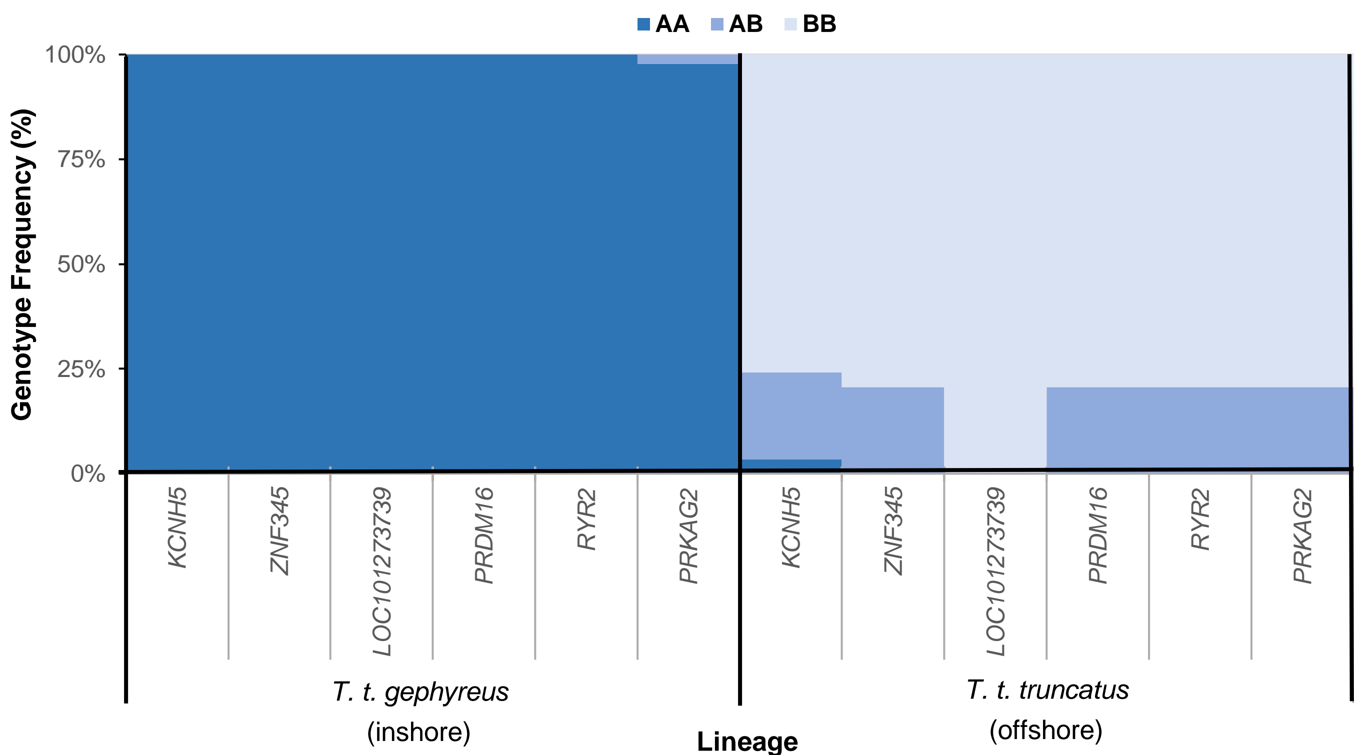


Figure 2.6 Genotype distribution for the six candidate genes implicated in early-stage evolution between offshore *T. t. truncatus* and inshore *T. t. gephyreus* in the southwestern Atlantic Ocean.

2.6.3.2 Functional Enrichment Analysis and Annotation

Of the 18,060 loci a total of 3,792 (20.99%) loci scored BLAST hits and were mapped and annotated. Twenty-seven of these loci were candidates for parallel evolution. A functional enrichment analysis found 90 categories significantly over-enriched in the parallel evolution candidate set (Appendix B: Table B.ii.5). This included glycosaminoglycan metabolic process

(GO:0030203), glycogen debranching enzyme activity (GO:0004133), walking behaviour (GO:0090659), mesonephric duct morphogenesis (GO:0072180), carbohydrate transport (GO:0008643), insulin binding (GO:0043559) and photoreceptor activity (GO:0009881), among many others (significance values provided in Appendix B: Table B.ii.5). Individual annotation of parallel evolution candidate loci identified 97 candidate genes as being within 20 KB of the respective locus as above (Appendix B: Table B.ii.6), while six candidate genes were identified from the early-stage evolution candidate loci (Appendix B: Table B.ii.7). Candidate genes were discovered to be related to several major bodily systems however, only candidates associated with the most important and relevant processes will be discussed in detail (but see Appendix B: Table B.ii.6 and 7 for full list of candidates).

2.7 Discussion

Since the Eocene, large-scale environmental and oceanographic restructuring in the world's oceans has influenced the rapid diversification of cetaceans (Steeman et al. 2009). With climate change now dramatically altering marine habitats worldwide, it is important to understand the principal drivers of genomic divergence and adaptation in marine organisms and how this may be affected with ongoing anthropogenic disturbance. The inshore-offshore pairs of bottlenose dolphin ecotypes in the Southern Hemisphere provides an excellent system to investigate genomic pathways that are particularly important to the adaptation and diversification of delphinids. A genomic dataset of over 18,000 high-quality SNPs was utilised to investigate some of the controversial phylogenomic relationships and genomic divergence within the genus *Tursiops*, as well as to identify evidence for environmental adaptation in these lineages. The hypothesis of genomic differentiation between ecotypes, as well as among the inshore lineages, was supported. This was reflected in proposed subspecies level divisions within the two currently recognised species, suggesting that ecotypic differentiation can lead to incipient speciation. Furthermore, modification to several major physiological systems were detected in the early stages of ecotypic divergence and were repeatedly selected for in the adaptation of inshore lineages across the Southern Hemisphere. The signal for selection found in the genes associated with these major bodily systems provides evidence for the adaptation of inshore bottlenose dolphins to their respective habitats, which may be affected by future environmental changes. These results highlight potentially critical adaptive pathways for marine vertebrates to successfully colonise new niche spaces, something that is becoming increasingly important in the face of ongoing climate change.

2.7.1 Genomic Variation

Variation in levels of genomic diversity among species can be caused by a series of factors, including demographic history (Ellegren and Galtier 2016). Founder events in particular can lead to low genomic diversity in a population. This occurs when a small group of individuals

colonises a new habitat resulting in limited gene pool for population expansion and an enhanced effect of genetic drift (Pardo et al. 2005; Ellegren and Galtier 2016). Thought to be a result of such founder events, inshore populations of bottlenose dolphins have been repeatedly reported to have substantially lower genetic diversity than their offshore counterparts (Hoelzel et al. 1998; Natoli et al. 2004; Lowther-Thieleking et al. 2015; Fruet et al. 2017). This was particularly true for *T. t. gephyreus* from the SWAO. This is consistent with previous estimates based on mitochondrial DNA (mtDNA) and microsatellite markers, showing that the SWAO lineage has the lowest genetic diversity recorded for any bottlenose dolphin population worldwide, being three times lower than the diversity of the adjacent offshore population (Fruet et al. 2014b; Fruet et al. 2017). The genomic data supports the hypothesis of a strong founder effect likely associated with the colonisation of inshore SWAO after the LGM (Fruet et al. 2017).

Current hypotheses for *Tursiops* diversification and radiation suggest a coastal Australasian origin for the genus, with subsequent colonisation of the pelagic realm and then repeated movement back into inshore habitats as the genus spread throughout the world's oceans (Moura et al. 2013). This is consistent with our finding of substantially higher genomic diversity in the Australian inshore lineages than in *T. t. gephyreus*, potentially reflecting differing colonisation and demographic histories. A substantially lower number of private alleles in *T. t. gephyreus* and a phylogeny with shorter branch lengths compared to other lineages also support this hypothesis, suggesting a more recent divergence from *T. t. truncatus* than for Australian inshore lineages. The small population size, restricted geographical range (Secchi 2007) and low genomic diversity makes the *T. t. gephyreus* lineage particularly susceptible to anthropogenic disturbances. Indeed, the International Union for the Conservation of Nature (IUCN) has recently classified them as Vulnerable due to high anthropogenic impacts and declining habitat quality (Vermeulen et al. 2019). Ongoing habitat degradation and other human impacts, such as bycatch, are likely to have major negative consequences for these dolphins (Secchi 2007; Daura-Jorge and Simões-Lopes 2011; Fruet et al. 2012). Small

cetaceans are apex predators and therefore, disturbances to their populations will undoubtedly have flow on effects to the ecosystems they inhabit.

2.7.2 Genomic Divergence

The taxonomy of cetaceans has long been a controversial topic. This is particularly true for the Delphinidae family and the classification of *Tursiops* species and their close relatives. Relatively recent species radiations resulting in incomplete lineage sorting and potential for hybridisation between delphinine species has created discordance between mtDNA and nuclear DNA markers, fuelling much of the debate (LeDuc et al. 1999; Amaral et al. 2012b; Moura et al. 2013; Gray et al. 2018). Clear genomic divergence was evident among the four sampled bottlenose dolphin lineages. The differentiation between SABD and *T. aduncus*, which are both currently recognised as *T. aduncus*, was at a similar (if not higher) level to that between *T. t. truncatus* and *T. t. gephyreus.* This was evident in both phylogenomic and population structure analyses. Combined with known morphological and osteological dissimilarities, as well as genetic differentiation (Charlton et al. 2006; Möller et al. 2008; Charlton-Robb et al. 2011), the SABD lineage, which was previously proposed as a separate species (the Burrunan dolphin, *T. australis*) (Charlton-Robb et al. 2011), seems to warrant subspecies level classification within *T. aduncus*. Based on the definitions given for species, subspecies and evolutionarily significant units in Taylor et al. (2017), a conservative subspecies classification is deemed most appropriate for both SABD and *T. t. gephyreus*. With many widely distributed, highly mobile cetacean species now known to exhibit strong population genetic structure at odds with their dispersal potential (see Hoelzel 1998; Hoelzel 2009), it is important to revisit the species classifications through studies of genomic divergence with next-generation sequencing. These techniques provide much greater power to detect patterns of differentiation and gain novel insights into the underlying causes of such divergences, which may need to be re-evaluated for several cetacean species, as well as potentially many other marine organisms.

It is thought that the coastal Australasian *Tursiops* ancestor colonised the pelagic environment and subsequently spread across the world's oceans and back into inshore habitats in several regions (Moura et al. 2013). Conflicting evidence exists surrounding the history of this divergence, with SABD initially suggested to be the ancestral lineage (Moura et al. 2013; Gray et al. 2018), but more recently found to be a sister group to *T. aduncus* (Moura et al. 2020). Support is provided here for the latter. Nevertheless, the ancestral coastal *Tursiops* form likely diverged into the pelagic *T. truncatus* form within the last million years (Gray et al. 2018). This is further supported by strong genomic divergence between *T. t. truncatus* and both *T. aduncus* and SABD (and longer phylogeny branch lengths within each of these lineages than in *T. t. gephyreus*). Further investigation is however, needed to better assess the divergence time between *T. t. truncatus* and the two inshore Australian lineages. Subsequent recolonisations of inshore environments around the world by *T. truncatus* are likely to have occurred with the opening of new coastal habitats with sea level rise after the LGM (Louis et al. 2014a; Louis et al. 2014b; Nykänen et al. 2019). This has led to several cases of inshore-offshore pairs of bottlenose dolphin ecotypes at varying stages of divergence, including the relatively recent formation of the inshore subspecies *T. t. gephyreus* in SWAO. The coastal Indo-Pacific form, *T. aduncus*, appears to be divided into several genomic stocks, including SABD in southern Australia, an Australasian lineage in eastern and western Australia and the Indo-Pacific region and other lineages along eastern Africa and in the Arabian Sea (Amaral et al. 2017; Gray et al. 2018), with further population subdivision within regions (e.g. Wiszniewski et al. 2010; Pratt et al. 2018). It should be noted that coastal bottlenose dolphins from eastern Africa represent the holotype lineage of *T. aduncus* (see Perrin et al. 2007), thought to have diverged from other *T. aduncus* ~342 - 327 thousand years ago (Moura et al. 2013; Gray et al. 2018). Samples from this region and the Arabian Sea were not available for this study.

While strong population structure is evident within the inshore lineages, the offshore form appears to maintain high gene flow throughout the Southern Hemisphere. Offshore bottlenose dolphins from across three ocean basins were found to be more genomically similar to each other than to the adjacent inshore populations. This is despite sightings of mixed groups with inshore dolphins in some regions (e.g. Fruet et al. 2017). With shared mtDNA haplotypes reported between the offshore dolphins from ASPSP in SWAO and those in the northeastern Atlantic Ocean (Querouil et al. 2007; de Oliveira et al. 2019), the pelagic connectivity may also extend between the two hemispheres. This study provides strong support for negligible reproductive exchange between inshore and offshore bottlenose dolphins in SWAO (but see de Oliveira et al. 2019), a conclusion reinforced by major morphological and osteological differences between them (Costa et al. 2016; Wickert et al. 2016; Fruet et al. 2017). Previous findings of substantially stronger population genetic structure in inshore populations than offshore are also confirmed here (Querouil et al. 2007; Hoelzel 2009; Möller 2012), although connectivity of the pelagic ecotype has not been documented on this scale previously (see Tezanos-Pinto et al. 2009). Population genomic structure on a much finer scale in the inshore than the offshore ecotype suggests social structure, natal philopatry and local adaptation may play a stronger role in inshore dolphin communities (Möller 2012). With repeated inshore colonisations in the genus *Tursiops*, currently at varying stages of divergence, this system provides a unique opportunity to investigate adaptations of delphinids to the inshore environment.

2.7.3 Genomic Basis of Ecotype Formation

Large-scale environmental changes have driven several mass extinctions and species radiations in the marine ecosystem over evolutionary history (Condamine et al. 2013). One of the most well documented examples of this is the effect of the Pleistocene climatic oscillations on marine organisms (e.g. Steeman et al. 2009). After restricted dispersal created by low sea levels during the glacial low-stand in the LGM, subsequent sea level rise during global

warming released coastal habitats worldwide providing opportunities for inshore colonisation (e.g. Portnoy et al. 2014; Silva et al. 2014). It is thought that during this time pelagic bottlenose dolphins moved back into the inshore environment to inhabit recently submerged coastal habitats (Moura et al. 2013; Louis et al. 2014b). By investigating how marine species, including dolphins, have historically adapted to new environments, we can better understand how ongoing climatic changes may impact adaptation, persistence and speciation now and into the future.

Typically, the inshore environment differs from the offshore in several key facets. Aside from the obvious differences in water depth, the inshore habitat may be more complex and resources in the area may be more reliable and evenly distributed (see Möller 2012). Fish assemblages can differ greatly between the two habitats, influenced by productivity, turbidity and salinity gradients, among other factors (e.g. Ooi and Chong 2011; Kopp et al. 2015). Similarities across inshore environments are likely to create similar selective pressures and may result in parallel adaptive evolution of taxa colonising and inhabiting these areas.

2.7.3.1 Ecotype Evolution

This study investigated the adaptive pathways that may be driving the relatively recent divergence of *T. t. gephyreus* from *T. t. truncatus*, as well as establishing support for potential parallel evolution of inshore bottlenose dolphin lineages across the Southern Hemisphere. This established the presence of genomic adaptations that are potentially crucial to both the short and long-term persistence of populations in inshore habitats and may be causing these lineages to diverge over time. Selective pressures in the inshore habitat appear to be driving strong directional selection in these lineages. Directional selection is a form of natural selection whereby favourable alleles are pushed toward fixation creating a reduction in genomic variation (Endler 1986). In inshore SWAO this is observed by almost complete fixation of the major allele in all six early-stage evolution candidate genes. In the offshore SWAO ecotype however, major allele homozygotes were almost completely absent, and heterozygosity was

much higher. This likely reflects divergent selective pressures between the two habitats and may indicate that balancing selection has a stronger role in the adaptation of offshore dolphins. With balancing selection actively retaining genomic variation in a population (Hedrick 2007), this may be important in allowing offshore delphinids to be suited to the wide range of habitat types and environmental conditions experienced over their typically much larger home ranges than of inshore dolphins (see Möller 2012). A very similar pattern was also detected for the top 10% of parallel evolution candidates, with almost complete homozygosity in inshore individuals at these loci and much higher heterozygosity in the offshore population. Furthermore, for 11 of the 14 top candidates there was near fixation of the same allele for each inshore lineage across three ocean basins. The replication of the same pattern in all three sampled inshore bottlenose dolphin lineages indicates that selective pressures may be similar across the inshore habitats, creating parallelism in the adaptive responses of these dolphins. Following the pattern of colonisation proposed for bottlenose dolphins, it may be that particularly favourable alleles for these candidates were in high frequency in the ancestral Indo-Pacific coastal population but lost their adaptive advantage in the colonisation of pelagic waters. Allele frequencies would have subsequently changed to allow offshore dolphins to become better adapted to a wide range of environmental conditions. Upon the recolonisation of the inshore habitat however, changes in selective pressures may have caused the original high frequency allele to again become the most adaptively favoured, pushing it toward fixation and creating a pattern of parallel evolution in the inshore ecotypes across the Southern Hemisphere. However, further investigation is needed to assess this hypothesis. This provides important information regarding the adaptation of inshore cetaceans, highlighting major adaptive pathways that are repeatedly being selected for across three ocean basins in the Southern Hemisphere.

2.7.3.2 Cardiovascular and Circulatory Systems

A number of candidate genes were found to be associated with adaptation of the cardiovascular and circulatory systems. In particular, allele frequency differences in the early-stage evolution candidate genes *PRKAG2* and *RYR2* suggests adaptation of the cardiac system is important for initial survival in the inshore habitat (for specific gene functions see Ahmad et al. 2005; Toischer et al. 2013). Several additional genes were implicated in parallel evolution of the inshore ecotype across the Southern Hemisphere – *CACNA1, JDP2, MYH11, NMRAL1, PDE1C, PDE9A*, *PLAT, PRKG1, RBM20, SEMA3E* and *TBX1*. These genes are involved in heart and blood vessel development and healthy functioning (Xu et al. 2004; Guo et al. 2012; Sakurai et al. 2012; Guo et al. 2013; Groen et al. 2014; Lee et al. 2015; Heger et al. 2018), heart muscle contraction (Kuang et al. 2012), haemoglobin concentration (Simonson et al. 2015) and in regulation of blood clotting (Browne et al. 1985; Schuliga et al. 2013). The *PRK* gene family (hosphoribulokinase/uridine kinases) appears to be particularly important, found in both the early-stage and parallel evolution analyses. *PRK* genes have also been found to be involved in the macroevolution of marine mammals to an aquatic lifestyle, particularly in regard to the nervous system (Foote et al. 2015; Zhou et al. 2015). Adaptation of the cetacean heart and circulatory system is possibly a reflection of a change in diving behaviour between the inshore and offshore dolphins. When diving, these animals cease breathing, reduce their heart rate and redirect blood oxygen supply from nonessential organs to the brain, heart and exercising muscles (Mirceta et al. 2013). Extended deep dives therefore, put significant stress on the body and often result in hypoxic conditions, whereby there is a reduction in oxygen delivery to the blood and tissues and potential DNA damage (see Mirceta et al. 2013; Tian et al. 2016). Accordingly, we found three parallel evolution candidate genes that are involved in DNA damage response – *DYRK1A, UBE2E2* and *USP10* (Zhang et al. 2016; Mizukami et al. 2017; Guard et al. 2019). The physiological adaptations of cetaceans to deep diving have been studied extensively at the macroevolutionary scale. Genomic studies have revealed positive selection on genes associated with cardiovascular system formation and regulation (McGowen et al. 2012; Nery et al. 2013b; Foote et al. 2015),

hypoxia tolerance (Yim et al. 2014; Tian et al. 2016), DNA repair and damage response (Zhou et al. 2013) and oxygen storage (McGowen et al. 2014) in the cetacean lineage. Storage of oxygen in the muscle and blood is essential for deep diving organisms and significantly higher blood volume and haemoglobin and myoglobin concentration have been recorded in many deep-diving species compared to their terrestrial and shallow-diving counterparts (Baldwin 1988; Kooyman and Ponganis 1998; Mirceta et al. 2013). Myoglobin and haemoglobin facilitate oxygen storage and transport in the muscles and have important roles in protecting the brain from hypoxic conditions (see Mirceta et al. 2013; Nery et al. 2013a). Haemoglobin levels and red blood cell counts have been reported to be significantly higher in offshore than inshore bottlenose dolphins in the Pacific and Atlantic Oceans, making their bodies more efficient at transporting and storing oxygen (Duffield et al. 1983; Hersh and Duffield 1990). Studies of the microevolutionary adaptations of dolphins to different diving behaviours have not previously broached the genomic basis to this evolution. There are however, many studies addressing adaptations of the cardiovascular system in relation to the evolution of hypoxia tolerance in other marine and terrestrial species (e.g. Kooyman and Ponganis 1998; Drabek and Burns 2002; Johnson et al. 2004). Several genes have been found to be involved in preventing hypoxia in high-altitude human populations (e.g. Simonson et al. 2015; Stobdan et al. 2015). Simonson et al. (2015) identified *NMRAL1,* a gene also found here, to be associated with haemoglobin concentration changes in humans from Tibet. Adaptation of the cardiovascular and circulatory systems may therefore, be a crucial step in the successful colonisation of, and persistence in, the inshore habitat by bottlenose dolphins and for dealing with changes to hypoxia-inducing behaviours in general. Furthermore, *RYR2* function has been shown to be affected by diet, as it is inhibited by excessive levels of the essential omega-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (Ismail 2005). These molecules are found in particularly high levels in oily fish such as herring, mackerel, salmon, tuna and sardines (Ismail 2005). Food webs in the freshwater ecosystem have very little available DHA compared to pelagic marine environments, a feature which has caused genomic adaptation in marine fish colonising the freshwater realm (Ishikawa et al. 2019).

Furthermore, the AMPK enzyme associated with *PRKAG2* is thought to regulate feeding behaviour and fatty acid and cholesterol metabolism in response to reduced energy levels (Carling 2004; Chantranupong et al. 2015). It is therefore, possible that cardiovascular adaptations of dolphins to the inshore environment are not only associated with changes in diving behaviour and exposure to hypoxic conditions, but also with differences in levels of omega-3 fatty acids in the ecotype diets.

2.7.3.3 Gastrointestinal System and Mouth

Modifications of the gastrointestinal system and mouth have occurred extensively across the evolution of both terrestrial and marine organisms. A vast diversity can be seen in the digestive systems and dentition of mammals depending on their food preferences. For example, the ruminant digestive system of cows and sheep is in stark contrast to that of closely related carnivores (Hume 2002). Vastly different shapes, strengths and sizes of teeth and jaw musculature are also seen within taxa depending on the target prey (e.g. in pinnipeds and elasmobranchs) (Motta and Huber 2004; Kienle and Berta 2016). Perhaps the most dramatic contrasts in dentition and prey choice within a taxon are exemplified by the cetaceans. This group is split into two parvorders based on dentition - Odontoceti (toothed whales and dolphins) and Mysticeti (baleen whales). Extensive diversity in jaw structure and the number, size and shape of teeth is also seen among odontocetes worldwide and is likely to have evolved as a response to differences in foraging strategies and preferred prey items (Werth 2006, 2007). Several genes potentially under parallel selection in the inshore bottlenose dolphin ecotype with functions related to these systems were found here. Namely, candidate genes *AGR2, BTC, EDA, GTF2IRD1*, *NKX2-3* and *TBX1* were discovered with broad functions related to the development of the jaw, mouth and teeth (Ohazama and Sharpe 2007; Catón et al. 2009; Han et al. 2018), stomach and gastrointestinal system (Howarth et al. 2003; Lin et al. 2016) and salivary glands (Mikkola and Thesleff 2003), as well as mucous secretion in the intestine (Park et al. 2009). Differing rostrum length and tooth size and number exists between

inshore and offshore bottlenose dolphins in the Southern Hemisphere, although the specific differences between ecotypes are not consistent across ocean basins (Hale et al. 2000; Kemper 2004; Wickert et al. 2016). Modification of these structures and systems is likely driven by differences in diet between the two ecotypes. While formal comparison of the diet of these dolphins has not yet been completed in SWAO and eastern Australia (but see Wiszniewski et al. 2009; Costa et al. 2016), stable isotope analysis provides evidence for dissimilarity in prey choice between inshore SABD and offshore *T. truncatus* in southern Australia (Gibbs et al. 2011). Dietary differences between inshore and offshore bottlenose dolphin ecotypes has also been documented in other regions around the world, including in the northeastern Pacific and northwestern Atlantic Oceans (Mead and Potter 1995; Diaz-Gamboa et al. 2018). Repeated selection on the candidate genes in the Southern Hemisphere inshore environment is therefore, potentially in response to differences in diet and may reflect broader adaptations of the mouth and gastrointestinal system than have been previously documented in phenotypic studies.

2.7.3.4 Adipogenesis and Energy Production

Fat reserves play a number of crucial roles in the survival of animals. For example, fat storage has critical importance in thermoregulation (Speakman 2018), buoyancy (Hagen et al. 2000), metabolism and energy production (Choe et al. 2016). In the short-term fat reserves can be affected by diet and activity levels (Schrauwen and Westerterp 2000). Genome level adaptations of this system however, may be largely dictated by long-term changes in temperature and diet. This has been extensively documented in cetaceans where the transition from a terrestrial to a fully aquatic lifestyle was coupled with major dietary changes and alteration of thermogenic requirements. Accordingly, several previous studies have found positively selected genes related to fat storage, lipid transport, metabolism and fatty acid synthesis and transport in cetaceans (McGowen et al. 2012; Sun et al. 2012; Nery et al. 2013b; Wang et al. 2015; Endo et al. 2018; Derous et al. 2019). Wang et al. (2015) documented the

importance of triacylglycerol adipose tissue to cetaceans, finding associated positive selection in the *PDE3B* gene, among others. Two members of this gene family were found here to be potentially involved in the parallel evolution of the inshore ecotype. In particular, *PDE1C*, which has functions in healthy heart functioning also has an important role in energy production (Han et al. 1999). Several other genes – *AGL, COX8A, GPC3, JDP2, NSDHL, OXCT1* and *RORA* – were discovered to be implicated in the parallel evolution of energy production pathways, as well as in adipogenesis and fat storage, and several associated processes (Caldas and Herman 2003; Nakade et al. 2007; Lau et al. 2008; Vial et al. 2011; Anushiravani et al. 2017; Li et al. 2018; Zeng et al. 2019). In addition, several associated GO terms were significantly over-enriched in the candidate gene dataset, including glycosaminoglycan and aminoglycan metabolic, catabolic and biosynthetic processes, carbohydrate transport and insulin binding, among several others. *PRDM16* was also identified as an early-stage evolution candidate, which has a key role in deposition of brown adipose tissue (Seale et al. 2007; Seale et al. 2008). Brown adipose tissue is essential for breakdown of food into energy for thermogenesis and/or lipid and glucose metabolism (Cannon and Nedergaard 2004) and can be affected by the amount of protein and fatty acids (e.g. EPA and DHA) in the diet (Cannon and Nedergaard 2004; Zhuang et al. 2019). It could also be crucial to the long-term adaptation of organisms to cold temperatures (Cannon and Nedergaard 2004; Li et al. 2014), with brown adipose tissue volume found to be significantly greater in human populations residing in chronically cold climates compared to those in warmer areas (Sazzini et al. 2014). Changes to adipogenesis and lipid and glucose metabolism have also been previously found to be important in microevolutionary adaptations of marine mammals. For example, these pathways were shown to be under differential selection between killer whale (*Orcinus orca*) ecotypes found in opposing climates and feeding on different diets (Foote et al. 2016), as well as between polar bears (*Ursus maritimus*) and brown bears (*U. arctos*; Liu et al. 2014). Similarly, lipid reserve level and body mass of southern elephant seals (*Mirounga leonina*) was shown to differ depending on their respective foraging zones, corresponding to different prey assemblages and thermal climates (Richard et al. 2016). Temperature profiles can potentially differ between

inshore and offshore habitats and coupled with discrepancy in diet and total body size between the two ecotypes (e.g. Charlton-Robb et al. 2011; Gibbs et al. 2011; Costa et al. 2016), this may create opposing thermogenic and energy production requirements for the dolphins. Genes associated with these processes may therefore, become increasingly important for the survival of species under anthropogenic climate change and ocean warming. Fat is also an important part of the sensory system for odontocete cetaceans, with specialised fat storages involved in echolocation (Gabler et al. 2018). It is therefore, possible that modification to adipogenic pathways could also be associated with changes in echolocation for inshore bottlenose dolphins.

2.7.3.5 Sensory Systems

Echolocation is one of the most remarkable adaptations made by odontocete cetaceans in their radiation throughout the world's oceans. This ability evolved around 34-36 MYA (Steeman et al. 2009), corresponding with changes to, and inactivation of, genes associated with the aquatic eye (McGowen et al. 2014). It has been suggested that evolution of the auditory system in these animals was coupled with a reduced reliance on the visual system, as echolocation proved more efficient for hunting in an aquatic setting (see McGowen et al. 2014). Further modifications to the aquatic eye and specialisation of echolocation among Odontoceti lineages has subsequently occurred (Au 2000; Meredith et al. 2013). Different clicking patterns have been observed between *T. aduncus* and *T. truncatus* (Wahlberg et al. 2011), and freshwater and oceanic dolphins (Gutstein et al. 2014), as well as between sympatric fish- and mammal-eating killer whale ecotypes; the latter thought to be as an adaptation to their different respective prey items (Barrett-Lennard et al. 1996). A number of other factors may affect the echolocation activity of odontocetes, including habitat complexity, water visibility, boat disturbance, behavioural and social context, the level of ambient noise and the body size of the animal (Wahlberg et al. 2011; Samarra and Miller 2015; La Manna et al. 2019). These factors are known to differ between the inshore and offshore habitats and

respective bottlenose dolphin ecotypes (see Möller et al. 2011; Costa et al. 2016; Diaz-Gamboa et al. 2018 among others), and could potentially be driving divergent adaptation of the auditory system. This is further supported by osteological differences between the ecotypes, with modification of the pterygoids between inshore and offshore dolphins (Kemper 2004; Perrin et al. 2011; Wickert et al. 2016). Perrin et al. (2011) hypothesised that this was reflective of differences in echolocation behaviour between the inshore and offshore forms. The genomic basis of echolocation differences between cetacean ecotypes has however, not been investigated until now. Here, three genes related to the development of ear structures and the perception of sound were found to be potentially implicated in the parallel evolution of the inshore ecotype – *CDH23, SEMA3E* and *TBX1* (Vitelli et al. 2003; Holme and Steel 2004; Fekete and Campero 2007). *CDH23* has previously been found to be associated with the development of high frequency hearing and echolocation in the evolution of cetaceans and bottlenose dolphins specifically (Shen et al. 2012a; Zhou et al. 2013; McGowen et al. 2014). Specific mutations in *CDH23* can also lead to the development of a disorder characterised by hearing loss and visual impairment in humans (Bolz et al. 2001). Significant over-enrichment in the GO term photoreceptor activity was found, as well as selection on a number of other candidate genes with functions in eye development and healthy functioning – *BTC, FREM2, PARD3* and *RORA* (Wei et al. 2004; Anand-Apte et al. 2010; Silveira et al. 2010; Zhang et al. 2019). *FREM2* in particular has a broad array of other functions, including potentially being involved in epidermal adaptations and kidney development (discussed below). Difference in the visual system of these dolphins has not been suggested on this scale before but could be influenced by differences in turbidity and light levels between the two habitats. This could also be affected by changes in diving behaviour between the ecotypes, with deep-diving pinniped species found to have bigger eyes that are more sensitive and able to adapt more quickly to low light levels than humans and their shallow-diving counterparts (Levenson and Schusterman 1999; Debey and Pyenson 2013). Structural modifications of both the auditory and visual systems of these dolphins thus, warrants further investigation.

2.7.3.6 Musculoskeletal and Integumentary Systems

While changes in the shape of the skull may be associated with echolocation, other adaptations of the musculoskeletal system are also present in bottlenose dolphins and other marine mammals. Modification of this system was of crucial importance to the colonisation of the aquatic system by marine mammals (de Buffrénil et al. 2010; Zhou et al. 2018a). Accordingly, a number of genes associated with muscle and bone development, particularly of the skull, were found to be significantly differentiated in each inshore-offshore comparison. This included the candidate genes *GPC3, GTF2IRD1*, *MBNL3, PIK3R1* and *SCUBE2* (for specific gene functions see Squillace et al. 2002; Tassabehji et al. 2005; Shinohara et al. 2012; Dwivedi et al. 2013; Lin et al. 2015). Skeletal differences between inshore and offshore bottlenose dolphins have been extensively researched, with the inshore ecotype typically having less vertebrae than those residing offshore (Hale et al. 2000; Kemper 2004; Wickert et al. 2016). *T. aduncus* and SABD have also been found to have shorter/smaller skulls than the offshore *T. truncatus* (Hale et al. 2000; Charlton-Robb et al. 2011). These differences are thought to be associated with opposing requirements for manoeuvrability and the manipulation of prey (Hersh and Duffield 1990; Perrin et al. 2011), but many of the skeletal modifications reported have unknown adaptive functions. Differences in bone density have also been found among cetacean species, likely as an adaptation to diving depth and the associated buoyancy requirements (see Foote et al. 2015). Subsequently, Zhou et al. (2018a) discovered several positively selected genes related to bone density in the common ancestor of cetaceans, identifying *PIK3R1* and another member of the *PIK* gene family (*PIK3CB)* to be highly correlated with different measures of bone compactness. While bone density changes have not been documented at the ecotype level previously, the repeated selection of *PIK3R1* across inshore lineages found here suggests this potential difference between ecotypes warrants

further investigation. Bone density changes have been also found between populations of threespine sticklebacks inhabiting environments with differing levels of freshwater input (Jamniczky et al. 2018). While there are obvious differences between these fish and bottlenose dolphins, this does provide support for the ability of bone densities to change over fine scales in response to different environmental selection pressures.

Musculoskeletal changes between inshore and offshore bottlenose dolphin ecotypes are possibly in response to differing diving behaviour and echolocation activity but may also have important roles in the adaptation of locomotion. Indeed, over-enrichment in the walking behaviour GO term was detected. Locomotion of cetaceans in their transition into the aquatic system was also heavily influenced by the development of blubber storages and adaptations of the skin (Reidenberg 2007; Wang et al. 2015). This has been important for the streamlining of the cetacean body to allow for more energy-efficient movement (Wang et al. 2015; Endo et al. 2018). Two genes with functions related to skin development – *FREM2* and *EDA* – were found to be implicated in the parallel evolution of the inshore bottlenose dolphin ecotype. Mutations in these genes can cause developmental problems with the epidermis, hair, teeth, exocrine glands (e.g. sweat and sebaceous glands), nails and craniofacial features in humans (Mikkola and Thesleff 2003; Nagaishi et al. 2012). These characteristics are all associated with major adaptations that have occurred in cetaceans and to a lesser extent in other marine mammals (e.g. sirenians and pinnipeds) (Reidenberg 2007; Gatesy et al. 2013; Lopes-Marques et al. 2019). The *EDA* gene has also been previously implicated in the parallel evolution of loss of armour in freshwater threespine sticklebacks (O'Brown et al. 2015), while *FREM2* may be involved in the evolution of pigmentation patterns in teleosts (Braasch et al. 2009). Phenotypic and genomic adaptations of the integumentary system (skin, hair, exocrine glands and nails) have not previously been documented on an ecotypic scale in marine mammals. It is possible that selection for *FREM2* and *EDA* genes are related to an interplay of several different selective pressures in the inshore habitat. This may include altered

requirements for manoeuvrability and locomotion, as well as potentially several other selective pressures affecting the integumentary system in inshore bottlenose dolphins. With both genes implicated in skin development processes in mammals and teleosts, this implies that they may play an important role in the adaptation of many different organisms to differing selective pressures in the aquatic environment.

2.7.3.7 Osmoregulation

Salinity adaptation is perhaps one of the most extensively studied processes in paired marine and freshwater populations of aquatic organisms. Previous studies have documented genes related to kidney function and osmoregulation to be under selection between freshwater and marine ecotypes of threespine sticklebacks (DeFaveri et al. 2011; Wang et al. 2014). Considerable alteration in the expression of several genes have also been reported in response to salinity changes, potentially enabling euryhaline lifestyles in marine organisms such as bull sharks (*Carcharhinus leucas*; Imaseki et al. 2019) and a number of teleost species (Yang et al. 2016; Su et al. 2019). Cetaceans inhabit almost all aquatic environments, including both hypersaline and freshwater. As such, they have had to adapt to a wide range of salinities, with genomic studies revealing positive selection on several genes related to kidney development and osmoregulation in the aquatic adaptation of dolphins (Sun et al. 2012; Gui et al. 2013; Nery et al. 2013b). Studies on the marine and freshwater ecotypes of finless porpoise (*Neophocaena* spp.) have also revealed genomic adaptation in genes associated with kidney function, urine formation and water and electrolyte balance pathways (Ruan et al. 2015; Zhou et al. 2018b), likely driving the distinct kidney structures found in the two forms (Ni and Zhou 1988; Zhou et al. 2018b). Specifically, Zhou et al. (2018b) reported differential selection on a potassium channel (*KCN*) gene between the porpoise ecotypes. This is of the same family as *KCNH5* found here to be putatively under selection in inshore SWAO bottlenose dolphins, implicating it in the early-stage ecotype evolution. This gene is also known to be involved in hypoxia tolerance in rats (Mironova et al. 2010). Over-enrichment in GO

terms related to kidney development, namely mesonephric duct morphogenesis, mesenchymal cell proliferation involved in ureteric bud development and cell proliferation involved in mesonephros development, was also found. Four genes – *AGR2, BTC, FREM2* and *GPC3* – were discovered to be associated with the potential parallel evolution of the inshore dolphins with roles in kidney development and functioning (Grisaru et al. 2001; Howarth et al. 2003; Xia et al. 2009; Saisawat et al. 2012). *FREM2* (and related gene *FREM1*) may be involved in the macroevolutionary adaptation of body fluid equilibrium pathways in dolphins (Sun et al. 2012). With a higher level of freshwater input in the inshore environment, salinity is expected to be lower and more variable than offshore. Specifically, *T. t. gephyreus* typically reside in a series of estuaries, lagoons and sheltered inshore habitats along the South American coast with strong freshwater input (Simões-Lopes and Fabian 1999; Fruet et al. 2011; Daura-Jorge et al. 2013). A similar pattern of residence is exhibited by inshore bottlenose dolphins in eastern Australia (e.g. Möller et al. 2002), but those in southern Australia are exposed to both hypersaline and low salinity environments (Bilgmann et al. 2007b; Passadore et al. 2018a; Pratt et al. 2018). This may therefore, be driving adaptation of the osmoregulatory system in these dolphins, primarily in the kidney. However, the structure of cetacean kidneys may not necessarily have evolved to cope with changes in osmotic stress, but rather in response to development of large body size and diving abilities (Nery et al. 2013b). Selection on these genes may therefore, also be associated with differences in diving behaviour and total body size between the inshore and offshore bottlenose dolphin ecotypes.

2.7.3.8 Brain Development and Nervous System

Cetaceans exhibit great variation in brain size relative to body size (Ridgway et al. 2016). Delphinids have the largest relative cerebellum and overall brain size within the cetacean lineage, also being approximately ten times larger than terrestrial cetartiodactyls of similar body size (Ridgway et al. 2016; Ridgway et al. 2018). While it was initially suggested that a large brain may have been a requirement for aquatic evolution, this has now been disproven

with both pinnipeds and sirenians displaying brain size and complexity approximately equal to their terrestrial counterparts (Marino 2007). Larger brain size requires a greater proportion of energy to be directed to the central nervous system and brain (Isler and Van Schaik 2006). Mammals meet these raised energy costs by increasing their energy intake and basal metabolic rate and/or reducing the allocation of energy put toward other processes (Mink et al. 1981; Isler and Van Schaik 2006). Subsequently, genes associated with brain and neural development and functioning, as well as lipid transport and metabolism have been found to be positively selected in the evolution of *T. truncatus* (McGowen et al. 2012). Several genes with functions related to the brain and nervous system were discovered to be possibly implicated in the evolution of the inshore ecotype. Specifically, significant differentiation in *KCNH5* and *ZNF345* genes was discovered between the inshore and offshore SWAO dolphins. Candidate genes *APH1B, CACNA1B, DYRK1A, EVL, MSI2, NKX2-2, NRXN3, NSG1, PARD3, PDE9A, PLXNA2, RORA* and *SHROOM4* were then found to be potentially involved in the parallel evolution of the inshore ecotype (see Briscoe et al. 1999; Ju and Wray 2002; Matthews and Sunde 2002; Sakakibara et al. 2002; Coolen et al. 2005; Mah et al. 2006; Heard-Costa et al. 2009; Wegiel et al. 2010; Yamaguchi et al. 2010; Sarachana and Hu 2013; Kroker et al. 2014; Armanet et al. 2015; Barford et al. 2017; Chen et al. 2017 for specific gene functions). *SHROOM4* and several *KCN, ZNF* and *CACN* genes have previously been documented to be involved in the evolution of marine mammals to the aquatic environment, particularly in regard to adaptation of the central nervous system (McGowen et al. 2012; Foote et al. 2015; Zhou et al. 2015; Zhou et al. 2018a; Huelsmann et al. 2019). The candidate gene *RORA*, encoding the retinoic acid-related orphan receptor-alpha protein, is also implicated in lipid and cholesterol metabolism (Sun et al. 2015). As larger brain size requires more energy, *RORA*, and several aforementioned candidate genes involved in energy production, may be important in this adaptation. This includes the gene *PDE1C*, of the same family as *PDE9A,* implicated here in memory deterioration (Kroker et al. 2014). Wang et al. (2015) also discovered that *PDE3B* was linked to adaptation of triacylglycerol metabolism regulatory pathways in cetaceans. Along with potentially more efficient energy production in cetaceans,

a diet much higher in calorie content compared to that of their terrestrial relatives, may have allowed the larger cetacean brain to form (Ridgway et al. 2018). Alteration to energy production pathways and a difference in diet between the bottlenose dolphin ecotypes may therefore, be closely linked to the adaptation of the brain and nervous system, as has been found for the wider cetacean lineage (McGowen et al. 2012). In birds, a larger brain has been reported to be important in colonising new habitats by enabling enhanced innovation and adaptability (Sol et al. 2005). As inshore habitats across the Southern Hemisphere show increased complexity and a wide diversity of environmental and oceanographic features in stark contrast to those encountered in the offshore realm, a large brain may be an important adaptation of these dolphins in ensuring successful colonisation of the new niche space. Furthermore, inshore bottlenose dolphins are known to have more complex behavioural and social systems than seen in the offshore ecotype (Möller 2012), even exhibiting population-specific prey handling techniques and tool use (e.g. Krützen et al. 2005; Finn et al. 2009). Prey handling, learning, tool use and behavioural and social complexity have all been previously implicated in the evolution of large brain size in mammals and birds (Marino 2005). They may therefore, be playing an important role in driving adaptation of the nervous system and brain in inshore bottlenose dolphins. With adaptation of this central bodily system recently implicated in the evolution of birds and both terrestrial and marine mammals, it is likely that this is a crucial step in the successful colonisation of new habitats.

2.7.4 Conservation Implications

The Delphinidae family is perhaps the most complicated phylogeny in the cetacean lineage. The genus *Tursiops* particularly, has a very controversial taxonomic history, with up to 20 species previously described but only two formally recognised species currently (Hershkovitz 1966; Committee on Taxonomy of the Society for Marine Mammalogy 2019). Genomic divergence within these lineages was found, corresponding to previous findings of negligible reproductive exchange between SABD and Indo-Pacific bottlenose dolphins (*T. aduncus*) in

Australian waters (Möller et al. 2008; Charlton-Robb et al. 2011), and between inshore (*T. t. gephyreus*) and offshore (*T. t. truncatus*) dolphins in SWAO (Wickert et al. 2016). A similar, if not higher, level of genomic divergence was found between SABD and *T. aduncus* as to that found between *T. t. truncatus* and *T. t. gephyreus*. It is therefore, proposed that a subspecies level classification for SABD within *T. aduncus* is appropriate (as also suggested by Moura et al. 2020). Inshore bottlenose dolphins typically reside in small, largely philopatric populations close to areas of high human disturbance (e.g. Daura-Jorge et al. 2011; Steckenreuter et al. 2012; Bossley et al. 2017). It is therefore, particularly important to define these taxonomic relationships to ensure that management strategies are well-informed about the vulnerability and adaptive capacity of these inshore dolphin lineages. In the event of major population declines, knowledge of species ranges and their ability to replenish endangered populations is especially crucial. This is particularly exemplified by our finding of extremely low genomic diversity in the potentially reproductively isolated inshore SWAO dolphins, suggesting that this population is especially vulnerable to population declines. With ongoing environmental changes throughout the world's oceans, it is important to understand how marine organisms may respond and how this could shape patterns of speciation. Several key pathways for the adaptation of bottlenose dolphins to contrasting selective pressures between the inshore and offshore ecosystems were highlighted. This includes adaptation of the nervous, gastrointestinal, cardiovascular, circulatory, sensory, musculoskeletal, integumentary and renal systems, as well as alterations to adipogenesis, osmoregulation, energy production, thermoregulation and brain and mouth development. Although not discussed in detail here, further research into ecotypic differences in genes associated with immune responses and reproduction is also warranted. These processes appear to be important to both the early stages of inshore colonisation, as well as the long-term success in this habitat and eventual genomic divergence along the speciation continuum. These results suggest that bottlenose dolphins have a vast capacity for adapting to changing selective pressures, but this is likely over an evolutionary scale of thousands of years. Anthropogenically accelerated climate change may therefore, pose a significant test for the adaptive capacity of these dolphins and

other long-lived marine vertebrates. The findings presented here are an important step in understanding the vast scope of potential adaptive responses by marine organisms.

2.7.5 Limitations and Future Directions

This study presents the first evidence of likely parallel evolution of genes associated with several major physiological systems in inshore bottlenose dolphins in the Southern Hemisphere. Despite relatively high power to detect signatures of selection (Manel et al. 2016), the reduced-representation nature of ddRADseq does yield relatively low genomic coverage (see Davey et al. 2011). As a result, several important genes involved in the adaptive divergence of bottlenose dolphin inshore and offshore ecotypes are likely to have been missed here (e.g. candidate genes associated with adaptations of the respiratory system). The use of whole-genome sequencing in future studies would allow a more comprehensive overview of ecotype formation in bottlenose dolphins and provide support to the results of this study. Additional research should also investigate the history of ecotype divergence in the Southern Hemisphere using coalescent-based demographic reconstructions (see Hein et al. 2004). This framework would allow candidate genes to be tested for the independent effects of adaptive divergence compared to whole-genome drift processes (Latta 2004). To further compliment this, future studies should endeavour to include representatives of the South African inshore bottlenose dolphins (*T. aduncus*) and inshore and offshore populations from across the Northern Hemisphere (*T. truncatus* (including *T. t. ponticus*) and *T. aduncus*). Enhanced collaboration between scientists across these study regions would allow the *Tursiops* phylogeny and patterns, as well as the underlying causes of genomic divergence, to be elucidated more completely.

2.8 Conclusion

This study addressed the genomic basis of divergence and ecotype formation in bottlenose dolphins (genus *Tursiops*) in the Southern Hemisphere. Genomic divergence and subsequent negligible reproductive connection was revealed between *T. t. truncatus* and *T. t. gephyreus* and between *T. aduncus* and SABD. Subspecies level classification is thus proposed for SABD within *T. aduncus* as currently recognised for *T. t. gephyreus* within *T. truncatus*. The inshore ecotype generally had reduced genomic diversity but was particularly low in *T. t. gephyreus* in SWAO, making them particularly vulnerable to anthropogenic disturbances. Differential adaptation between the inshore and offshore ecotypes was evidenced by putative selection on several genes and over-enrichment in GO terms associated with most major bodily systems. This suggests that differing selective pressures between the two habitats are likely creating contrasting adaptations in the dolphins, potentially driving genomic divergence of inshore and offshore ecotypes, and leading to species and subspecies level differentiation. On the other hand, similar environmental characteristics in inshore habitats across the Southern Hemisphere, appears to be driving parallel evolution in the inshore bottlenose dolphin ecotype. With climate change continuing to alter marine ecosystems, this research informs about the potential pathways for cetaceans and other marine vertebrates to adapt to a changing environment and how this may drive population and higher-level divisions. This is an important first step in investigating genomic adaptation in marine organisms, which warrants further investigation at a whole-genome level.

Chapter 3 : Seascape Genomics of Indo-Pacific Bottlenose Dolphins (*Tursiops aduncus*) in Eastern Australia

3.1 Contributions

Eleanor Pratt – conception of study and design of methods, DNA extraction and ddRAD library preparation, bioinformatics and analysis, writing of thesis.

Luciana Möller – primary supervisor – conception of study design, guidance in interpretation, collection of samples, drafting and revision of thesis.

Luciano Beheregaray – associate supervisor – guidance in design and interpretation, drafting and revision of thesis.

Jonathan Sandoval-Castillo – guidance for laboratory methods and bioinformatics and assistance with analysis.

Chris Brauer – guidance and assistance with analysis.

3.2 Permits and Ethics Approvals

Biopsy samples were obtained under licences from the Department of Environment and Climate Change (Licence Number: S10763) and Marine Parks Authority (Permit Number: PSGLMP 2008 ⁄ 003), and under approval by the Macquarie University Animal Ethics Committee (AEC Reference Number: 2007⁄013) as per Wiszniewski et al. (2010 and 2012).

3.3 Abstract

Microevolution in many coastal marine organisms is thought to be driven by adaptation to local environmental and oceanographic conditions, making them particularly vulnerable to changes in the oceans. This hypothesis can now be empirically tested due to the recent emergence of cost-effective genomic techniques and spatial statistical methods that enable the study of adaptation in many non-model species. Here, a seascape genomics framework is applied to investigate population genomic structure and genome-environment associations (GEAs) in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*; IPBD) along the east coast of Australia. Utilising over 14,000 SNPs, the presence of fine-scale population genomic structure was established, as well as contrasting patterns of differentiation between putatively adaptive and neutral datasets. Genomic differentiation of IPBD appears to be driven by intertwined factors, including spatial scales, neutral processes (e.g. genetic drift), local ocean circulation, broad-scale habitat heterogeneity and productivity gradients. Habitat and possibly prey specialisation, reinforced by strong social structure, is likely impacting on neutral population genomic structure of these animals in eastern Australia. At the adaptive level, these factors also appear to be promoting physiological adaptations and differentiation. Functional annotation revealed candidate genes potentially important in the adaptation of IPBD to seascape heterogeneity and associated prey assemblage changes across the region. This included genes involved in bone and heart development, and the elongation of fatty acid chains into molecules needed for healthy bodily functioning. Given the status of the EAC as a global hotspot for climate change, selective pressures on local biodiversity are changing rapidly and are expected to increase in the future. This study pioneers the assessment of population genomic differentiation and putative adaptation to seascape heterogeneity in a near-top predator along the EAC and contributes to our understanding of the adaptive capacity of coastal dolphins under scenarios of rapid environmental change.

3.4 Introduction

Heterogeneous environmental conditions are typical across both spatial and temporal scales and are significant evolutionary drivers of populations (Nevo 2001; Schluter 2001). Environmental differences among habitats creates opposing selective forces impacting on local biodiversity. Consequently, natural selection tends to favour traits that are best suited to the local conditions (Williams 1996). Where species span a heterogeneous landscape or seascape, this can lead to genomic differentiation and subsequent reproductive isolation as local populations become more specialised to their particular habitats (Schluter 2001). This level of population adaptive divergence is typically termed microevolution (Jablonski 2000). Microevolution encompasses the creation of new species and divergent populations as a taxon colonises and adapts to new habitats and environmental conditions. The emergence of landscape genetics, and more recently landscape genomics, allows scientists to study microevolution in more detail (Joost et al. 2007). This framework utilises GEAs to study genome-wide neutral and adaptive variation in species spanning heterogeneous landscapes (Grummer et al. 2019). It allows the identification of environmental factors acting as selective agents on different populations, as well as regions of the genome that are potentially affected by these selective forces (i.e. under adaptive selection). While the application of this framework in terrestrial ecosystems is well established (e.g. Holderegger and Wagner 2006; Rellstab et al. 2015), the same is not true for marine ecosystems (Grummer et al. 2019). Referred to as seascape genomics, this approach is receiving growing attention, particularly for species of commercial interest or conservation concern (e.g. Benestan et al. 2016; Rodriguez-Zarate et al. 2018; Sandoval-Castillo et al. 2018). By studying environmental factors that are potentially driving population differentiation and adaptation, scientists can also gain a better understanding of how these species will respond to future changes to their habitats (Grummer et al. 2019). This is particularly important in the face of ongoing human-induced climate change, with the world's ecosystems already undergoing major alterations. The oceans are predicted to become warmer and more acidic, with simultaneous changes to

ocean circulation, and the strength and position of upwellings (Gregory et al. 2005; Poloczanska et al. 2013). Environmental factors such as SST, salinity and primary productivity have been found to be associated with population differentiation and adaptation in marine organisms, including invertebrates, teleosts and marine mammals (Amaral et al. 2017; Diopere et al. 2017; do Prado et al. 2018; Bernatchez et al. 2019; Teske et al. 2019). Assessing the adaptive capacity of marine populations is therefore, needed to make informed management decisions under climate change scenarios.

Cetaceans provide an opportunity to evaluate the effects of selective pressures imposed by the environment at both macro- and microevolutionary scales. Approximately 53 MYA cetaceans began their macroevolutionary transition from a terrestrial to a fully aquatic lifestyle (Thewissen et al. 2009; McGowen et al. 2014). They have subsequently radiated throughout the world's oceans and now inhabit most aquatic habitats, from tropical to polar waters and from open-ocean to freshwater environments (Jefferson et al. 2011). With the adaptation of cetaceans to a vast range of environmental conditions and niche spaces, major differences in morphology, physiology, prey choice and hunting techniques have evolved among species (Jefferson et al. 2011). Differences in these characteristics can also be found within species, typically related to local adaptation of populations to their particular niche space (e.g. Westgate 2007; Foote et al. 2009; Fruet et al. 2017). Local adaptation and associated levels of genetic differentiation between cetacean populations are often driven and reinforced by complex social behaviours typical of many species (Möller 2012). This can include the strong matrilineal bonds and cultural transmission of prey handling techniques found in genetically divergent populations of killer whales (*Orcinus orca*; Moura et al. 2015; Foote et al. 2016), as well as sex-biased dispersal observed in a number of cetacean species, including bottlenose dolphins (*Tursiops* sp.; Möller and Beheregaray 2004; Bilgmann et al. 2007b) and spotted dolphins (*Stenella* sp.; Adams and Rosel 2006). Oceanographic and environmental characteristics such as ocean currents, primary productivity, salinity and SST have also been suggested to

influence population genetic structuring of many cetacean species, including common dolphins *(Delphinus delphis*; Möller et al. 2011; Amaral et al. 2012a), franciscanas (*Pontoporia blainvillei*; Mendez et al. 2010), humpback dolphins (*Sousa* spp.; Mendez et al. 2011; Amaral et al. 2017) and bottlenose dolphins (*Tursiops* sp.; Natoli et al. 2005; Wiszniewski et al. 2010; Amaral et al. 2017; Pratt et al. 2018).

While the influence of certain environmental characteristics on local adaptation and population genetic structure has long been suggested for cetaceans, formally testing the relationship between these factors has been largely out of reach. With the rise of seascape genomics (Grummer et al. 2019), scientists can now empirically test the strength of association between particular environmental variables and population differentiation in non-model marine organisms, such as cetaceans. This is of particular use for Indo-Pacific bottlenose dolphins *(Tursiops aduncus*; hereafter referred to as IPBD) due to their preference for coastal waters and embayments that are environmentally heterogeneous. The distribution of *T. aduncus* along the east coast of Australia spans a region that is characterised by strong anthropogenic impacts to the marine environment. This area is home to the largest cetacean-based tourism industry in Australia (Steckenreuter et al. 2012) and has been deemed a “global hotspot” for ocean temperature change, with warming in this area three to four times that of the global average (reviewed in Suthers et al. 2011). The local marine environment is heavily influenced by the EAC, a western boundary current that moves southward along eastern Australia then deflects toward New Zealand (Suthers et al. 2011; Macdonald et al. 2013). Worldwide, many other boundary currents are experiencing similar rapid warming (Wu et al. 2012). Boundary currents, including the EAC, often create latitudinal gradients in environmental variables such as SST, productivity and salinity, which can affect the spatial connectivity of local biota (e.g. Carr and Kearns 2003; Suthers et al. 2011; Baltazar-Soares et al. 2014; Everett et al. 2014; Oke et al. 2019). In eastern Australia these impacts can be seen in another small cetacean, the common dolphin (*D. delphis*; Möller et al. 2011), as well as in coastal fishes (Shaddick et

al. 2011) and several species of invertebrates (Hoskin 2000; Banks et al. 2007; Piggott et al. 2008). Previous research of IPBD in the region based on microsatellite DNA markers demonstrated fine-scale population genetic structure (Wiszniewski et al. 2010) similar to that in common dolphins (Möller et al. 2011). Complex population genetic structure over small spatial scales is typical of coastal bottlenose dolphins worldwide (e.g. Sellas et al. 2005; Ansmann et al. 2012; Fruet et al. 2014; Pratt et al. 2018). This appears associated with local adaptation to their specific habitats (Möller et al. 2007; Wiszniewski et al. 2010; Amaral et al. 2017), but such a hypothesis has not been properly tested given that previous studies were based on a few DNA markers incapable of detecting signals of selection. With the development of GEA analyses this can now be empirically tested by identifying regions of the genome that are potentially under selection and therefore, indicative of adaptation to environmental or oceanographic heterogeneity. Scientists can now distinguish between patterns of differentiation driven by neutral compared to putatively adaptive loci. That enables a better understanding of the relative roles of selection versus neutral processes (e.g. genetic drift, gene flow and mutation) in shaping genetic variation. Neutral loci are DNA regions that do not directly affect the fitness of an individual or population. These loci are therefore, under the influence of the aforementioned neutral processes. Adaptive loci on the other hand, do putatively have an effect on fitness and thus alleles at these loci may be actively selected for or against by natural selection (Holderegger et al. 2006). It is important to differentiate between these two processes in order to understand the drivers of population genomic differentiation and connectivity of IPBD over the seascape. This allows informed management decisions and predictions of the adaptive capacity of these dolphins in the face of climate change to be made.

This chapter generates and uses genome-wide data to investigate the relationship between population genomic differentiation of IPBD (*T. aduncus*) and environmental gradients along the New South Wales (NSW) coast in eastern Australia using a seascape genomics approach. It is hypothesised that neutral population structure will follow a pattern of isolation by distance

(IBD) demonstrating the influence of the EAC and the lack of obvious geographical barriers to dispersal of cetaceans along eastern Australia. Adaptive population structure on the other hand, is suggested to occur in association with environmental and oceanographic variation and may reflect adaptation of dolphins to local conditions. DdRADseq was utilised to develop a genomic dataset of 14,466 SNPs. While the population genetic structure of the IPBD in this region has previously been studied with microsatellite and mitochondrial DNA markers (Möller et al. 2007; Wiszniewski et al. 2010), the large genome-wide dataset utilised here substantially improves the power to detect neutral patterns of genomic differentiation (Luikart et al. 2003). This is particularly important in providing better resolution to the fine-scale dolphin substructure detected in Hunter Region (Möller et al. 2007; Wiszniewski et al. 2010). The genomic basis of adaptive divergence is also investigated through GEA analyses, empirically testing for the association between genomic differentiation and 26 environmental, topological and oceanographic variables. Informed by the GEA analyses, putatively adaptive and neutral loci were separated for population genomic structure and diversity analyses. With the EAC and other similar boundary currents predicted to warm and strengthen significantly in the future (Suthers et al. 2011; Van Gennip et al. 2017), the population structure of *T. aduncus* and other marine species are likely to change over time. This study pioneers the application of seascape genomics to elucidate evolutionary adaptation and resilience of bottlenose dolphins across a rapidly warming and vulnerable coastal marine ecosystem.

3.5 Methods

3.5.1 Sample Collection

Free-ranging IPBD were biopsy-sampled at nine locations along the east coast of Australia between 1998 and 2007 (Figure 3.1). Skin and blubber samples were collected using either a hand-held biopsy pole (Bilgmann et al. 2007a) or a remote biopsy gun system (Krützen et al. 2002). Visible biopsy wound marks on the animal's body and identification of recognisable dorsal fin characteristics were used to minimise the risk of resampling individuals. Dependent calves were not sampled. Biopsy samples were preserved in either 100% ethanol or a salt-saturated solution of 20% DMSO and later stored at -80ºC. Photo-identification studies of Port Stephens (PS) and Jervis Bay (JB) bottlenose dolphins confirm the high level of residency in these communities (Möller et al. 2002; Wiszniewski et al. 2009). While similar studies have not been completed for all sampled locations in eastern Australia, natal philopatry and small home ranges are typical of inshore bottlenose dolphins worldwide (see Möller 2012). Microsatellite data generated by Wiszniewski et al. (2010) were used to select 111 samples for genomic analysis based on pairwise relatedness estimates <0.5 (theoretical value for first-order relatives) calculated in *GenAlEx* (Peakall and Smouse 2006, 2012) using the Queller and Goodnight (1989) estimator.

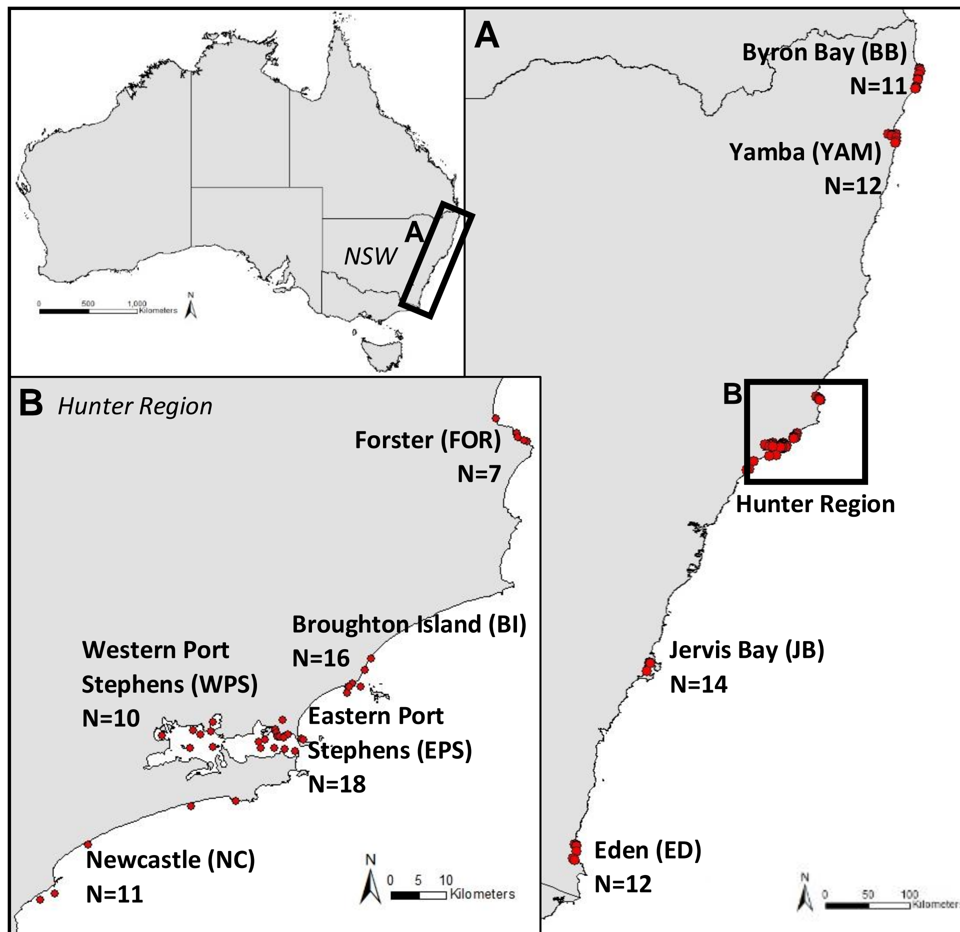


Figure 3.1 Indo-Pacific bottlenose dolphin (*T. aduncus*) sampling sites along the New South Wales (NSW) coast in eastern Australia. N = sample size before filtering.

3.5.2 Genomic Laboratory Methods

3.5.2.1 DNA Extraction

DNA was extracted from biopsy samples using the salting-out protocol as detailed in Chapter 2 (Sunnucks and Hales 1996), and checked for quality and quantity using gel electrophoresis, a NanoDrop 1000 spectrophotometer (Thermo Scientific) and a Qubit 2.0 fluorometer (Life Technologies). Sexes of individual dolphins were known from the work of Wiszniewski et al. (2010).

3.5.2.2 ddRAD Library Preparation

Libraries were prepared following a ddRAD protocol modified from Peterson et al. (2012) as per conditions in Brauer et al. (2016) (detailed in Chapter 2). Libraries for this chapter were multiplexed with 96 individually barcoded samples per lane, which were sequenced at the SAHMRI as per Chapter 2. Genomic data for JB and many PS samples were already available from Chapter 2.

3.5.3 Bioinformatics

Demultiplexing and processing of raw data files was completed using *dDocent* v.2.2.19 (Puritz et al. 2014) as described in Chapter 2. See Appendix B.iii.1 and 2 for all *dDocent* scripts used throughout this chapter. *VCFtools* was used to filter the VCF created by the *dDocent* pipeline using custom BASH scripts (Appendix C.iii.1) for the filtering steps outlined in Appendix C: Table C.ii.1 (modified from Brauer et al. 2016). The resulting 14,466 high-quality SNPs were then aligned to the *T. aduncus* genome downloaded from NCBI (GCA_003227395.1 ASM322739v1), to confirm that the final dataset of SNPs for downstream analysis did not contain exogenous DNA.

3.5.4 Neutral Population Genomic Diversity and Structure

Outlier/candidate loci identified by GEA analysis using redundancy analysis (RDA) or any of the additional outlier detection methods (see below) were considered as potentially adaptive and were thus removed from the full dataset to create the putatively neutral dataset. *Arlequin* v.3.5.2.2 (Excoffier and Lischer 2010) was used to calculate molecular diversity indices for the neutral dataset, including %PL, H_E and H_O for dolphins at each sampling location. Wright's inbreeding coefficient (F_{IS}) at the population level was calculated as (H_E - H_O) / H_E (Wright 1922). To assess population structure at the neutral level, *Arlequin* was used to estimate pairwise genomic differentiation (F_{ST}) and corresponding significance levels among sampling

locations based on 10,000 permutations. To account for biases potentially created by multiple testing, the significance levels were corrected using the Benjamin and Yekutieli's (2001) method (i.e. B-Y correction) (see Narum 2006). This resulted in an alpha (α) level of 0.0120. Both non-model and model-based methods were used to further test for neutral population genomic structure. Non-model methods included PCA run using the packages *adegenet* and *FactoMineR* in R and a DAPC using *adegenet* (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011). The number of clusters with the lowest BIC value estimated by DAPC was considered the most statistically supported number of groups (i.e. populations) in the dataset. The model-based method *Admixture* v.1.3.0 (Alexander et al. 2009) uses a maximum-likelihood approach to estimate the most likely K in the dataset. This was run with default settings and testing for K values between one and 10 (number of sampling locations plus one). The K value with the lowest cross validation (CV) error was determined to be the most statistically supported number of neutral populations. *Arlequin* was used to carry out an AMOVA testing the level of genomic variance explained by these four populations. IBD was tested for using a Mantel test implemented in the R package *vegan* (Oksanen 2011). Pairwise geographical distances among sampling sites were measured in Google Maps with the closest along-shore distance taken, while linearised F_{ST} ($F_{ST}/(1- F_{ST})$) was used as the measure of population genomic distance. Along-shore distance was used as this is the most likely path of travel for coastal bottlenose dolphins moving among regions (L. Möller pers. comm.). Contemporary migration (in the last two generations) among the four putative populations was investigated using *BayesAss3* (Wilson and Rannala 2003). This is a Bayesian multilocus method that uses a Markov chain Monte Carlo (MCMC) technique modified for application with genomic markers. To reduce computational time 5,000 SNPs were randomly selected from the neutral dataset for use in the analysis. Mixing parameters for allele frequencies, inbreeding coefficient and migration rates were adjusted until acceptance rates were between 20 and 60% (0.32, 0.33, 0.32 for each parameter, respectively) as are suggested to be optimal (Rannala 2007). To assess chain convergence three runs were conducted, each with the same parameters of one hundred million iterations and ten million burn-in steps, but with 500,

1,000 and 5,000 seeds for each run, respectively. These were then checked for congruence and convergence in *Tracer* v.1.7.1 (Rambaut et al. 2018). There were no substantial differences between any of the three runs and as such the results of only the run with 1,000 seeds are shown here.

3.5.5 Genotype-Environment Association

Loci potentially under selection due to association with local environmental and oceanographic variation were identified using a GEA approach. Eleven ecological variables, including a number of topographic features, were selected due to hypothesised associations between these variables and bottlenose dolphin population structure in Australian waters (Bilgmann et al. 2007b; Wiszniewski et al. 2010; Zanardo et al. 2018) and elsewhere (Natoli et al. 2005; Amaral et al. 2017). The variables were: salinity, SST, primary productivity, chlorophyll A concentration (chloA), current velocity, topographic relief, seafloor slope, seafloor rugosity, bathymetry and distance to closest mouth of a major river (Appendix C: Table C.ii.2). The average annual mean, maximum, minimum and range of each variable were used whenever possible, resulting in a total of 26 variables obtained from a variety of sources (Appendix C: Table C.ii.2). ChloA data were not available for the westernmost reaches of the PS embayment and was therefore, extrapolated from the geographically nearest available data point (i.e. from east PS) (Appendix C: Figure C.i.1Cii). The distance to river mouth data were calculated using ArcGIS to measure the distance between each individual sample and the mouth of the closest major river, defined as those with a relevance category of one or two in the publicly available data layer from the NSW Government Spatial Services (more information given in Appendix C: Table C.ii.2). Standardisation of environmental data was done in Microsoft Excel using the basic calculation of =standardize(x, mean, standard deviation).

To establish which ecological variable/s are likely to be driving population genomic differentiation across the study region, a multivariate RDA approach was implemented in the R package *vegan* (Oksanen 2011). This approach was chosen as it has higher power compared to other popular GEA techniques and can deal well with IBD, variation in sample sizes, selection strengths and demographic histories, providing consistently low false-positive and high true-positive rates (Forester et al. 2018; reviewed in Grummer et al. 2019). The RDA was done at the individual level to avoid potential biases related to dolphin movement, which could be created by taking an average value for each sampling locality. In RDA however, pairwise distances among samples are calculated as straight lines and therefore, using individual XY coordinates can introduce biases where there are complex coastlines and large embayments (e.g. PS). To control for this, pairwise seascape distances among individuals (i.e. no land crossing) were first determined using ArcGIS before transforming the distance matrix to Moran's eigenvector maps (MEM) using the R package *memgene* (Galpern et al. 2014). The MEM axes were then used as spatial variables in the RDA to represent spatial genomic patterns across multiple geographic scales (see Galpern et al. 2014). While IPBD in NSW most likely travel primarily along the coastline (L. Möller pers. comm.), the transformed seascape distances used here are expected to be an adequate representation of spatial differences among IPBD. Ecological variables explaining a significant (p <0.05) portion of variation in the genomic data were identified using a forward selection approach. Collinearity among the variables was minimised by removing highly correlated parameters one-by-one until all those remaining had a conservative variance inflation factor (VIF) of <3, as previously used in other cetacean studies (Christiansen et al. 2013; Goldbogen et al. 2015). Only the variables retained after these two processes were included in the final RDA model. Heterogeneity in these variables across the study region was then visually inspected in ArcGIS. Significance of the overall model and of each individual ecological variable were calculated through 1,000 analysis of variance (ANOVA) permutations. Loci were identified as being potentially under selection if they had a score greater than three standard deviations (SD) from the mean locus scores, which were calculated across all loci for each of the first

three RDA axes (Forester et al. 2016). The correlation between allele frequencies of candidate SNPs and each of the retained variables was calculated to find which ecological variable/s each candidate SNP was most highly associated with.

3.5.6 Adaptive Population Genetic Diversity and Structure

Loci identified as potentially under selection by the RDA were extracted from the full genomic dataset to form the putatively adaptive dataset (hereafter referred to as the adaptive dataset). Adaptive molecular diversity and population genetic structure were then assessed using the same methods as detailed for the neutral population genomic diversity and structure section. Briefly, this included using *Arlequin* to calculate genetic diversity measures (H_E, H_O and %PL). Wright's inbreeding coefficient (F_{IS}) was not estimated for the adaptive dataset due to violation of the assumption of neutrality. *Arlequin* was also used to estimate pairwise genetic differentiation (F_{ST}) and corresponding B-Y corrected significance levels among sampling locations. Adaptive population structure was further investigated by re-running PCA, DAPC and *Admixture* with putatively adaptive loci only. *Admixture* was again run with default settings, testing for values of *K* between one and 10. The *K* value with the lowest CV error was identified as the most statistically supported number of populations at the adaptive level. Due to the putatively adaptive nature of this dataset the assumption made by *Admixture* of the loci being in Hardy-Weinberg Equilibrium (HWE) is expected to be violated (Funk et al. 2012). This analysis was still included to allow for a comparison to the disclosed neutral population genomic structure. A Mantel test in the R package *vegan* was used to test for IBD. Geographical distances were calculated in Google Maps using the closest along-shore distance, being the most likely path of travel for coastal IPBD as explained above. Adaptive genetic differentiation (linearised F_{ST}) was used to represent genomic distances among sampling localities.

3.5.7 Functional Enrichment Analysis and Annotation

A functional enrichment analysis using the Fisher's exact test was used to identify GO annotation terms significantly over- and/or under-represented in the candidate dataset. First, 300 bp flanking sequences either side of each SNP (601 bp in length) were extracted from the *T. aduncus* genome (as used in the bioinformatics steps) for all 14,466 SNPs in the full dataset. The BLAST was then run using *blastn* and the NCBI nucleotide database (Altschul et al. 1990; Sayers et al. 2019). Blast2GO (Conesa et al. 2005) was then used to map and annotate all "blasted" loci. Expectation (e) values of 1E-6 and 1E-3 were used for the BLAST and annotation steps, respectively. A Fisher's exact test was then conducted in Blast2GO with an alpha value of 0.05. To further investigate adaptations that may be occurring in conjunction with environmental gradients, the putative functions of loci that were moderately to highly correlated (<-0.4 or >0.4) with either (or both) of the significant environmental variables in the RDA (primary productivity range and mean chloA) were investigated. These candidate loci were run in the NCBI web BLAST search against the *T. truncatus* genome assembly (NIST Tur_tru v1 Reference Annotation Release 101) using *blastn*. A threshold e-value of <1E-10 and an identity of >90% were used to select the most reliable candidates. Candidate genes were selected if they were within 20 kilobases (KB) of the query sequence (as previously used for southern Australian bottlenose dolphins; Batley et al. 2019). Putative gene functions and associated GO terms were then found using the Swiss-Prot database in UniProtKB (Boutet et al. 2007; The UniProt Consortium 2018).

3.5.8 Additional Outlier Tests for Selection

To further limit the presence of SNPs responding to selection in the neutral dataset, three additional methods for detecting outlier loci were implemented to compliment the RDA. Informed by exploratory analyses with the full dataset, each method was utilised to identify putative outliers across the four inferred major populations (northern NSW, Hunter Coast, PS and southern NSW; see Results), as well as for each pairwise comparison between these

populations. First, the coalescent-based FDIST method (Beaumont and Nichols 1996) employed by *Arlequin* was used. This was run under the hierarchical island model with 100,000 simulations and 100 demes for each of 10 groups (Excoffier et al. 2009). Using the *p.adjust* function in the R package *plyr* (Wickham 2011), p-values were corrected for FDR to avoid biases due to multiple testing. A Bayesian outlier test was then run in *Bayescan* v.2.1 (Foll and Gaggiotti 2008), with 100,000 iterations using prior odds of 10 and a burn-in period of 50,000. For these two analyses, SNPs were considered as outliers if they had a FDR of <10% (i.e. q-value <0.1). Finally, the RandomForest machine-learning method was implemented using the *rfPermute* and *randomForest* v.4.6-14 packages in R (Breiman 2001; Archer 2016). The method used by RandomForest is described in Chapter 2 and was implemented here with 125,000 trees and default settings for the proximity and importance parameters. The number of randomly chosen SNPs tested for each split of the tree (mtry) was set to the value that minimised the out-of-bag error rate as well as computational time (as suggested by Brieuc et al. 2018). As there is no standard process for the selection of candidate loci in RandomForest, the protocol established by Batley et al. (2018) was used. This involved plotting the importance value distributions, with loci above the upper elbow of the distribution curve selected as candidates. Loci identified to be potentially under selection by any of these three methods or GEA analysis were removed from the full dataset to form the putatively neutral dataset as detailed above.

3.6 Results

Biopsy samples of 111 free-ranging IPBD from the east coast of Australia were ddRAD sequenced (Figure 3.1). Samples were sequenced across six lanes (in combination with samples for Chapters 2 and 4), with a total of 193,782,853 raw reads (average of 2,018,571 reads per individual (±1,316,471 standard deviation)). One individual with >20% missing data was removed from the dataset; this sample invariably had a very low number of reads compared to other samples. Missing data was on average 2.4% (±1.5% standard deviation) for the retained samples. Further filtering to obtain data of the highest quality (Appendix C: Table C.ii.1), left 14,466 SNPs for 110 individuals for downstream analysis. When mapped to the *T. aduncus* genome, 99.65% of these SNPs (n= 14,416) aligned. Although the dataset was not filtered for only those SNPs aligning to the *T. aduncus* genome, the very high percentage of alignment indicates that this will not bias results.

3.6.1 Neutral Population Genomic Diversity and Structure

A total of 323 candidate SNPs were identified by GEA analysis (see below) and the three additional outlier detection methods. This included 281 RDA, 22 *Arlequin*, 16 *Bayescan* and 30 RandomForest candidate SNPs. Twelve outliers were detected in more than one of the three additional outlier detection methods and 14 outlier SNPs were also identified as candidates by RDA. These loci were subsequently removed from the full dataset to form the neutral dataset of 14,143 SNPs. Neutral genomic diversity across all localities was much lower than observed in the adaptive dataset (Appendix C: Table C.ii.3). This appears to differ among sampling locations irrespective of inferred habitat type or population (as defined below). There were largely negligible differences in inbreeding levels between each sampling location, with low levels found across localities and particularly in PS (Appendix C: Table C.ii.3).

Fine-scale neutral population genomic structure was found in IPBD along the NSW coast in eastern Australia. *Admixture* showed clear genomic division between northern (Yamba and

Byron Bay (BB)), central (Hunter Region - PS east and west, Newcastle, Forster and Broughton Island (BI)) and southern (JB and Eden) sampling regions, with three as the most statistically supported number of populations (Figure 3.2Ai1 and 2). Further subdivision within the Hunter Region was revealed when analysing the pattern of structure at the highly supported $K = 4$ (Figure 3.2Ai3). It should be noted that Hunter Region refers to all sites in the central NSW population (PS east and west, Newcastle, Forster and BI), while the Hunter Coast refers to only the central open-coast localities (Newcastle, Forster and BI). Three neutral populations was supported by PCA and DAPC, but PCA did also show the lower level genomic division between Hunter Coast and PS dolphins (Figure 3.2Aii and iii). The presence of four hierarchically structured populations was supported by AMOVA reporting that 4.49% (p <0.001) of the variation was explained by population division, compared to only 1.24% (p <0.001) among sampling locations within populations (Appendix C: Table C.ii.4).

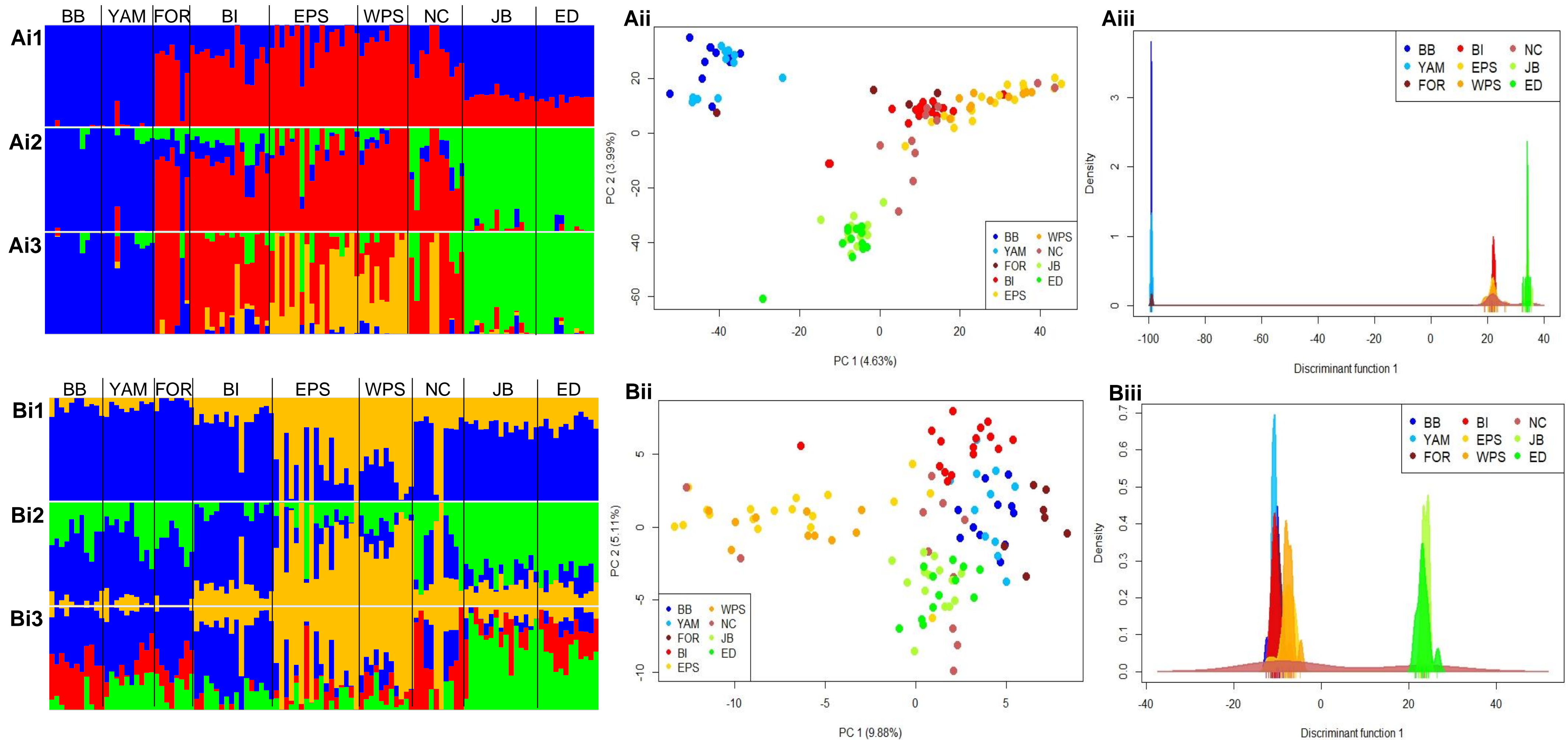


Figure 3.2 Population genomic structure of Indo-Pacific bottlenose dolphins along the east coast of Australia based on 110 individuals and **A)** 14,143 SNPs in the neutral dataset and **B)** 281 outlier SNPs in the adaptive dataset. **i)** *Admixture* plots whereby each column represents an individual dolphin and the proportion of a given colour represents the percentage probability of that dolphin belonging to a given population. Vertical black lines mark the start of a new sampling location. Population structure is shown at **1)** $K = 2$ populations (neutral = highly supported; adaptive = low support); **2)** $K = 3$ (neutral = most supported; adaptive = highly supported); and **3)** $K = 4$ (neutral = highly supported; adaptive = most supported). **ii)** Principal component analysis (PCA); and **iii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from north (left/top) to south (right/bottom) and colours correspond to neutral population membership. Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3).

Genomic differentiation (F_{ST}) across the study region was moderate, with an average of 0.051 (Figure 3.3). All pairwise comparisons were significant, except those between the eastern and western PS dolphins and between BI and Forster. Significant F_{ST} values between localities within each of the four main populations suggested further fine-scale subdivision. The highest level of differentiation was recorded between the northernmost and southernmost sample sites, BB and Eden at 0.091. A Mantel test subsequently found IBD to be affecting neutral population genomic structure ($r = 0.8795$, $p < 0.001$, Figure 3.4). Contemporary migration was revealed to be most significant from the Hunter Coast into the PS population (Appendix C: Figure C.i.2). Migration north from the southern NSW population into the Hunter Coast and PS was also more substantial than between northern NSW and the other populations. While each population had a high percentage of non-migrants (mean = 90.13%), the PS dolphins had the lowest proportion at 81.59%, consistent with the high level of immigration from the Hunter Coast (Appendix C: Figure C.i.2).

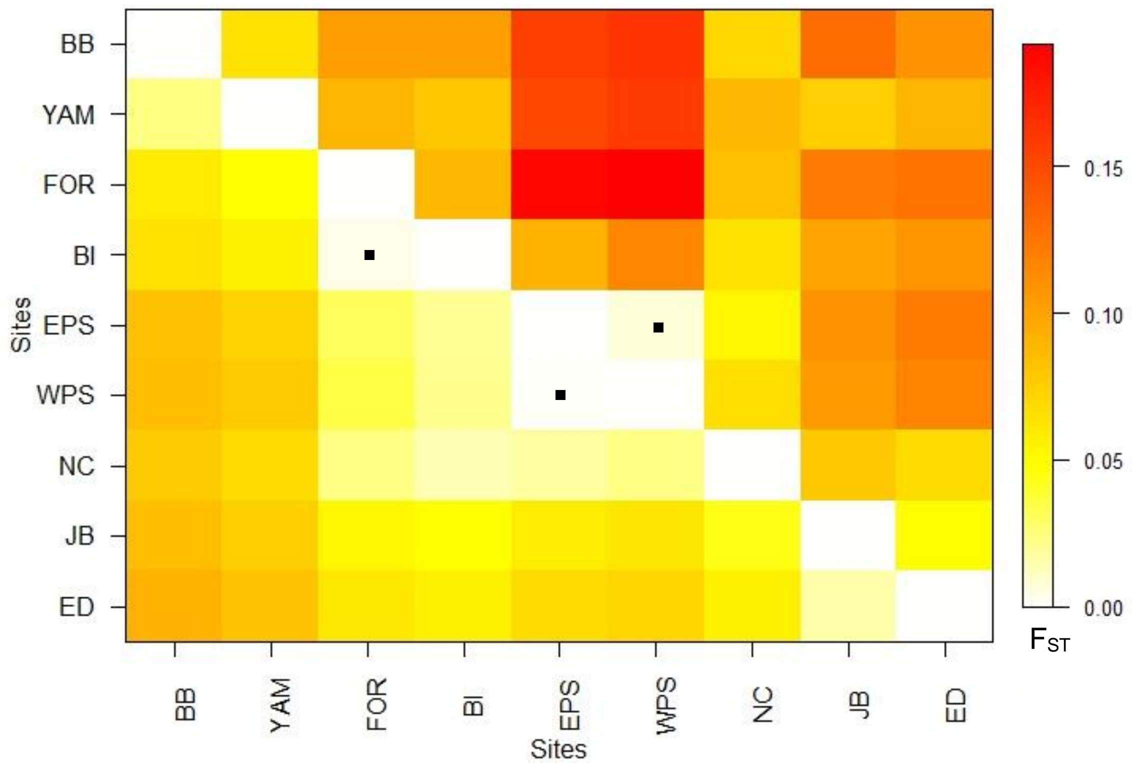


Figure 3.3 Heat map of pairwise genomic differentiation (F_{ST}) between sampling sites of Indo-Pacific bottlenose dolphins along the east coast of Australia (n=110 individuals) as estimated by *Arlequin*. Estimates based on the adaptive dataset (281 SNPs) are found in the top half of the matrix, while estimates based on the neutral dataset (14,143 SNPs) are in the bottom half. Comparisons non-significant at the B-Y corrected alpha value 0.0120 are marked with a black square (▪). Average F_{ST} for the neutral dataset was 0.051 and for the adaptive dataset was 0.099. Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3).

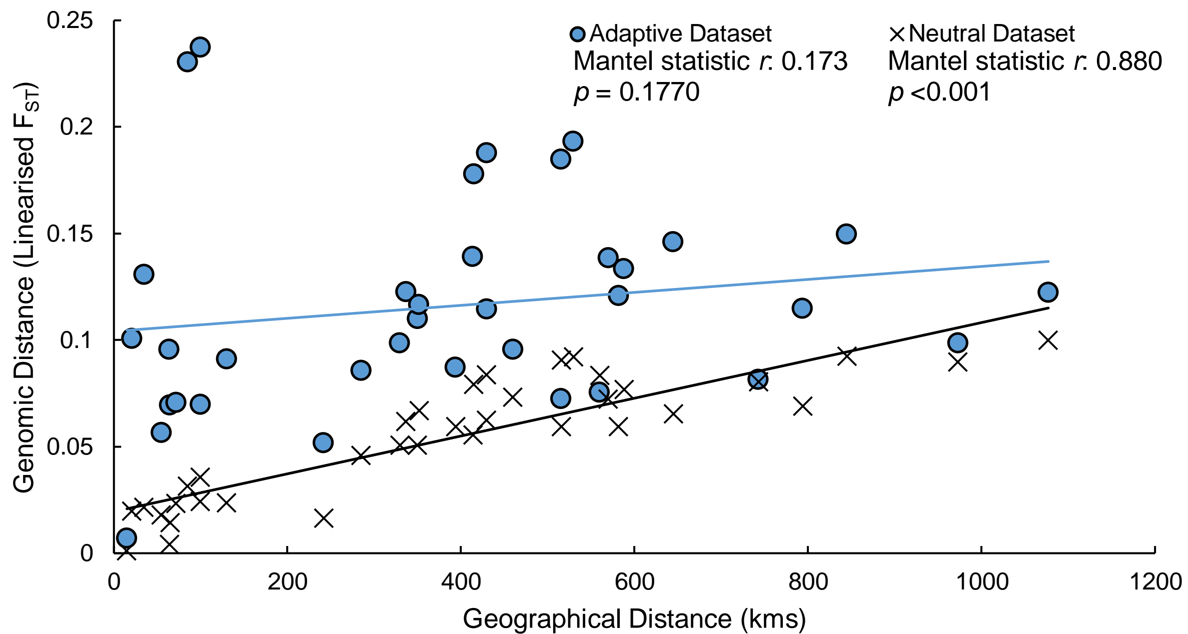
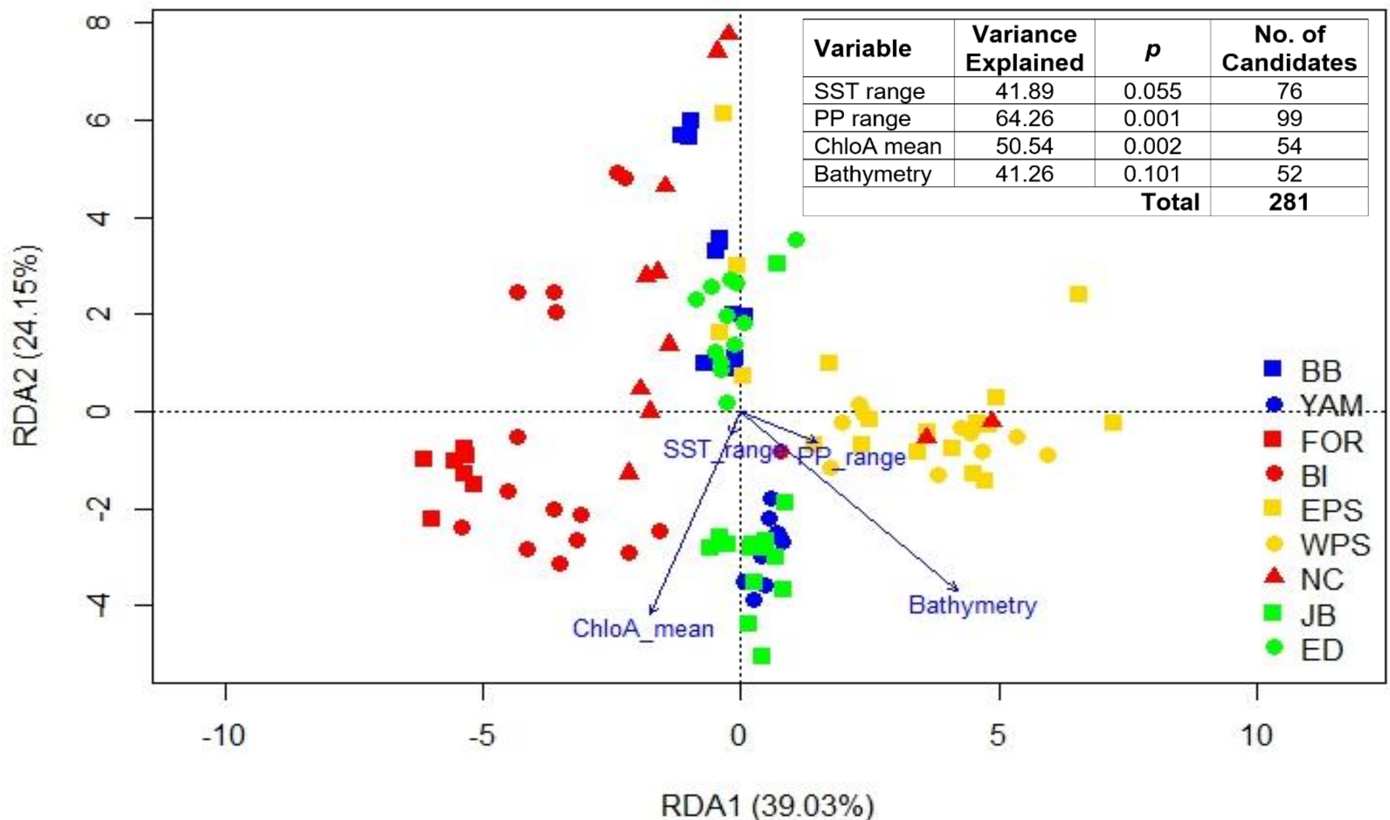


Figure 3.4 Mantel test for analysis of isolation by distance (IBD) in Indo-Pacific bottlenose dolphins from nine sampling localities along the east coast of Australia. Correlation between neutral (crosses) and adaptive (blue circles) genomic distance (linearised F_{ST}) and geographical distance along the coast (kms) among sampling locations is displayed.

3.6.2 Genotype-Environment Associations

Twenty-six individual environmental parameters (based on eleven variables) were selected for GEA analysis. The forward selection procedure identified four ecological variables significantly associated ($p < 0.05$) with genomic variation: SST range, primary productivity range, mean chloA and bathymetry. All four variables were under the VIF threshold of three, implying very low levels of collinearity (Appendix C: Figure C.i.3). As such, no further variables were removed before running the final model. Heterogeneity in these four variables was visually evident across the study region and between Hunter Coast and PS regions in particular (Appendix C: Figure C.i.1). The overall model was significant ($p = 0.001$), with 9.55% of variance in the genomic dataset explained by the spatial variables and 4.81% explained by the four environmental parameters (Figure 3.5; Appendix C: Table C.ii.5). Of these parameters, primary productivity range and mean chloA were significant ($p < 0.05$). There were 281 loci with scores ±3SD from the mean of at least one of the constrained RDA axes. These loci were retained as potential candidates for selection responding to variation in the

environment. When plotted, the primary division on the first RDA axis (39.03% of the constrained variance) was associated with primary productivity range and bathymetry, leading to a separation between the Hunter Coast and PS populations. The second RDA axis (24.15% of the constrained variance) highlights a division within northern NSW and southern NSW populations based on mean chloA and to a lesser extent SST range (Figure 3.5).



Variable	Variance Explained	*p*	No. of Candidates
SST range	41.89	0.055	76
PP range	64.26	0.001	99
ChloA mean	50.54	0.002	54
Bathymetry	41.26	0.101	52
		Total	**281**

Figure 3.5 Redundancy analysis (RDA) in genotype-environment association (GEA) analysis, displaying the influence of the four retained ecological variables on individual genomic variance coloured by sampling locality. Overall variance in the genomic dataset explained by the model was 215.54 (p = 0.001) (residual variance = 3,839.07), with 9.55% explained by space and 4.81% explained by the four ecological variables (see Appendix C: Table C.ii.5 for more information). Table inset details the variance explained, significance (p) and number of candidate loci most highly correlated with each of the four retained variables. Details of the ecological variables used are provided in Appendix C: Table C.ii.2 (SST = sea surface temperature, PP = primary productivity, ChloA = Chlorophyll A concentration). Sampling location abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3) and are ordered from north (top) to south (bottom) in the legend. Colours correspond to neutral population membership.

3.6.3 *Adaptive Population Genetic Diversity and Structure*

The 281 GEA candidate loci were extracted from the full dataset to form the putatively adaptive dataset. Genetic diversity was relatively high at all sampling localities, with Newcastle and PS embayment dolphins having considerably higher diversity than dolphins elsewhere in the study region (Appendix C: Table C.ii.3). Strong adaptive population structure was revealed for IPBD along the NSW coast. *Admixture* analysis revealed three adaptively divergent populations, with major adaptive differentiation between the embayment population of PS and the open-coast dolphins (Figure 3.2Bi1) (note: open-coast communities are referred to here as dolphins from sampling localities outside of the PS embayment). Dolphins sampled to the north and south of PS appear to be differentiated and further division within these groups is also evident (Figure 3.2Bi2 and 3). PCA corroborated these findings, showing strong separation of PS dolphins from others and clear differentiation between the open-coast dolphins found to the south and north of PS (Figure 3.2Bii). DAPC supported the presence of these three adaptive populations (i.e. north of PS, PS and south of PS), but also highlighted major differentiation between individuals from the southernmost sites (JB and Eden) and all other localities and suggested that PS dolphins were more similar to those from northern sites (Figure 3.2Biii). Differentiation estimated using the F_{ST} index was found to be moderate across the region, with an average of 0.099, substantially higher than that recorded for the neutral dataset (0.051) (Figure 3.3). All pairwise comparisons of sampling localities were significant, except between east and west PS. The pattern of divergence highlighted the adaptive differences between PS and other localities, as well as that the southernmost sites are highly differentiated at the adaptive level (Figure 3.3). The Mantel test suggested that IBD does not play a major role in the adaptive divergence of IPBD in coastal NSW ($r = 0.1726$, $p = 0.1770$) (Figure 3.4).

3.6.4 Functional Enrichment Analysis and Annotation

Of the 14,466 loci blasted, a total of 823 loci (5.7%) scored BLAST hits and were mapped and annotated, of which 2.3% were GEA candidate loci. Functional enrichment analysis revealed 50 GO terms to be significantly over-represented in the putatively adaptive dataset compared to the full dataset (Appendix C: Table C.ii.6). This included GO terms such as bone morphogenetic protein signalling pathway (GO:0030509), MHC protein binding (GO:0042287), galactosyltransferase activity (GO:0008378), cation transmembrane transporter activity (GO:0008324) and several associated terms (for significance values see Appendix C: Table C.ii.6). Twenty-four loci moderately to highly correlated with mean chloA and/or primary productivity range were functionally annotated using the NCBI BLAST. From these 24 loci, 11 candidate genes were identified within 20KB of the candidate SNP (Appendix C: Table C.ii.7). Candidate genes associated with mean chloA had broad functions involved in regulating receptor function. By contrast, those associated with primary productivity range were involved in a variety of processes, including the development and functioning of vital bodily systems. Specifically, several candidate genes, including *VEGFA*, *SMYD1, OSTF1* and *THNSL2,* showed dramatic increases in the frequency of the minor allele in the central region, particularly in the PS community. Allele frequencies for these genes in the northern and southern communities were much more similar to one another (Appendix C: Figure C.i.4). The candidate gene *ELOVL2* on the other hand, showed a pattern of allele frequency change closely correlated with the primary productivity range gradient (Figure 3.6).

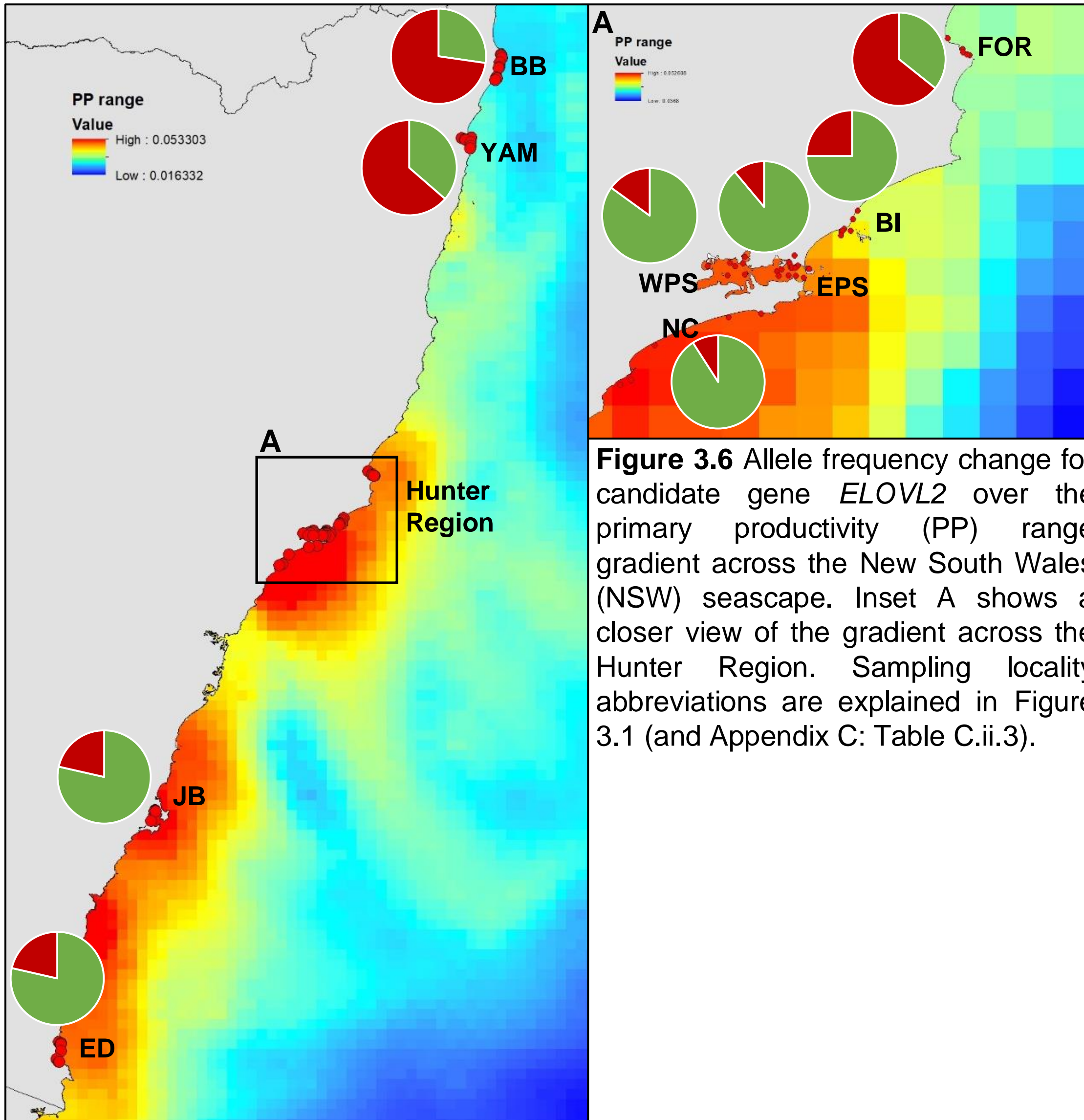


Figure 3.6 Allele frequency change for candidate gene *ELOVL2* over the primary productivity (PP) range gradient across the New South Wales (NSW) seascape. Inset A shows a closer view of the gradient across the Hunter Region. Sampling locality abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3).

3.7 Discussion

A seascape genomics approach was used to assess population genomic structure of IPBD and its relationship with environmental heterogeneity along eastern Australia. Spatial and environmental gradients are suggested to influence connectivity and adaptation of IPBD in this region. Specifically, the hypothesis of IBD and ocean circulation impacting on the neutral population structure of IPBD in eastern Australia was supported, while fine-scale adaptive differentiation was found to be associated with environmental heterogeneity over spatial scales as small as approximately 10 kilometres. This was particularly driven by variation in productivity and habitat type. Genomic differentiation is thus potentially occurring through a complex interplay of prey specialisation, social structure and physiological adaptations to differing diets and habitat characteristics. The high-quality genomic dataset generated here substantially improved the power and resolution to investigate fine-scale population genomic structure and allowed GEA to be formally tested. With environmental conditions potentially influencing the evolutionary trajectory of this near-top marine predator, further climate change impacting on these dolphins may have top-down impacts on marine ecosystems.

3.7.1 Genomic Variation

Levels of genomic variation in a population are affected by demographic history through changes in effective population size (N_e) over time due to founder events, strong bottlenecks and/or climatic change and habitat alterations (Bjørnstad and Grenfell 2001; Ellegren and Galtier 2016). Founder events have been suggested to drive the pattern of low genomic variation in estuarine and embayment populations of bottlenose dolphins compared to pelagic populations (Möller et al. 2007; Louis et al. 2014b; Bayas-Rea et al. 2018). Bottlenose dolphin social structure and natal philopatry is also typically very strong in sheltered habitats and may be a contributing factor to the lower levels of genomic diversity often recorded (Möller 2012; Pérez-Alvarez et al. 2018), as previously documented for IPBD in the PS embayment (Wiszniewski et al. 2010). In contrast, it was found that the PS communities had similar levels

of neutral genomic diversity to dolphins from other localities and diversity levels appear to vary irrespective of habitat or population. This probably reflects the influence of recent gene flow from the Hunter Coast population into PS, and perhaps also the strength and type of selection at play. Given that diversity levels based on the candidate dataset were substantially higher than neutral diversity across all localities, it is hypothesised here that balancing selection maintains higher levels of genomic variation compared to the genome-wide baseline, as heterozygosity in genes affecting fitness is actively retained (Charlesworth 2006; Hedrick 2007). On the other hand, natal philopatry and local adaptation through directional selection, combined with occasional long-range migration, could perhaps also lead to high levels of heterozygosity. Distinguishing between these alternative hypotheses requires an in-depth study to characterise both the strength and the different modes of selection. Such study, which is outside the scope of this thesis, is not only analytically challenging but requires information about the adaptive value of candidate SNPs and phenotype-genotype interactions (Bernatchez 2016) that is not currently available for bottlenose dolphins.

3.7.2 Population Structure and Adaptation

Population differentiation of both terrestrial and marine species has long been tied to changes in environmental conditions and the associated selective forces (Briggs 1974; Cowen and Sponaugle 2009). In the marine system ocean circulation, SST, salinity and productivity are common environmental factors influencing population genomic differentiation (Benestan et al. 2016; Diopere et al. 2017; Leon et al. 2018; Bernatchez et al. 2019). Gradients in these and other variables can drive physiological adaptations and lead to ecological barriers to gene flow among populations. Indeed, despite a general lack of obvious physical boundaries to the movement of cetaceans, coastal bottlenose dolphins typically show fine-scale population structure (Möller et al. 2007; Wiszniewski et al. 2010; Fruet et al. 2014b; Bayas-Rea et al. 2018; Pratt et al. 2018). This suggests that there are underlying factors influencing the genomic differentiation of bottlenose dolphins that are not immediately apparent. Potential

explanations are discussed below in two sections, one focusing on neutral demographic signal and the other on putatively adaptive genetic variation.

3.7.2.1 Neutral Differentiation

3.7.2.1.1 Spatial Influence of the East Australian Current

In eastern Australia, the connectivity and neutral substructure of IPBD populations appears to be largely driven by space and patterns of ocean circulation. The EAC strongly impacts the latitudinal gradients of SST, productivity and salinity across the region and in turn influences the distribution and gene flow of local biota (Keane and Neira 2008; Suthers et al. 2011; Oke et al. 2019). Along its NSW range, this current has been classified into three major water masses (Keane and Neira 2008; Oke et al. 2019). Population genomic structure of IPBD at the neutral level closely mirrored these water masses, forming three major populations, as previously found for these dolphins by Wiszniewski et al. (2010), and also for common dolphins (*D. delphis*) (Möller et al. 2011) based on microsatellite markers. Wiszniewski et al. (2010) previously referred to these populations as northern NSW, Hunter Region and southern NSW and these will also be referred to here. Distinct larval fish assemblages have also been found in each of the three water masses (Keane and Neira 2008). Despite the absence of physical barriers to the dispersal of marine organisms in eastern Australia, marine molluscs (*Bedeva hanleyi* (Hoskin 2000), *Donax deltoids* (Miller et al. 2013), and *Haliotis coccoradiata* (Piggott et al. 2008)), sea urchins (*Centrostephanus reodgersii*) (Banks et al. 2007; Banks et al. 2010), and several species of squid (Brandt 1983) also demonstrate latitudinal patterns of population structuring and abundance reflective of EAC circulation. This is also true for IPBD with a strong signal of IBD, likely reflecting the influence of social structure, thought to be an important driver of neutral population structure in inshore dolphins worldwide (Hoelzel 2009; Möller 2012). Coastal bottlenose dolphin populations typically exhibit high levels of female-dominated natal philopatry (Möller and Beheregaray 2004; Möller 2012), and long periods of maternal care, which allows them to become specialised to local resources (Hoelzel 2009). This includes the development and cross-generational transmission of specific prey handling

techniques, as seen in many populations of dolphins (e.g. Sargeant et al. 2005; Weiss 2006; Finn et al. 2009; Pittman and Durban 2012). Behavioural changes are often associated with habitat and environmental discontinuities and as such resource specialisation and natal philopatry can restrict gene flow among populations (Hoelzel 2009). This allows neutral processes, such as genetic drift and mutation to create genome-wide differentiation that may mimic the influence of environmental heterogeneity (Storz 1999; Whitlock 1999). Sociality and resource specialisation of these dolphins has not been studied intensively across NSW, but strong social structure has been documented in the PS and JB communities (Möller et al. 2001; Möller et al. 2002; Möller and Beheregaray 2004; Möller et al. 2006; Wiszniewski et al. 2009). With distribution and abundance of several species, including potential IPBD prey items impacted by the EAC (e.g. squid; Brandt 1983), it could be that the neutral structure evident across this region is mainly a result of resource specialisation and social structure of the IPBD communities; a similar process has been suggested to be occurring for common bottlenose dolphins (*T. truncatus*) in European waters (Natoli et al. 2005). While often alluded to, ocean currents have rarely been explicitly implicated in having an indirect effect on population differentiation in small cetaceans through their prey assemblages. This complex interplay may be occurring more frequently than currently documented. Indeed, in several regions inhabited by bottlenose dolphins, population genetic structure of fish species have been revealed to be affected by patterns of ocean circulation, including in Mediterranean (*Serranus cabrilla*; Schunter et al. 2011), Scandinavian (*Platichthys flesus*; Hemmer-Hansen et al. 2007), and New Zealand waters (*Pagrus auratus*; Bernal-Ramirez et al. 2003). Climate change is predicted to alter the position and strength of ocean circulation systems (Bakun et al. 2015; Van Gennip et al. 2017), and thus could have major impacts on both mid-lower order trophic groups and the higher order predators relying on them.

3.7.2.1.2 *Broad-scale Habitat Type Change*

Neutral differentiation was also reported between the PS and open-coast IPBD communities. This coincides with a major habitat type change across just tens of kilometres and highlights the hierarchical nature of the neutral population structure of IPBD across the region (although this could not be detected by DAPC due to restrictions of the method). Historical neutral population genetic divergence between dolphins from the Hunter Coast and PS regions was thought to be influenced by a founder event after the colonisation of the embayment (Wiszniewski et al. 2010). The historical signal could perhaps be diluted by the high level of contemporary migration moving from the larger Hunter Coast population into PS, consistent with sightings of movements of dolphins between the two areas (Möller et al. 2002). Wiszniewski et al. (2010) also suggested high levels of migration between PS and Hunter Coast based on microsatellite markers, but in the opposite direction. These opposing results are likely due to the strong neutral gene flow between the two regions, with this large SNP dataset providing more power to clarify the pattern and direction of contemporary migration. Coalescent-based demographic analyses using the powerful SNP dataset generated here represent a valuable research direction for future studies assessing demographic histories in IPBD associated with geomorphological changes in coastal habitats.

The presence of bottlenose dolphin fission-fusion societies documented among dolphins from each of the sampled sites along the Hunter Coast (Möller et al. 2007). This may be driving neutral genomic differentiation between PS and open-coast dolphins and could be related to habitat and prey assemblage variation and associated IPBD resource specialisation. Specifically, bathymetry differs markedly between these two regions. Differences in depth profile has been shown to be associated with strong social structure of inshore bottlenose dolphins in Australia (Zanardo et al. 2018; Diaz-Aguirre et al. 2019). In both studies the disclosed social and genetic division was linked to local adaptation to the different water depths and associated habitat types and prey assemblages. Worldwide, bottlenose dolphins

also exhibit strong genomic structure between inshore and offshore populations reflecting adaptations to selective pressures of the shallow versus deep-water environments (e.g. Hoelzel et al. 1998; Oudejans et al. 2015; Fruet et al. 2017; Chapter 2). In addition to bathymetric differences, seasonal upwellings in the PS embayment create short-lived plankton blooms that influence fish abundance in the region (Hallegraeff and Jeffrey 1993). Productivity changes are known to significantly alter fish assemblages worldwide (e.g. Ward et al. 2006; Condie et al. 2011; Santora et al. 2017; Tiedemann et al. 2017), including between the three major EAC water masses in NSW (Keane and Neira 2008). Furthermore, lower density of IPBD in the western reaches of PS may be reflective of reduced productivity in the area (Wiszniewski et al. 2009). Productivity is suggested to be associated with population genetic structure of small cetaceans by influencing foraging behaviour and social structure (Mendez et al. 2010; Mendez et al. 2011). It is therefore, possible that primary productivity and depth profile are impacting on the neutral population structure of IPBD by reinforcing social structure through specialisation on differing prey assemblages between Hunter Coast and PS. Indeed, Möller et al. (2007) stated that while the feeding ecology of these dolphins has not been studied, habitat differences related to substrate types (i.e. mud versus sand) between the two areas suggests that differing prey and feeding specialisations between the communities are likely. Habitat type change has been previously suggested to influence social structuring and subsequent neutral genomic differentiation in other similar systems (Hoelzel 1998; Amaral et al. 2012a; Möller 2012; Fruet et al. 2014b; Moura et al. 2014).

A complex interplay of the impact of the EAC circulation, IBD and broad-scale habitat changes are likely driving IPBD resource specialisation and social structure, and in turn influencing neutral genomic differentiation across NSW coastal bottlenose dolphins. This is reminiscent of killer whales (*Orcinus orca*), with both social and ecological factors influencing the strong population genomic structure seen among ecotypes (Hoelzel and Moura 2015). A similar process leading to genetic differentiation over fine spatial scales has been suggested for

rough-toothed dolphins (*Steno bredanensis*) around several central Pacific Ocean islands (Albertson et al. 2017), and false killer whales (*Pseudorca crassidens*) in the northern Pacific Ocean (Martien et al. 2014). Ongoing human-induced climate change and overfishing of prey stocks are therefore, likely to have dramatic impacts on cetaceans, potentially promoting further fragmentation of populations.

3.7.2.2 Adaptive Differentiation

It is largely assumed that habitat type differences and climate change will only indirectly affect megafauna by impacting on prey species distribution and abundance (Bakun et al. 2015; Sydeman et al. 2015). However, this study provides evidence that habitat type and environmental variation may also result in physiological adaptations and associated adaptive differentiation in upper trophic level predators. The signal of population differentiation detected in the adaptive dataset is likely due to adaptive divergence rather than the influence of spatial variables, supported by an insignificant Mantel test result for IBD in this dataset. With the RDA finding adaptive differentiation correlated with primary productivity and chloA gradients in eastern Australia, it may be that the associated differences in selective pressures are driving physiological adaptations in local bottlenose dolphins.

3.7.2.2.1 Open-coast Population Divergence

Over evolutionary history, the diversification of marine species, and cetaceans specifically, has been closely associated with periods and regions of high productivity in the world's oceans (Fordyce 1980; Berger 2007; Steeman et al. 2009; Pyenson et al. 2014; Barragán-Barrera et al. 2017). Several studies of population genetic structure of small cetaceans have also noted heterogeneity in productivity to be potentially driving population differentiation (Natoli et al. 2005; Mendez et al. 2010; Mendez et al. 2011; Amaral et al. 2012a; Amaral et al. 2017). This is also the case for many fish and invertebrate species worldwide (e.g. Nanninga et al. 2014; Diopere et al. 2017; Barahona et al. 2019). It is thus not surprising that gradients in productivity

(primary productivity and chloA) could be associated with the adaptive differentiation of IPBD in NSW. While primary productivity range and mean chloA were included in the final RDA model, they were highly correlated with other variables (e.g. mean, range, maximum and/or minimum, removed during forward selection; data not shown). Primary productivity and chloA metrics will therefore, be referred to as general primary productivity and chloA gradients. It appears that primary productivity in particular is influencing adaptive differentiation between the northern, central and southern populations and also within the Hunter Region (central) population. Gradients of chloA on the other hand, were shown to be influencing fine-scale adaptive divergence within the northern NSW and southern NSW populations, with differential upwelling and circulation patterns previously documented within each of these regions (Hallegraeff and Jeffrey 1993; Oke and Middleton 2001; Lee et al. 2007).

While several studies have inferred an influence of productivity on the population genetic structure of small cetaceans (Natoli et al. 2005; Mendez et al. 2010; Mendez et al. 2011; Amaral et al. 2012a; Amaral et al. 2017), very few have investigated the genomic basis to potential physiological adaptations of the dolphins to productivity gradients. In the Gulf of Mexico however, spikes in phytoplankton productivity were found to be imposing strong selective pressures on the resident common bottlenose dolphins (*T. truncatus*; Cammen et al. 2015a; Cammen et al. 2015b). While algal blooms across the NSW coast have not yet been as deadly as in the Gulf of Mexico, strong variation in productivity might act as a selective agent on local bottlenose dolphins. The primary productivity range gradient in NSW was followed closely by allele frequency changes in a homolog of the candidate gene, *ELOVL2*. Minor allele homozygotes were substantially more prevelant in dolphins in the north, corresponding to areas of high primary productivity variation, compared to the populations further south. On the other hand, homozygotes at the major allele were almost absent from the northern region compared to being present in about 70% of individuals sampled from BI southwards. This suggests a potential role of directional selection on dolphins responding to

differing selective pressures in the northern and central/southern areas of NSW, subsequently pushing the two variants toward fixation in the opposing regions. *ELOVL2* codes for a protein that elongates short polyunsaturated fatty acids into important bioactive long-chain molecules, such as EPA and DHA (Monroig et al. 2016). These molecules are essential for the development and normal functioning of the brain and body (Horrocks and Yeo 1999; Siriwardhana et al. 2012). The ability of an individual to convert short fatty acids into long bioactive chains is dependent on specific *ELOVL* gene variants (Monroig et al. 2016). The closely related gene, *ELOVL3*, was inactivated in the cetacean stem lineage and is associated with the disappearance of the sebaceous glands and subsequent adaptation of thermoregulatory and lipid deposition pathways in the macroevolution of cetaceans (Huelsmann et al. 2019; Lopes-Marques et al. 2019). *ELOVL2* may be under differential selection in IPBD in NSW also in response to differing lipid metabolism requirements. With variation in this gene correlated with the productivity gradient in NSW, it may be that physiological adaptations of IPBD is a response to preferred prey itemsacross the region. Expression of this gene is indeed affected by differences in diet, with salmon fed opposing diets showing significantly altered *ELOVL2* expression profiles (Morais et al. 2009). Both EPA and DHA are found in high quantities in fatty fish such as salmon, tuna and mackerel (Horrocks and Yeo 1999); all potential prey items of coastal bottlenose dolphins (e.g. Gibbs et al. 2011). DHA levels are much higher in the marine food chain than the freshwater and genomic adaptation has been shown in freshwater populations of several teleost species to cope with this depletion (Ishikawa et al. 2019). Although yet to be studied, it may be that variation in IPBD diet across the study region has created differences in the requirements for DHA and EPA to be synthesised in the body compared to sourced through their diet and has subsequently driven differential selection in *ELOVL2*.

Genes involved in fatty acid biosynthesis and metabolism have been previously implicated with diet-related genomic differentiation in mammals. Specifically, physiological adaptation

and genetic differentiation in killer whale (*O. orca*) ecotypes, polar bears (*Ursus maritimus*) and human populations (Greenlandic Inuit people and native Siberians) were associated with modifications of fatty acyl and lipid metabolism pathways in response to differences in the lipid content of their respective diets (Clemente et al. 2014; Liu et al. 2014; Fumagalli et al. 2015; Foote et al. 2016). This suggests that differences in the lipid content in the diet of mammals can drive genomic differentiation and physiological adaptations. While the ecological mechanism has been proposed for other cetaceans (Amaral et al. 2012a; Möller 2012; Fruet et al. 2014b; Moura et al. 2014), genomic evidence of potential physiological adaptations was lacking until now. Additional research on prey distributions and the feeding ecology of IPBD along NSW is needed to fully evaluate this hypothesis.

3.7.2.2.2 Port Stephens Embayment Divergence

Adaptive differentiation between the PS and open-coast IPBD populations was correlated with primary productivity and depth gradients. Four candidate genes of particular note were found: *VEGFA, SMYD1*, *OSTF1* and *THNSL2*. For each of these genes, major increases in the frequency of the minor allele were found in the PS region and homozygotes at this allele were substantially more common in PS than at other locations. These genes are associated with the development of the cardiac, circulatory and skeletal systems. Specifically, *SMYD1* is an important factor in the formation of cardiac muscle tissue (cardiomyogenesis) and early heart development (Sirinupong et al. 2010; Du et al. 2014). *VEGFA* is involved in vascular development stimulated by hypoxia, which acts to increase blood circulation to areas lacking in oxygen (Nagy et al. 2007; Claesson-Welsh and Welsh 2013). *THNSL2* and *OSTF1* genes on the other hand, are involved in bone development pathways (Rifas and Weitzmann 2009; Vermeren et al. 2017). This was also supported by several significantly over-enriched GO terms associated with bone morphogenetic proteins. Changes in the skeletal, cardiac and circulatory systems of cetaceans have been linked to living and hunting at different depths (see Chapter 2). This includes modification of blood and muscle oxygen storage capabilities in bottlenose dolphins and other diving marine mammals (Hersh and Duffield 1990; Mirceta et

al. 2013), and changes to the density and structure of bones to adjust buoyancy in killer whales and bottlenose dolphins (Foote et al. 2015). Several other marine species, including pinnipeds (Van Citters et al. 1965; Hindell et al. 1992), scorpion fishes (Scorpaenidae family; Yang et al. 1992), and dogfish (Squalidae family; Treberg et al. 2003) also exhibit adaptations of the heart and circulatory systems to enable diving behaviour in deep environments. The depth gradient between the Hunter Coast and PS regions is however, not remarkable (Appendix C: Figure C.i.1) and bathymetry was not significant in the final RDA model. Despite this, the shallower conditions in PS may still be driving minor adaptations for hunting and survival in the local IPBD. This suggests that even small changes in depth may have impacts on the physiological requirements of cetaceans.

The *OSTF1* gene also has an important role in an individual's response to osmotic stress, with expression levels in mice changing under hypertonic conditions to protect the cells from damage (Fiol et al. 2007). Similar changes occur in eels (*Anguilla japonica*) transferred between fresh and saltwater (Tse et al. 2012). Accordingly, functional enrichment analysis revealed over-representation in inorganic cation transmembrane transporter activity and cation transmembrane transporter activity GO terms in the candidate dataset. Although waters within the PS embayment are largely of a marine salinity level (Schröder-Adams et al. 2014), resident dolphins move in and out of the local rivers where lower salinity levels are experienced (Möller et al. 2002). Allele frequency changes at the *OSTF1* gene may therefore, be influenced by the use of the riverine systems by PS dolphins compared to those along the open-coast, as well as dealing with lower salinity levels during periods of heavy rain fall and river outflow to the embayment. While salinity and distance to nearest major river mouth did not register as significant drivers of IPBD genomic differentiation in the RDA, this may be influenced by a lack of sampling beyond the river mouths. Selection on genes involved in ion transport pathways is common in the adaptation of many marine species to estuarine and freshwater environments. This includes several teleosts (Czesny et al. 2012; Kozak et al.

2014), blue mussels (genus *Mytilus*; Lockwood and Somero 2011), inshore bottlenose dolphins (Chapter 2) and freshwater finless porpoises (*Neophocaena asiaeorientalis asiaeorientalis*; Ruan et al. 2015; Yuan et al. 2018). This demonstrates that while the PS dolphins do not reside in freshwater permanently, the use of this habitat on occasion may still be driving adaptation to less saline environments. Even small alteration to the environment and/or distribution of marine animals may therefore, impose disproportionately strong selective pressures on them. This may particularly be the case where strong social structure is reinforcing genomic differentiation at both neutral and adaptive levels, as possibly the case for bottlenose dolphins along the Hunter Coast. With climate change already affecting species distribution and salinity levels of the world's oceans, adaptation of ion transport pathways may become increasingly important to marine species.

3.7.3 Conservation Implications

This study furthers our understanding of the relationship between marine organisms and the environments they live in. Not only does the use of genomic techniques allow us to have significantly higher power to detect fine-scale population structure (Cammen et al. 2016; Vendrami et al. 2017; Attard et al. 2018), we can also differentiate between adaptive and neutral variation (Kelley et al. 2016). This was particularly important in highlighting the adaptive divergence of PS from the open-coast NSW IPBD communities and differentiation between the northern and southern regions. This is the first time a seascape genomics framework has been applied to IPBD, enabling us to move past pattern-based suggestions to formally test GEAs. Several studies have highlighted that this approach significantly increases the power to detect genes under selection (Forester et al. 2018; Grummer et al. 2019). In particular the RDA modelling method employed here has been previously shown to be one of the most powerful approaches currently available for reduced-representation genomic datasets (Forester et al. 2018).

With marine environments changing rapidly, it is important to uncover the most influential selective forces on population differentiation in marine species. In line with worldwide changes to ocean currents, the EAC is showing alteration in circulation patterns and becoming warmer and saltier in response to climate change (Cai et al. 2005; Ridgway 2007; Suthers et al. 2011). The EAC influences the entire ecosystem along Australia's east coast, including regulation of upwellings and areas of enhanced productivity (Lee et al. 2007; Everett et al. 2014). With ocean circulation and productivity suggested here to have major effects on the genomic differentiation of IPBD, climate change is likely to have a strong impact on the genomic differentiation of local dolphins and the wider marine communities. Furthermore, physiological adaptations to prey assemblages appears to be occurring in the NSW IPBD and thus these changes could be even more detrimental to marine mammals than first thought. This is likely to impact on entire food webs through both top-down and bottom-up processes. Management and conservation strategies for these dolphins should be developed to accommodate for changes to the selective forces impacting across the region. Specifically, habitat fragmentation must be prevented across the NSW region in order to preserve the hierarchical population genomic structure, while also facilitating gene flow between adjacent populations. This will likely enhance the adaptive capacity and resilience of these dolphins, providing them with the genomic tools to potentially adapt to climatic change. It is important to acknowledge the strong genomic divergence in bottlenose dolphins across the region, in particular segregation of the northern, central and southern populations and further division between the central Hunter Region communities. While these populations are significantly differentiated there is still a moderate to high level of gene flow and admixture among dolphins along the NSW coast and therefore, it is also important to note that management decisions in one region are likely to impact on IPBD from other regions. By utilising this genomic dataset recommendations can now also be based on adaptive differences between populations. In particular, adaptive differentiation of the PS dolphins from the open-coast communities is evident. This should be considered in management strategies as stressors and selective forces are likely to affect the embayment and open-coast regions differently. This will allow for the conservation of

potentially important genomic diversity within the east coast of Australia IPBD metapopulation. In managing these populations policy makers must consider how changes in the EAC and patterns of productivity across NSW are likely to affect the local dolphins in the future. Processes by which IPBD are potentially adapting to their local habitat and prey were identified, with similarities highlighted between the adaptation of these dolphins and other mammals based on diet. Ongoing climate change, as well as overfishing, is already having pronounced effects on prey distribution and abundance and in turn is likely to impact on marine mammals and other top predators. Without modelling, it cannot be foreseen how changes to the ecosystem will impact marine species. A holistic and predictive approach considering whole food webs and the way ocean circulation systems, such as the EAC, will change in the future, is needed to effectively conserve marine species. In this way the complex array of factors that are, and will be, impacting on population genomic differentiation and adaptation into the future can be accounted for and mitigated.

3.8 Conclusion

A seascape genomics framework was utilised to evaluate environmental and spatial effects on population genomic structure of IPBD across eastern Australia. Patterns of neutral and adaptive genomic differentiation differed across the study region. Local ocean circulation, broad-scale habitat type change and gradients in productivity were revealed to be impacting the IPBD population genomic structure across the NSW seascape. Driven by IBD and EAC circulation, habitat and prey specialisation coupled with strong social behaviours are likely to be affecting this pattern at the neutral level. At the adaptive level however, variation in habitat and diet appear to be driving physiological adaptations in local dolphins. A number of candidate genes were identified to be involved in adaptation of several major physiological systems and pathways, including the synthesis of important long-chain fatty acids. This could be related to differences in diet in association with productivity and prey assemblage variation across the NSW seascape. Many of the pathways and body processes potentially under selection in IPBD populations have been previously found to be important in the adaptation of other mammals and marine species to particular habitats and diets, and thus could have important implications to wider marine communities. Furthermore, this study demonstrates that marine mammals may adapt to particular habitats and prey items through not only behavioural and social mechanisms, but also physiological, prompting this research to be applied to other predators. With climate change being especially pronounced in the marine realm, including Australia's east coast, it is becoming increasingly important to evaluate the adaptive capacity of marine species. This ensures that management strategies are well-informed about how the environment is affecting marine ecosystems and thus how climate change is likely to impact on them into the future.

Chapter 4 : Seascape Genomics of Southern Australian Bottlenose Dolphins (*Tursiops* cf. *australis*)

4.1 Contributions

Eleanor Pratt – conception of study and design of methods, collection of Gulf St. Vincent samples, DNA extraction and ddRAD library preparation, bioinformatics and analysis, writing of thesis.

Luciana Möller – primary supervisor – conception of study design, guidance with interpretation, assistance with collection of Gulf St. Vincent samples, drafting and revision of thesis.

Luciano Beheregaray – associate supervisor – guidance in design and interpretation, drafting and revision of thesis.

Kerstin Bilgmann – collection of Spencer Gulf, St. Francis Island and Western Australian samples.

Nikki Zanardo – collection of Adelaide samples, and assistance in collecting other Gulf St. Vincent samples.

Fernando Diaz-Aguirre – collection of Coffin Bay samples.

Jonathan Sandoval-Castillo – guidance for laboratory methods and bioinformatics and assistance with analysis.

Chris Brauer – guidance and assistance with analysis.

4.2 Permits and Ethics Approvals

Biopsy samples were collected with Ministerial Exemption from Primary Industries Resources South Australia (PIRSA), exemptions #9902648, #9902714 and #9902601, with permits #K25761-6 and #E26171 from the Department of Environment, Water and Natural Resources (DEWNR), South Australia and with permit from the Department of Environment and Conservation, Western Australia (#SF008961). Animal ethics approvals were acquired from the Flinders University Animal Welfare Committee, projects #E310, #E375 and #E326.

4.3 Abstract

Environmental variation has long been thought to influence fine-scale genetic differentiation in coastal marine organisms. With the growing application of genomic methods for non-model species, associations between genetic variants and environmental variables can now be empirically tested. A high-quality ddRAD dataset of over 8,000 SNPs was utilised for a seascape genomic analysis of southern Australian bottlenose dolphins (*Tursiops* cf. *australis*; SABD). Population genomic structure was elucidated using neutral and putatively adaptive loci separately, identifying a hierarchical metapopulation and adaptively divergent populations, respectively. GEA analysis identified 241 SNPs with a putative signal of selection and revealed that SST and salinity gradients along the southern Australian coast are key drivers of adaptive divergence in these dolphins. In particular, contrasting environmental conditions between and within South Australia's gulfs and embayments may be driving adaptive divergence of dolphins across scales from thousands of kilometres to less than one hundred. Functional enrichment analysis and candidate gene annotation revealed functions related to sodium-activated ion transport, kidney development, adipogenesis and thermogenesis. Modification of these genes is likely associated with the adaptation of bottlenose dolphin's osmoregulatory and thermoregulatory systems in response to marked salinity and SST variation, respectively. Coastal bottlenose dolphins from southern Australia are a distinct dolphin lineage, have a restricted range and are impacted by several human activities, making them particularly vulnerable to population declines. This study clarifies spatial and environmental drivers of genomic divergence across a large section of the lineage's range, as well as pathways of local adaptation to be considered in evolutionary management of dolphin populations under ongoing climatic and habitat changes.

4.4 Introduction

Over evolutionary time, changing environmental conditions and the colonisation of new habitats have been the primary drivers for the radiation of species (Wellborn and Langerhans 2015; Stroud and Losos 2016). The formation of new species and population-level differentiation is referred to as microevolution (Jablonski 2000). As species expand their geographical range, populations commonly become more specialised to their respective habitats and niches. This can in turn create subdivision within a lineage, which is exacerbated in cases where the species exhibits strong social structure, natal philopatry and/or small home ranges (Storz 1999). This form of ecological speciation is also often termed divergence with gene flow and does not necessarily require physical isolation for the creation of new species and lineages (see Cooke et al. 2012). Differing selective pressures in adjacent regions can drive natural selection to create differentiation among populations as they adapt to the local conditions even in the absence of a hard-physical barrier. Neutral processes, such as mutation and genetic drift, are likely to aid in the differentiation of populations, particularly where population sizes are small (Willi et al. 2007). Both adaptive and neutral processes are therefore, important factors to consider when investigating the formation of species and population differentiation within species.

In the marine environment, gene flow and resulting population structure can be highly variable depending on the life history characteristics of the species. High gene flow and large-scale dispersal have been thought to be typical of marine organisms. This paradigm is now being questioned with many species of fish and invertebrates revealing much finer-scale population structure than previously suggested (e.g. Hoffman et al. 2012; Teske et al. 2015). Species of marine megafauna, including dolphins, pinnipeds and turtles, also have been documented to have population structure seemingly at odds with their highly mobile nature (e.g. Möller et al. 2011; Matsuzawa et al. 2016; Ahonen et al. 2016; Amaral et al. 2017). Environmental heterogeneity in the marine environment is thus potentially much more influential in shaping

population structure than previously thought. Landscape genomics presents an exciting framework to investigate the role of environmental heterogeneity in shaping the microevolution of populations. This field of study investigates genome-wide neutral and adaptive variation of populations across heterogeneous landscapes to address novel or previously intractable questions, such as forecasting adaptive capacity under environmental change (Grummer et al. 2019, pp. 1). While terrestrial landscape genomics has quickly become common place, the seascape equivalent is lagging behind (Grummer et al. 2019). Marine-based GEA studies have up until now been largely focused on commercially important species of fish and invertebrates (e.g. Sandoval-Castillo et al. 2017; Diopere et al. 2018), but are beginning to be utilised for a greater diversity of marine species (e.g. Rodriguez-Zarate et al. 2018; Teske et al. 2019). The seascape genomics framework allows for empirical testing of the relationship between particular environmental, topographic and oceanographic variables, and genomic variation. Patterns of genomic variation based on neutral, compared to putatively adaptive loci, can also be untangled to reveal populations that may be strongly influenced by natural selection. Neutral loci are defined here as those that have no apparent effect on the fitness of an individual or population. Adaptive loci are those that putatively do have an effect on fitness and therefore, alleles at these loci can potentially be selected for or against by natural selection (Holderegger et al. 2006). The separate consideration of adaptive and neutral loci is a major asset of this approach, allowing scientists to better understand the processes shaping population genomic structure. This also gives us the opportunity to evaluate the adaptive potential of species in response to ongoing ocean and climate change (Grummer et al. 2019).

Cetaceans provide an excellent case for which to study both macroevolution and the lower level microevolutionary differentiation as lineages become specialised to specific niches. Cetaceans transitioned from a land-based to a fully aquatic lifestyle approximately 53 MYA (McGowen et al. 2014). They have since radiated throughout the world’s oceans and diversified into at least 86 species (Reeves et al. 2003). Within the parvorder Odontoceti

(toothed whales and dolphins), Delphinidae is the most diverse family with at least 37 species of dolphins and porpoises (Fordyce 2009; Amaral et al. 2012b). This family originated approximately 26 MYA, but began to radiate dramatically in the late Miocene (~11 MYA), coinciding with the extinctions of several families of small odontocetes and changes in ocean circulation and patterns of productivity (Steeman et al. 2009). The relatively recent radiation of this taxon have led to much confusion and disagreement in the taxonomic classification of delphinid species (Amaral et al. 2012b). Incomplete lineage sorting due to recent rapid speciation events and hybridisation has created discordance between the species trees estimated by different genetic markers and analytical techniques (Amaral et al. 2012b). This is especially evident in the subfamily Delphininae and even more so in the genus *Tursiops* (bottlenose dolphins) where up to as many as 20 species have previously been described (Hershkovitz 1966), although only two are formally recognised today (Committee on Taxonomy of the Society for Marine Mammalogy 2019).

Common bottlenose dolphins (*Tursiops truncatus*) are found in tropical and temperate waters worldwide in both coastal and pelagic environments, while in the Indo-Pacific, the coastal-type is classified as *T. aduncus*, the Indo-Pacific bottlenose dolphin (Möller and Beheregaray 2001; Shirakihara et al. 2003; Perrin et al. 2007; Jayasankar et al. 2008). Several other species and subspecies have been proposed worldwide, with mixed levels of acceptance from cetacean scientists. One such species is the Burrunan dolphin (*T.* cf. *australis*), proposed by Charlton-Robb et al. (2011) based on multiple lines of evidence, including genetic and morphological characteristics (Charlton et al. 2006; Möller et al. 2008; Charlton-Robb et al. 2011). This however, has recently been found to represent a sister taxon to other *T. aduncus,* suggesting it should be considered at a subspecies level instead (Moura et al. 2020; Chapter 2). Due to ongoing controversy around the validity of this taxon, this lineage will be hereafter referred to as SABD. This lineage is endemic to the coastal waters of southern Australia and thought to largely exist in small, restricted populations with significant population genetic structure over

fine geographical scales (Bilgmann et al. 2007b; Charlton-Robb et al. 2015; Pratt et al. 2018). With strong levels of philopatry (inhabitation of the natal environment or group for the duration of their lives) exhibited by bottlenose dolphins (e.g. Möller & Beheregaray 2004), it is likely that they are highly specialised to local resources and environmental conditions (see Louis et al. 2014). While this is beneficial to populations inhabiting a stable environment, SABD reside in close proximity to a number of rapidly changing, highly urbanised areas with heavy impacts on the marine ecosystem. This includes habitat disturbance and destruction, pollution (chemical, physical and noise), commercial and recreational fishing, tourism and climate change (e.g. ocean warming and heatwaves, and/or changes in circulation patterns) (Bilgmann et al. 2008; Charlton-Robb et al. 2015; Bossley et al. 2017; Bilgmann et al. 2019). As a result, they are currently classified as Endangered in Victoria under the Victorian Flora and Fauna Guarantee Act (DSE 2013). This is not uncommon for coastal dolphins, with many species being threatened by anthropogenic influences (Reeves et al. 2003). The vulnerability of SABD and many coastal cetacean species to population declines and extinction is exacerbated by a lack of knowledge about their ecology and demographics.

Accumulating research on SABD in recent years is helping to fill the gaps in knowledge surrounding this lineage (see Passadore et al. 2017, 2018a, 2018b; Zanardo et al. 2016a, 2016b & 2018; Diaz-Aguirre et al. 2018; Bilgmann et al. 2007b, 2019). The population structure of these dolphins has previously been investigated using traditional genetic techniques, including microsatellites and mtDNA (Bilgmann et al. 2007b; Charlton-Robb et al. 2015; Pratt et al. 2018). While these studies were an important first step, the methods used have now been superseded by genomic techniques. The use of genomics in this space allows for better resolution and higher power to detect patterns of population genomic structure over small spatial scales. In addition, the seascape genomics approach can potentially be particularly useful here, with SABD ranging over a highly heterogeneous seascape, including several separate bioregions with steep gradients in SST, salinity and primary productivity, among

other factors (IMCRA Technical Group 1998; Edyvane 1999; Day et al. 2008). The southern Australian coastline is characterised by several vastly different habitats. In the west, the Great Australian Bight dominates the seascape. This is an approximately 1,800 kilometre stretch of open ocean with strong wave action and very few sheltered bays (Edyvane 2000). In South Australia on the other hand, two large inverse estuaries, GSV and SG, provide relatively protected habitat for local marine species (Petrusevics 1993; Kämpf et al. 2010). These gulfs are however, starkly different to one another, with a strong gradient of hypersalinity in SG (Appendix D: Figure D.i.1D) and a more uniform habitat in GSV (IMCRA Technical Group 1998; O'Connell et al. 2016). Differing levels of environmental heterogeneity in the southern Australian seascape has been suggested to be associated with the variable pattern of both fine and large-scale population genetic structure in SABD (Pratt et al. 2018). For example, using microsatellites Pratt et al. (2018) found that SABD living in the environmentally variable SG had substantially more population genetic sub-structuring than those residing in the more homogeneous GSV. This is typical of coastal bottlenose dolphins, with genetic differentiation and social structure previously suggested to be influenced by environmental heterogeneity (Natoli et al. 2005; Wiszniewski et al. 2010; Möller 2012; Amaral et al. 2017). With high levels of anthropogenic influence and ongoing climate change affecting the local habitat of SABD, it is important to establish how these dolphins have adapted to different environmental conditions. This can then help to inform management strategies in regard to how SABD may respond to future changes in their environment.

The aims of this chapter are to investigate the population genomic structure of coastal bottlenose dolphins in southern Australia and to explore the relationship between patterns of genomic divergence and environmental heterogeneity using a seascape genomics approach. This chapter tests the hypothesis of fine-scale population genomic structure and adaptive divergence in the SABD lineage in southern Australia. While neutral population structure is hypothesised to reflect IBD and the influence of social structure, adaptive differentiation may

occur in response to strong environmental gradients and habitat type changes in this region. DdRADseq was used to create a dataset of 8,081 high-quality SNPs. GEA analysis was then used to empirically test for associations between SABD genomic differentiation and 24 environmental, oceanographic and topological variables. Informed by this analysis, candidate loci were separated from the putatively neutral dataset and population genomic structure and diversity were then investigated separately for the two marker sets. This study builds upon previous genetic studies of SABD to provide a better understanding of the threats faced by these dolphins, as well as their capacity to adapt to changing environmental and oceanographic conditions.

4.5 Methods

4.5.1 Sample Collection

Biopsy samples were collected from 139 free-ranging SABD at 12 locations between 2004 and 2015 (Figure 4.1). The samples spanned over six bioregions characterised by differing biological, oceanographic and environmental variables (see DEH 2006). Biopsy samples consisted of skin and blubber and were obtained by a hand-held biopsy pole (Bilgmann et al. 2007a) or a remote biopsy gun (Krützen et al. 2002). To reduce the risk of resampling, individuals were observed for fresh biopsy wounds or scars and recognisable dorsal fin characteristics. No dependent calves were sampled. Samples were stored in either 90% ethanol or a salt-saturated solution of 20% DMSO at -80ºC. Previous studies of Adelaide and Coffin Bay (CB) bottlenose dolphins revealed strong residency and small home ranges in these communities (Zanardo et al. 2016b; Passadore et al. 2018a; Diaz-Aguirre et al. 2019). As this is a typical feature of inshore bottlenose dolphin populations worldwide (see Möller 2012), a similar pattern is expected to be characteristic of the other dolphin communities sampled here. Microsatellite data available from Pratt et al. (2018) were used to select individuals for genomic analysis with the aim of avoiding closely related individuals. This was done based on a Queller and Goodnight (1989) estimator calculated in *GenAlEx* (Peakall and Smouse 2006, 2012), and by selecting only one individual of each pair that had an estimated relatedness value of ≥0.5 (theoretical value for first-order relatives).

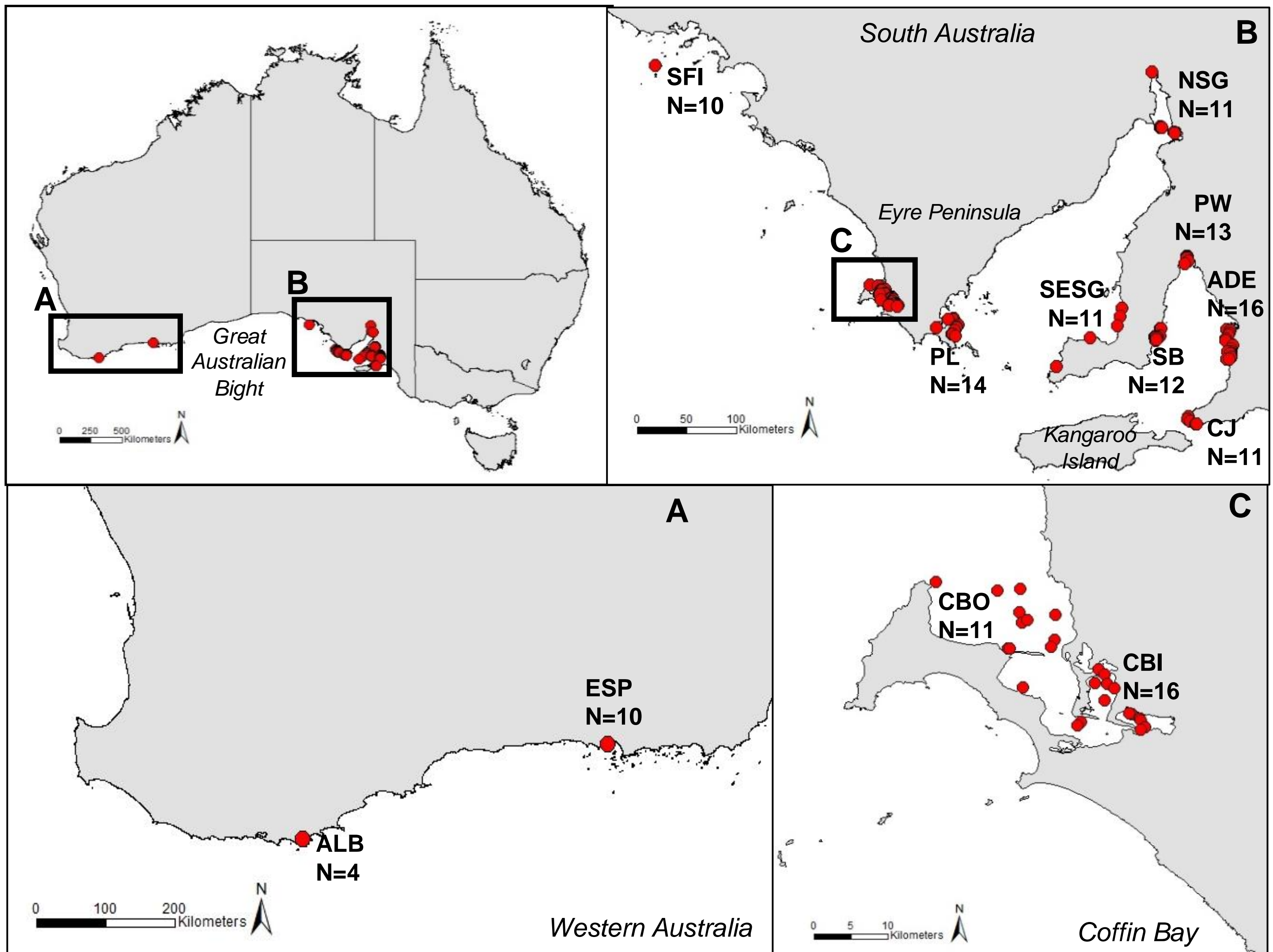


Figure 4.1 Sampling sites of southern Australian coastal bottlenose dolphins. From west to east sampling locality abbreviations are: ALB (Albany), ESP (Esperance), SFI (St. Francis Island), CBO (outer Coffin Bay), CBI (inner Coffin Bay), PL (Port Lincoln), NSG (Northern Spencer Gulf), SESG (Southeast Spencer Gulf), SB (Stansbury), PW (Port Wakefield), ADE (Adelaide) and CJ (Cape Jervis). Note: N= sample size before filtering.

4.5.2 Genomic Laboratory Methods

4.5.2.1 DNA Extraction

The salting-out protocol was used to extract DNA from the skin of the selected individuals as per Chapter 2 (Sunnucks and Hales 1996). Quality and quantity of DNA was checked using a Qubit 2.0 fluorometer (Life Technologies), NanoDrop 1000 spectrophotometer (Thermo Scientific) and gel electrophoresis. The sex of each dolphin was available from previous genetic studies (Bilgmann et al. 2007b; Pratt et al. 2018).

4.5.2.2 ddRAD Library Preparation

Libraries were prepared for ddRADseq using a modified protocol from Peterson et al. (2012) as described in Brauer et al. (2016) (detailed in Chapter 2). Each library was multiplexed to 96 samples and sequenced at the SAHMRI as per Chapter 2.

4.5.3 Bioinformatics

Raw data files were demultiplexed and processed using *dDocent* v.2.2.19 (Puritz et al. 2014). Details of this pipeline are found in Chapter 2. All *dDocent* scripts used here are in Appendix B.iii.1 and 2. Custom BASH scripts (Appendix D.iii.1) were used to run *VCFtools* to filter the resulting VCF as detailed in Appendix D: Table D.ii.1 (modified from Brauer et al. 2016). Loci were then mapped against the *T. aduncus* genome, downloaded from NCBI (GCA_003227395.1 ASM322739v1). Only loci that aligned to this genome were retained for analysis.

4.5.4 Neutral Population Genomic Diversity and Structure

Loci that were identified as potential candidates for selection by GEA analysis or any of the three additional outlier detection methods (see below) were removed from the full 8,081 SNP dataset to form the putatively neutral dataset, hereafter referred to as the "neutral dataset". For each of the 12 sampling sites molecular diversity indices, such as H_E, H_O and %PL, were calculated in *Arlequin* v.3.5.2.2 (Excoffier and Lischer 2010) using the neutral dataset. Wright's inbreeding coefficient (F_{IS}) was calculated for each sampling location as (H_E - H_O) / H_E (Wright 1922). The model-based maximum-likelihood method implemented in *Admixture* v.1.3.0 (Alexander et al. 2009) was used to investigate neutral population genomic structure. This was run in the command line with default settings, testing for *K* between one and 16 (number of sampling locations and social groups as informed by previous studies (Zanardo et al. 2018;

Diaz-Aguirre et al. 2019)). The *K* value with lowest CV error was selected as the most likely number of populations in the dataset. Two non-model methods were also used to investigate the population genomic structure of SABD. A PCA was run in R using the packages *adegenet* and *FactoMineR* to visually inspect the clustering of individual samples (Jombart 2008; Jombart and Ahmed 2011; Francois et al. 2015). A DAPC in *adegenet* was then used to statistically test for the number of genomic "clusters" in the dataset (Jombart et al. 2010). Informed by DAPC results, both non-model analyses were re-run separately for sites to the east and west of Eyre Peninsula to enable detection of potential lower levels of differentiation.

Pairwise genomic differentiation (F_{ST}) between sampling locations was then calculated for the neutral dataset in *Arlequin* based on 10,000 permutations. Significance levels for the estimates were corrected using the B-Y correction to reduce biases potentially created by multiple testing (see Benjamini and Yekutieli 2001; Narum 2006). This resulted in an alpha (α) level of 0.0105. An AMOVA was run to determine the significance of the population structure identified by the methods above (i.e. genomic variation explained by the division of 1) Albany, Esperance, St. Francis Island (SFI) and outer Coffin Bay (CBO); 2) inner Coffin Bay (CBI); 3) SG; and 4) GSV). A Mantel test was then run in the R package *vegan* (Oksanen 2011) to test for IBD at the neutral level. This tests for correlation between genomic (linearised F_{ST} calculated as $F_{ST}/(1- F_{ST})$) and geographical distance among sampling locations. Pairwise geographical distances among sampling locations were measured in Google Maps as the closest along-shore distance between localities. This was done as this is the most likely path of travel for coastal bottlenose dolphins moving along the southern Australian coastline, as confirmed by aerial surveys (Bilgmann et al. 2018). *BayesAss3* (Wilson and Rannala 2003) was used to estimate contemporary migration rates (i.e. over the past two generations) among five inferred neutral populations (i.e. 1) Albany, Esperance and SFI, 2) CBO, 3) CBI, 4) SG and 5) GSV). To investigate the potential role of CBO as a connectivity corridor between the Great Australian Bight and CBI, CBO dolphins were considered as a separate population for the purposes of this analysis only. *BayesAss3* has been modified for use with genomic markers,

but to reduce computation time a subset of 5,000 SNPs were randomly selected from the neutral dataset for the analysis. Allele frequency, inbreeding coefficient and migration rate mixing parameters were adjusted to optimise acceptance rates (i.e. between 20 and 60% (0.31, 0.34, 0.38 for each parameter, respectively), as suggested by Rannala (2007)). Chain convergence was assessed by running the program three times with the same parameters of one hundred million iterations and ten million burn-in steps, but a differing number of seeds for each run (500, 1,000 and 1,500, respectively). Congruence and convergence among the three runs was then checked in *Tracer* v.1.7.1 (Rambaut et al. 2018). Results of the three runs were very similar and as such the results of only the run with 1,000 seeds are shown here.

4.5.5 Genotype-Environment Association

A seascape genomics approach utilising GEA analysis was run on the full dataset to identify loci that may be under selection due to ecological variation across the study region. Data for nine oceanographic, environmental and topological variables were selected for use based on hypothesised and previously known associations with population genetic structure of bottlenose dolphins (Natoli et al. 2005; Bilgmann et al. 2007b; Wiszniewski et al. 2010; Zanardo et al. 2018). These variables were SST, salinity, current velocity, chloA, primary productivity, bathymetry, seafloor slope, seafloor rugosity and topographic relief. Averaged annual mean, maximum, minimum and range values were utilised for each variable, wherever possible, resulting in a total of 24 variables. Data were downloaded from a variety of sources as detailed in Appendix D: Table D.ii.2. All ecological data were then standardised using the basic calculation implemented in Microsoft Excel: = standardize (x, mean, standard deviation).

Associations between ecological variables and genomic differentiation in SABD was investigated using RDA implemented in the R package *vegan* (Oksanen 2011). The RDA method is a multivariate approach that takes into account multiple selective factors and was

shown to perform better than other GEA methods (e.g. latent factor mixed models) over a range of demographic scenarios (Forester et al. 2018). This technique has also been found to work well with varying levels of selection and when IBD is present, providing a balance between low false-positive and high true-positive rates across different sampling designs and sample sizes (Forester et al. 2018; reviewed in Grummer et al. 2019). The RDA was conducted at the individual level, utilising the XY coordinates taken for each individual sample. Using XY coordinates in RDA can however, be problematic as pairwise distances between each sample are calculated as straight lines. The complexity of the southern Australian coastline, including large gulfs and embayments, renders this a highly inaccurate measure of spatial distance between samples. To better represent the actual distance among samples, XY coordinates were transformed by first calculating pairwise seascape distances (i.e. not crossing land) in ArcGIS. These were then used by the R package *memgene* (Galpern et al. 2014) to create MEMs. Combined, the MEM axes represent spatial genetic patterns at multiple geographic scales, which were then used as the "space" variable in the RDA (see Galpern et al. 2014 for more information). Although aerial surveys have suggested that SABD travel along the coastline in southern Australia (Bilgmann et al. 2018), transformed seascape distances are likely to provide an adequate representation of spatial differences among these dolphins to be used in the RDA. A forward selection procedure was then used to select the ecological variables explaining a significant (p <0.05) portion of variation in the genomic data and to reduce collinearity among variables. This was followed by VIF analysis to eliminate any residual collinearity. This involved removing individual variables until all those remaining had a conservative VIF of <3, as previously used in other cetacean studies (O'Brien 2007; Christiansen et al. 2013; Goldbogen et al. 2015). Only variables retained after the forward selection and the VIF analyses were included in the final RDA model. Each of the retained variables were then visually inspected for geographic heterogeneity in ArcGIS. Significance of the overall model and each individual explanatory variable were calculated through 1,000 ANOVA permutations. Loci were identified as candidates for being under selection if they had a score greater than three SD from the mean locus scores, which were calculated across all

loci for each of the first three RDA axes (Forester et al. 2016). Correlation of the allele frequency of these candidate SNPs with each of the retained environmental variables was calculated to establish which variable/s each candidate was most associated with.

4.5.6 Adaptive Population Genetic Diversity and Structure

Loci from the full dataset that were identified by the RDA as potential candidates for being under selection were used to form the putatively adaptive dataset (hereafter referred to as the "adaptive dataset"). Molecular diversity measures were calculated in *Arlequin* as per the neutral dataset, including H_E and H_O and %PL. Wright's inbreeding coefficient (F_{IS}) was not calculated for the adaptive dataset as the assumption of neutrality is violated. *Admixture* was again run with default settings, for *K* between one and 16, with the *K* value with the lowest CV error identified as the most supported number of adaptively differentiated populations. As this dataset is putatively under selection, the assumption of HWE made by *Admixture* is thus violated (Funk et al. 2012). This analysis was therefore, run only for comparative purposes, to investigate potential differences between neutral and adaptive population genetic structure. Free from the same assumptions, PCA and DAPC were then run in R, for all localities and then again splitting the sites to the east and west of Eyre Peninsula as informed by the initial DAPC results. Pairwise genomic differentiation (F_{ST}) between sampling locations was calculated for the adaptive dataset in *Arlequin* based on 10,000 permutations and a B-Y corrected significance level of 0.0105. An AMOVA was then run to determine the significance of the putative population structure. A Mantel test was used to test for the presence of IBD at the adaptive level, using the same pairwise distances as calculated for the neutral dataset and the adaptive linearised F_{ST} as a measure of genomic distance.

4.5.7 Functional Enrichment Analysis and Annotation

Flanking sequences 300 bp in length were extracted from the *T. aduncus* genome (as used during filtering) for all 8,081 SNPs, resulting in a 601bp length fragment containing each SNP. *Blastn* was then used to perform a BLAST search from the nucleotide database available through the NCBI (Altschul et al. 1990; Sayers et al. 2019). An e-value of 1E-6 was used. The resulting XML file was then loaded into Blast2GO (Conesa et al. 2005), and all "blasted" loci were then mapped and annotated. An e-value of 1E-3 was used for the annotation. Blast2GO was then used to perform a functional enrichment analysis using a Fisher's exact test with an alpha value of 0.05 to identify GO terms that were potentially over- or under-represented in the putatively adaptive dataset compared to the reference (full) dataset. To further investigate how environmental variables are potentially impacting the adaptive genomic differentiation of SABD, functional annotation for individual candidate loci of particular interest was conducted. Specifically, this included candidates that were moderately to highly correlated (<-0.4 or >0.4) with one or both of the top two environmental variables explaining the most variation in the genomic dataset. A BLAST search against the *T. truncatus* genome (NIST Tur_tru v1 Reference Annotation Release 101) was run in the NCBI web blastn tool. Only candidates with an e-value of <1E-10 and identity of >90% were considered as reliable. Candidate genes were identified within 20 KB of the query sequence (as previously used for SABD; Batley et al. 2019). Gene functions were then investigated using the Swiss-Prot database in UniProtKB (Boutet et al. 2007; The UniProt Consortium 2018).

4.5.8 Additional Tests for Selection

To ensure that there was no residual signature of selection in the neutral dataset which could be associated with variables not tested by the RDA, three additional outlier tests were used to identify and remove any further loci that could be potentially under selection in the SABD genome. Outliers were tested between the four major populations identified by exploratory data analysis with the full dataset (specified above), as well as for each possible pairwise comparison among these four populations. Firstly, *Bayescan* v.2.1 (Foll and Gaggiotti 2008)

was used to implement Bayesian outlier analyses, with 100,000 iterations, a burn-in period of 50,000 and prior odds of ten. Employing the coalescent-based FDIST method (Beaumont and Nichols 1996), *Arlequin* was then run using the hierarchical model (Excoffier et al. 2009) with 100,000 simulations and 100 demes. The number of groups to simulate was set at the number of localities in the comparison, plus one. P-values were adjusted for FDR using the *plyr* R package (Wickham 2011) to control for biases created through multiple testing. *Bayescan* automatically controls for this, returning FDR corrected q-values, rather than p-values. For both *Arlequin* and *Bayescan* analyses, a SNP was considered as a candidate outlier if they had a q-value of <0.1, corresponding to a FDR of 10%. Thirdly, RandomForest, using the *rfPermute* and *randomForest* packages (v.4.6-14), was implemented in R (Breiman 2001; Archer 2016). Missing data are not tolerated by the machine-learning algorithm used and therefore, it was imputed using the *na.roughfix* function before beginning the analysis. The method used by RandomForest is briefly explained in Chapter 2 and was implemented with 125,000 trees and default settings for the proximity and importance parameters. The number of randomly chosen SNPs tested for each split of the tree (mtry) was set to the value that minimised the out-of-bag error rate and computational time (as suggested by Brieuc et al. 2018). As there is no formal method for the selection of candidate SNPs in RandomForest, the importance value distributions were plotted and loci above the upper elbow of the distribution curve were selected as candidates (as per Batley et al. 2018).

4.6 Results

DdRAD sequences for 139 SABD were obtained from biopsy samples collected from the southern coast of Australia (Figure 4.1). Samples were sequenced across three lanes with other samples from Chapters 2 and 3, with a total of 410 million raw reads (average of 2,720,842 reads per individual (standard deviation ± 2,631,273)). A low number of reads was highly correlated with high missing data in the dataset. To mitigate this issue, eight individuals were removed during the filtering process due to having >20% missing data. The final sample size was 131 SABD with an average of 2.9% missing data (standard deviation ± 3.2%). After a series of filtering steps to obtain the highest quality data for further analysis as detailed in Appendix D: Table D.ii.1, 8,104 SNPs were retained. These loci were then aligned to the *T. aduncus* genome, with a 99.65% alignment rate, leaving 8,081 SNPs for analysis (Appendix D: Table D.ii.1).

4.6.1 Neutral Population Genomic Diversity and Structure

The GEA analysis (see below) and three additional outlier detection methods identified a total of 264 candidate SNPs. The RDA identified 241 candidate SNPs while *Arlequin, Bayescan* and RandomForest detected 9, 12 and 19 outliers respectively. There were six outlier SNPs identified in more than one of the three additional outlier detection methods and a further 11 outliers were also detected by RDA. These loci were subsequently removed from the dataset, leaving 7,817 putatively neutral SNPs for downstream analysis. Neutral genomic diversity was relatively high across the study region, but as expected was substantially lower compared to the adaptive dataset (Appendix D: Table D.ii.3). Patterns of neutral genomic diversity did not appear to be consistent with inferred changes in habitat type across the different diversity measures (e.g. embayment vs open-coast populations). Patterns of inbreeding in subpopulations were relatively low in all sampled sites except for NSG dolphins (Appendix D: Table D.ii.3). *Admixture* returned highly supported values between two and five genomic clusters, likely reflecting a hierarchical metapopulation (Figure 4.2Ai1-4). This inference was

also supported by PCA (Figure 4.2Aii). Clear differentiation of GSV from other SABD sites was evident, while SG was shown to be more connected at the neutral level to the sites to the west of Eyre Peninsula (Figure 4.2Ai2 & 3). It appears that there is neutral gene flow creating high levels of admixture between CBI and the Western Australian and SFI dolphins, with CBO acting as a link between the two (Figure 4.2Ai3 & 4). DAPC supported the presence of only two genomic populations, with a split between sites to the east and west of Eyre Peninsula (Figure 4.2Aiii). To investigate this division further, sites to the east and west of Eyre Peninsula were analysed separately using both PCA and DAPC methods. West Eyre Peninsula sites then displayed a clear differentiation between CBI and the four other sites (Appendix D: Figure D.i.2Ai & ii). DAPC and PCA results were however, not congruent for the east Eyre Peninsula dolphins. PCA showed neutral differentiation between the two gulfs (SG and GSV) (Appendix D: Figure D.i.3Ai), while DAPC did not support this division, with only one cluster to the east of Eyre Peninsula (data not shown due to limitations of the method). When forced to split into two genomic clusters, DAPC separated individuals of the two gulfs correctly, although this was not statistically supported (Appendix D: Figure D.i.3Aii). AMOVA supported the presence of four populations (Western Australia/SFI/CBO, CBI, SG, GSV), but the percentage explained among them (4.06%; p <0.001) was only slightly higher than among sample sites within populations (3.44%; p <0.001) (Appendix D: Table D.ii.4A). The non-congruence among methods is likely a reflection of the hierarchical nature of neutral gene flow of SABD across southern Australia, further supported by moderate to high levels of differentiation (F_{ST}) among all pairwise sampling localities (Figure 4.3). Average F_{ST} across the region was moderate at 0.072, with the highest levels of neutral differentiation between Western Australia sites and those from CBI eastwards, a pattern which was not evident in the adaptive dataset (Figure 4.3). IBD was found to be highly significant for the neutral dataset (r = 0.8682, p <0.0001) (Figure 4.4). *BayesAss3* revealed varying levels of contemporary migration across the southern Australian seascape, suggestive of a hierarchical metapopulation structure (Appendix D: Figure D.i.4). Eastward gene flow was on average lower than westward, with particularly high levels of contemporary gene flow estimated out of GSV into SG and from CBI

into the CBO population. The CBO and SG populations had relatively low proportions of non-migrants, reflecting higher gene flow into these areas, while the embayment CBI community had the highest proportion of non-migrants.

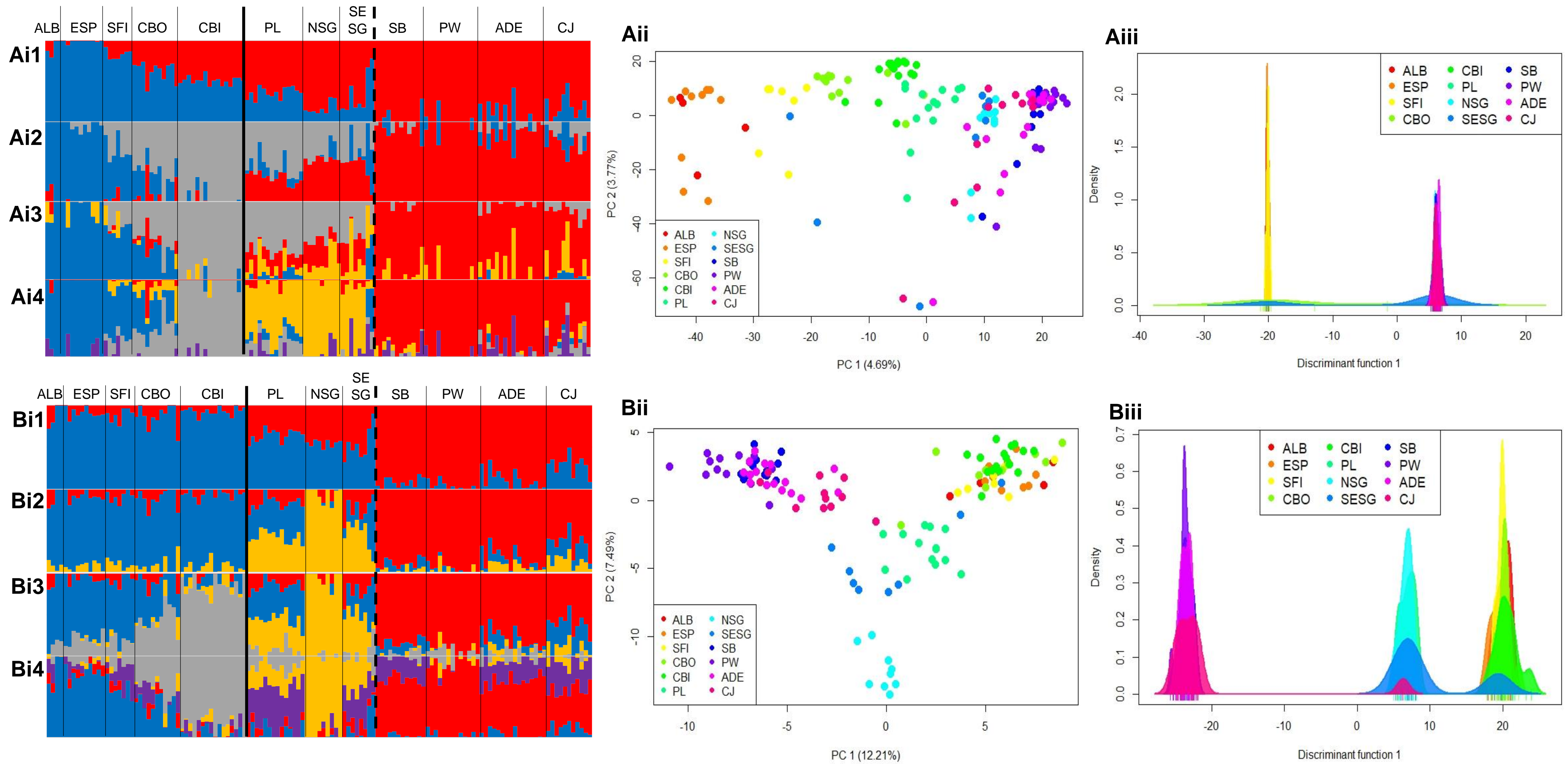


Figure 4.2 Population genomic structure of southern Australian coastal bottlenose dolphins based on 131 individuals and **A)** 7,817 SNPs in the neutral dataset and **B)** 241 outlier SNPs in the adaptive dataset. **i)** *Admixture* plots whereby each column represents an individual dolphin and the proportion of a given colour represents the percentage probability of that dolphin belonging to a given population. Vertical black lines mark the start of a new sampling location, with the thick line representing the split between sites to the east and west of Eyre Peninsula and the dashed line marking the split between Spencer Gulf and Gulf St. Vincent. Population genomic structure is shown at **1)** *K*=2 populations (neutral = highly supported; adaptive = low support); **2)** *K*=3 (neutral = highly supported; adaptive = low support); **3)** *K*=4 (neutral = most supported; adaptive = most supported); and **4)** *K*=5 (neutral = highly supported; adaptive = highly supported). **ii)** Principal component analysis (PCA); and **iii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from west (left/top) to east (right/bottom). Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

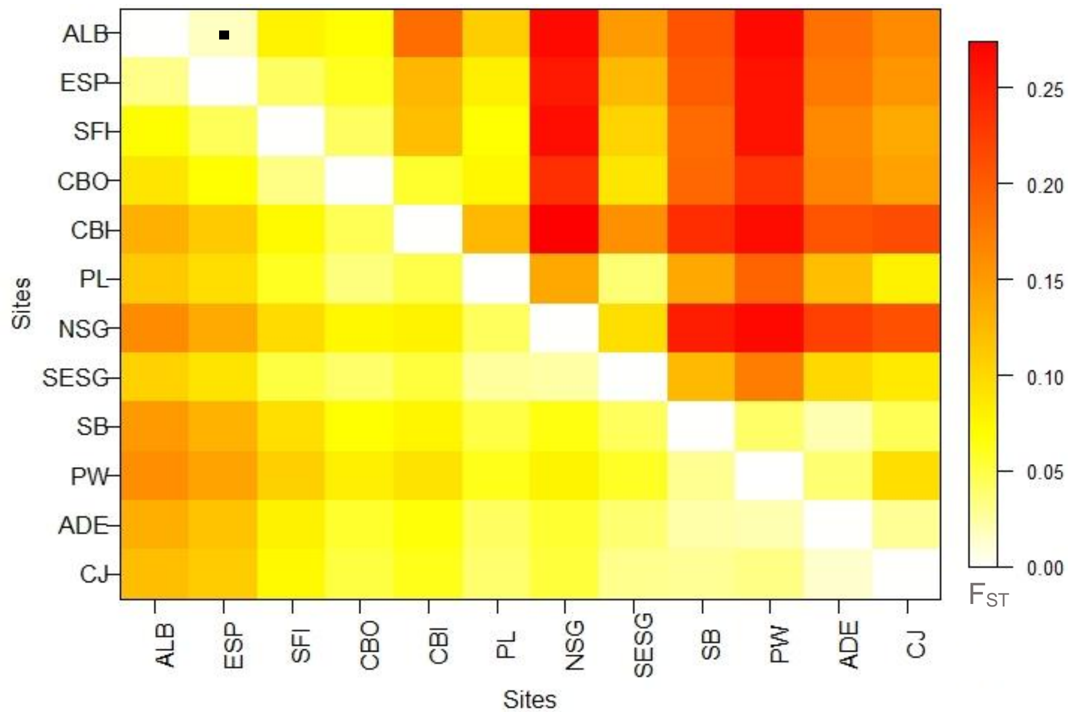


Figure 4.3 Heat map of pairwise genomic differentiation (F_{ST}) between sampling localities of southern Australian coastal bottlenose dolphins based on 131 individuals as estimated by *Arlequin*. F_{ST} estimates based on 241 adaptive SNPs can be found in the top half of the matrix, while estimates based on the 7,817 neutral SNPs are in the bottom half. The black square (▪) denotes the one estimate that was non-significant at the B-Y corrected alpha value 0.0105 across all pairwise comparisons for both datasets. Average F_{ST} for the neutral dataset was 0.072 and for the adaptive dataset was 0.147. Sampling location abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

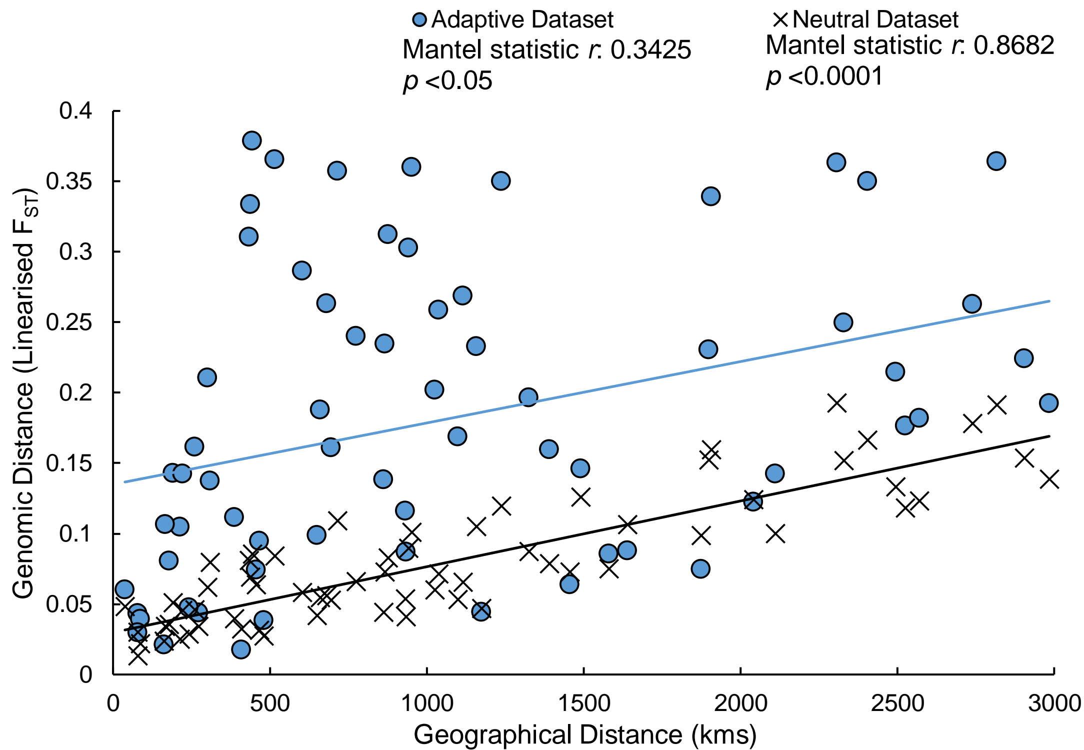
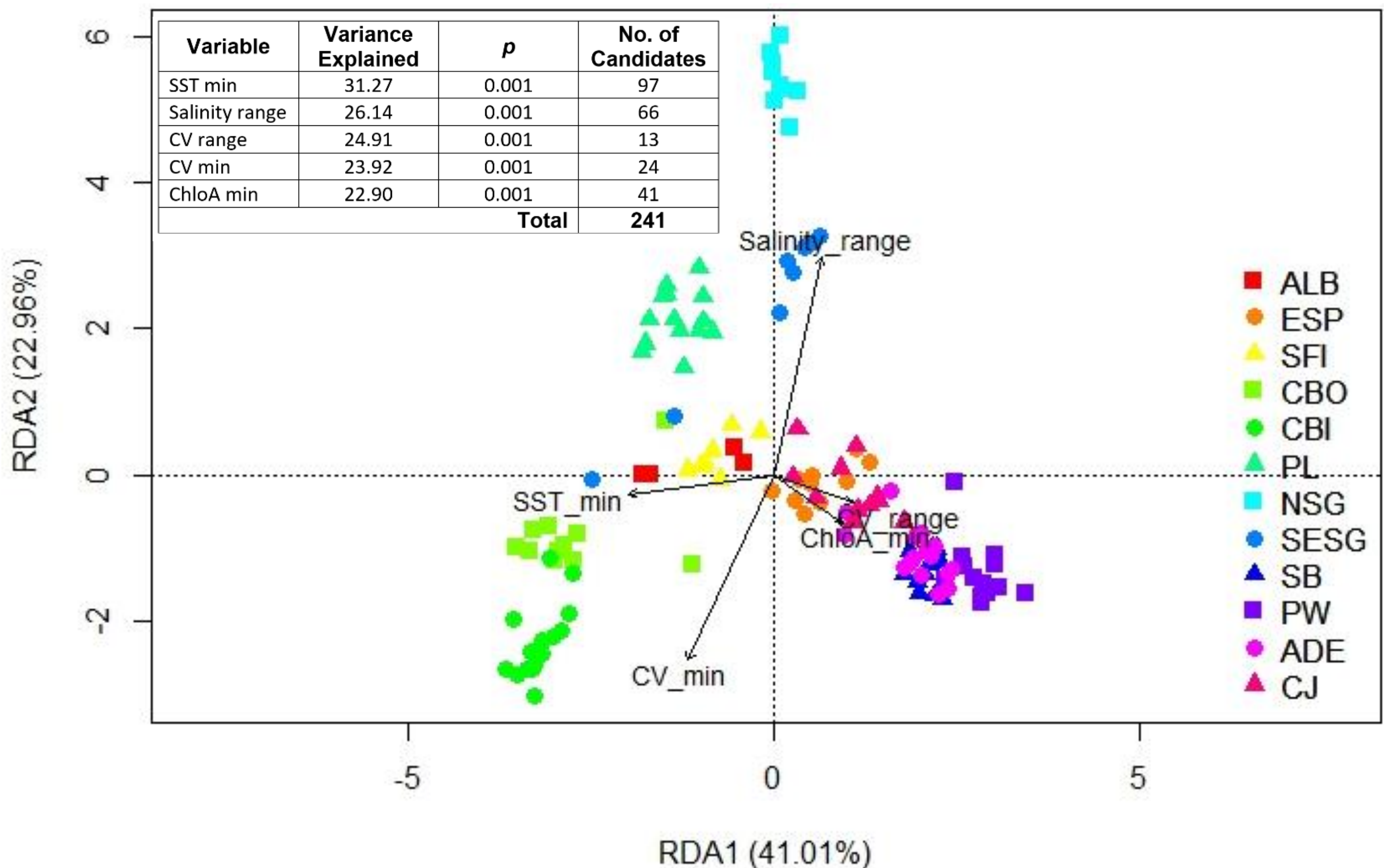


Figure 4.4 Mantel test for isolation by distance (IBD) in southern Australian coastal bottlenose dolphins. Correlation between neutral (crosses) and adaptive (blue circles) genomic distance (linearised F_{ST}) and along-shore geographical distance (kms) among sampling localities is displayed.

4.6.2 Genotype-Environment Association

A total of 24 environmental, topological and oceanographic variables were used in the RDA as part of seascape genomics GEA analysis (Appendix D: Table D.ii.2). Forward selection identified nine variables significantly associated ($p < 0.05$) with genomic variation. Collinearity of these nine variables was tested and variables were removed until those remaining had a VIF of <3 (Appendix D: Figure D.i.5). The final five variables included in the RDA model were minimum SST, salinity range, current velocity range and minimum, and minimum chloA. Stark gradients in these five variables was clear upon visual inspection of the environmental maps in ArcGIS (Appendix D: Figure D.i.1). The overall RDA model was significant ($p = 0.001$), with 5.68% of the genomic variance in the dataset explained by the spatial variables and 9.30% explained by the retained ecological variables (Figure 4.5; Appendix D: Table D.ii.5). All five

ecological parameters were significant at $p = 0.001$, with minimum SST and salinity range explaining most of the variation. The RDA identified 241 loci with scores ±3SD from the mean of at least one of the constrained RDA axes and therefore, potential candidates for selection in response to environmental heterogeneity. On the first RDA axis (41.01% of the constrained variance) the inner and outer CB dolphins are shown to be separated from the other sampling localities based on a potential association with minimum SST and to some extent, minimum current velocity. Genomic divergence of SABD communities in GSV on the other hand, was associated with current velocity range, minimum SST and minimum chloA. On the second axis (22.96% of the constrained variance), salinity range is associated with the genomic divergence of SG dolphins, particularly those in NSG (Figure 4.5).



Variable	Variance Explained	*p*	No. of Candidates
SST min	31.27	0.001	97
Salinity range	26.14	0.001	66
CV range	24.91	0.001	13
CV min	23.92	0.001	24
ChloA min	22.90	0.001	41
		Total	**241**

Figure 4.5 Genotype-environment association (GEA) redundancy analysis (RDA) testing for the association between the five retained ecological variables and individual genomic differentiation coloured by sampling locality. Overall variance in the genomic dataset explained by the model was 186.10 (p = 0.001) (1,702.40 residual variance), with 5.68% explained by space and 9.30% explained by the five ecological variables (see Appendix D: Table D.ii.5 for details). Table inset details the variance explained, significance (p) and number of candidate loci most highly correlated with each of the five retained variables. Additional information of the ecological variables used are provided in Appendix D: Table D.ii.2 (SST = sea surface temperature, CV = current velocity, ChloA = chlorophyll A concentration; also note min = minimum). Sampling location abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3) and are ordered from west (top) to east (bottom) in the legend.

4.6.3 Adaptive Population Genetic Diversity and Structure

The 241 loci identified as putatively under selection by the RDA were extracted to form the adaptive dataset. Molecular diversity at the adaptive level was high throughout the study region, with no obvious association with habitat type or putative neutral population (Appendix D: Table D.ii.3). The highest diversity was recorded in southeastern SG (SESG), while the

lowest level was estimated for Port Wakefield (PW). *Admixture* estimated the presence of four adaptive populations, with a very clear separation of GSV dolphins from all other sampling localities (Figure 4.2Bi1-3). While the entire SG also seems to be differentiated, NSG dolphins appear to be adaptively divergent from the southern SG (PL and SESG) sites, with the southern dolphins having stronger affinity to dolphins further west (Figure 4.2Bi3). These findings were supported by PCA, showing clear adaptive differences between GSV and SG and the more admixed sites to the west of Eyre Peninsula (Figure 4.2Bii). PCA also highlights the clear separation of NSG from the southern SG dolphins and to a lesser extent PW from the other GSV dolphins. The low-level separation of PW from other GSV dolphins was also evident when inspecting $K = 5$ in the *Admixture* graphs (Figure 4.2Bi4) and when only the localities to the east of Eyre Peninsula were run in PCA (Appendix D: Figure D.i.3Bi). While at the neutral level DAPC detected only two genomic populations, at the adaptive level DAPC supported the presence of four adaptively differentiated clusters (Figure 4.2Biii) but did not detect NSG as distinct, until the east Eyre Peninsula sites were analysed separately (Appendix D: Figure D.i.3Bii).

To the west of Eyre Peninsula, the sheltered embayment dolphins inhabiting CBI appears to be segregated from the Western Australian and SFI dolphins, with the CBO individuals acting as a link between the two (Figure 4.2Bi3 & 4). Although DAPC suggested CBI as a separate population, it also indicated a close association between these dolphins and those further west (Figure 4.2Biii). Both PCA and DAPC run with the west Eyre Peninsula sites strongly supported the finding that CBI is adaptively divergent from the other west Eyre Peninsula dolphins (Appendix D: Figure D.i.2Bi & ii). This was further supported by AMOVA, with 12.03% of the variation (p <0.001) explained by population division, compared to 5.29% (p <0.001) among sampling localities within each population (Appendix D: Table D.ii.4B). This was substantially higher than the variance explained among and within populations for the neutral dataset. Differentiation among sampling localities at putatively adaptive DNA regions was also tested using F_{ST}. The average F_{ST} value at the adaptive level was substantially higher than for

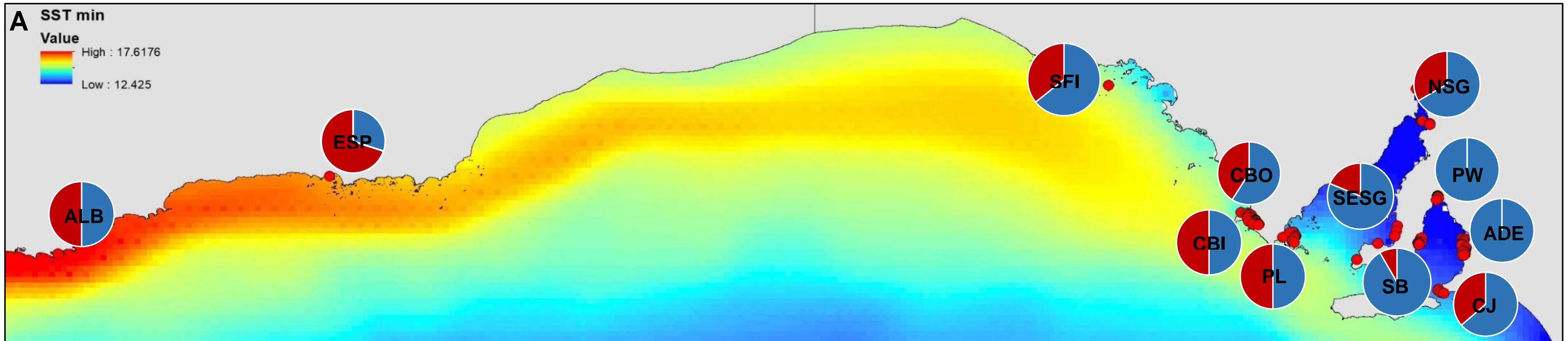
the neutral dataset (0.147 compared to 0.072). Values ranged from the non-significant comparison of Albany and Esperance at 0.017, to a highly significant 0.275 between dolphins from NSG and CBI (Figure 4.3). Genetic and geographical distances were found to be moderately correlated (r = 0.3425, p <0.05), although this was substantially less than that recorded for the neutral dataset (r = 0.8682, above) (Figure 4.4).

4.6.4 Functional Enrichment Analysis and Annotation

Of the 8,081 SNPs, a total of 453 loci (5.6%) scored BLAST hits and were mapped and annotated, of which 3.3% were GEA candidates. Enrichment analysis found 215 GO terms significantly (p <0.05) over-represented in the putative adaptive dataset compared to the full dataset (no terms were under-represented) (Appendix D: Table D.ii.6). Notable GO terms that were significantly over-enriched included temperature homeostasis (GO:0001659), adaptive thermogenesis (GO:1990845), cellular response to carbohydrate stimulus (GO:0071322), positive regulation of muscle organ development (GO:0048636), ion gated channel activity (GO:0022839), ion transmembrane transport activity (GO:0015075), positive regulation of cell proliferation involved in kidney development (GO:1901724), as well as several other associated GO terms (significance values are in Appendix D: Table D.ii.6).

To further investigate individual gene functions, a total of 82 candidate loci that were moderately to highly correlated with salinity range and/or minimum SST (30 and 54, respectively, with two overlapping) were analysed with BLAST. Of these, 38 candidate genes were identified within 20KB of a candidate SNP (Appendix D: Table D.ii.7). Of particular interest was one candidate locus identified in an exonic region of the *CMKLR1* gene, which was negatively correlated with SST minimum. Allele frequency plotted across the southern Australian coastline revealed an increase in homozygosity and frequency of the major allele closely corresponding to decreasing SST moving east (Figure 4.6A). The major allele reaches near-fixation in GSV, with 94% of inner-GSV dolphins being homozygotes for this variant, compared to just 20% of dolphins to the west of Eyre Peninsula (Figure 4.6B).

On the other hand, several candidate loci that were highly correlated with salinity range were associated with genes involved in ion transport. This included the genes *KCNT2* and *SLC22A18*, as well as in an exonic region of *RYR2*. Variation in the allele frequencies of these genes in the dolphins across the southern Australian coast shows a sharp increase in the frequency of the minor allele in SG, particularly NSG (Appendix D: Figure D.i.6Ai, Bi and Ci). For all three genes, minor allele homozygotes are only present in SG and are in particularly high frequency in NSG (Appendix D: Figure D.i.6Aii, Bii and Cii).



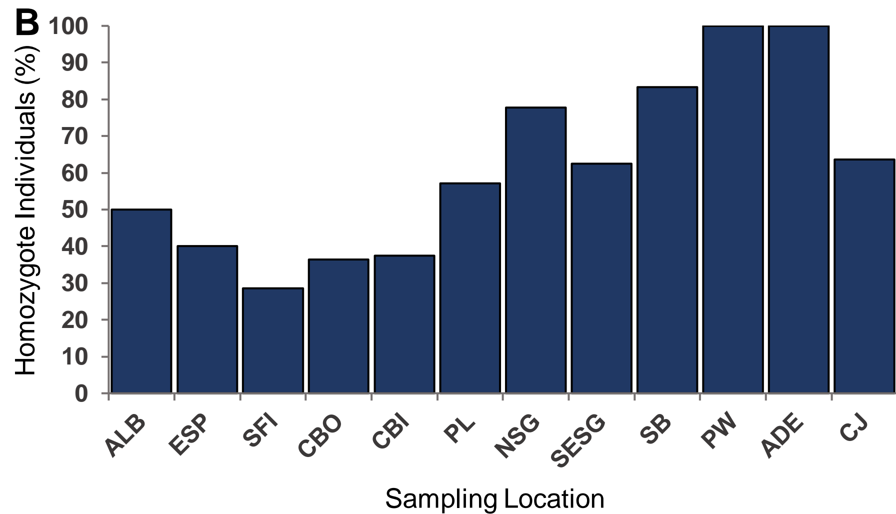


Figure 4.6 Variation in candidate gene *CMKLR1*, showing **A)** allele frequency change over the strong minimum sea surface temperature (SST) gradient across the southern Australian seascape. Blue and red pie sections correspond to the frequency of occurrence at each sampling location of the major and minor alleles, respectively; and **B)** the percentage of southern Australian coastal bottlenose dolphins at each sampling locality found to be homozygotes at either allele. Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

4.7 Discussion

Anthropogenic climate change is affecting all ecosystems on earth, with a complex interplay of numerous changes in the marine ecosystem. This includes an increase in the frequency of climate extremes, such as marine heatwaves, and an overall trend toward warmer, more acidic oceans (Poloczanska et al. 2013). While it is generally expected that upper trophic species may be mostly affected by climate change through indirect impacts on the food web and habitats (e.g. Bakun et al. 2015; Sydeman et al. 2015; Scheffers et al. 2016), adaptation of these species to environmental conditions at the genomic level is relatively understudied. Application of the landscape genomics framework into marine systems now allows for the relationship between environmental features and genomic differentiation and adaptation to be formally tested. Associations between genomic differentiation of SABD and the distribution of environmental gradients along southern Australia were investigated. Adaptive and neutral datasets showed different patterns of population genomic structure and molecular diversity suggestive of local adaptation across the region. The hypothesis of fine-scale neutral population structure in bottlenose dolphins associated with IBD and social structuring along coastal southern Australia was supported. Adaptive divergence on the other hand, was found to be in association with environmental differentiation and habitat type changes as hypothesised. A number of candidate genes moderately/highly correlated with variation in salinity range and/or minimum SST were identified. These appear to be involved in cellular ion transport and adipogenesis, respectively. The evidence for spatial adaptive divergence and likely physiological adaptations associated with environmental heterogeneity contrasts with previous suggestions that megafauna species are likely to be mainly indirectly affected by environmental and climatic change through impacts on prey species (e.g. Bilgmann et al. 2007b; Möller et al. 2011; Bakun et al. 2015; Sydeman et al. 2015).

4.7.1 Genomic Variation

Extensive variation in population genomic diversity within species is thought to be impacted by a number of factors, including the demographic history of a particular species (Romiguier et al. 2014; Ellegren and Galtier 2016). Demographic fluctuations are typically due to environmental and ecological disturbances, including climate change and anthropogenic influences, or variations in N_e resulting from a strong bottleneck and/or a founder event (Bjørnstad and Grenfell 2001; Ellegren and Galtier 2016). Population demographics, including the effect of potential founder events, social structure and natal philopatry have been suggested to be key drivers of lower genomic variation in coastal compared to pelagic bottlenose dolphin populations worldwide (Möller et al. 2007; Möller 2012; Louis et al. 2014b; Bayas-Rea et al. 2018). Here however, differences in population genomic diversity in SABD were not associated with broad-scale changes in habitat type (e.g. embayment/gulf versus open-coast) for either neutral or adaptive datasets. It is also worth noting, that while there is a considerable density of SABD over the small NSG region (Bilgmann et al. 2019), field observations suggest that the NSG dolphin community is small and highly philopatric (Bilgmann and Möller, personal comm.). Due to the small size of this local community, the relatively low number of samples secured for the NSG provides a reasonable representation of genomic diversity in dolphins inhabiting the region. This may be particularly true given the large number of independent markers used here (e.g. see Gaughran et al. 2018).

While the neutral and adaptive datasets showed a similar pattern of genomic variation across the study region, neutral diversity was lower than putative adaptive diversity for each sampling locality. This suggests a potential influence of balancing selection, whereby allelic diversity in genes affecting fitness is actively retained (Charlesworth 2006; Hedrick 2007). Directional selection on the other hand, typically leads to a reduction in diversity through the process of selective sweeps (Hedrick 2007). Alternatively, higher diversity levels in the adaptive dataset may be as a result of high natal philopatry and local adaptation through directional selection

on the SABD genome, combined with occasional migration between populations. Further studies are required to determine the extent to which both directional and balancing selection are impacting these dolphins. As mentioned in Chapter 3, distinguishing between different modes of selection and identifying their relative strengths is hampered by the lack of information about adaptive phenotypes in inshore bottlenose dolphins and how such phenotypes interact with varying environments.

4.7.2 Environmental Drivers of Genomic Differentiation

It is well established that changes in the environment and the opening of new niche space is closely associated with periods of rapid species radiation (Stroud and Losos 2016). In the marine realm, periods of ocean warming and cooling throughout the Cenozoic greatly impacted on the strength and position of upwelling systems, having major implications for the diversification of many marine species (Norris et al. 2013). While it is well documented how environmental changes have influenced species radiations over evolutionary time, diversification of marine species at the population level to contemporary environmental changes has only recently started to be investigated. Evidence is provided here supporting a strong influence of environmental gradients and changes in broad-scale habitat type in shaping population genomic structure in a highly mobile near-top predator, the SABD. Population differentiation in small cetaceans has often been linked to local adaptation to specific prey types, which is then reinforced by social structure and philopatry (Hoelzel 2009; Tezanos-Pinto et al. 2009; Möller 2012). This is particularly common for bottlenose dolphin populations inhabiting embayments and other sheltered coastal habitats (Curry and Smith 1997; Wiszniewski et al. 2010; Fruet et al. 2014b; Louis et al. 2014a). Here, the clear distinction identified among gulf, embayment and open-coast communities in southern Australia indicates that broad-scale habitat type changes impact SABD population genomic structure. With the use of a high-quality genomic dataset of over 8,000 SNPs, we can now move past habitat type related inferences of population structure and empirically test for

influences of specific environmental and oceanographic variables on the bottlenose dolphin genome. The RDA revealed a greater proportion of the genomic variation being explained by the environment (9.3%) compared to space (5.7%). This suggests that selection in response to environmental and oceanographic gradients is having a strong impact on SABD adaptive differentiation. While still significant, the strength of IBD was considerably lower for the adaptive dataset than the neutral, further supporting the above inference.

A potential soft barrier to the dispersal of SABD in the waters off Eyre Peninsula was identified, which was previously suggested for this species based on microsatellite markers (Bilgmann et al. 2007b; Pratt et al. 2018). This was thought to be related to the presence of a strong upwelling system south of Eyre Peninsula, which heavily influences fish distribution in the region (Dimmlich et al. 2004; Kämpf et al. 2004). The influence of this barrier could however, be somewhat confounded by the clear separation of the two gulfs from open-coast sites to the west in terms of both environmental conditions and SABD population genomic structure. Nevertheless, differences in the environmental conditions either side of Eyre Peninsula appear to be influencing patterns of genomic differentiation in several other species of marine organisms. This includes common dolphins (*Delphinus delphis;* Bilgmann et al. 2014), Australian sea lions (*Neophoca cinerea*; Lowther et al. 2012), mulloway (*Argyrosomus japonicus;* Barnes et al. 2015), and potentially the marine clam, *Lasaea australis* (Li et al. 2013). The diversity of these examples points to the likely role of upwelling systems in influencing population genomic divergence of entire marine communities (also see Kelly and Palumbi et al. 2010). Due to the contrasting patterns of differentiation on either side of Eyre Peninsula, these regions are hereafter discussed separately.

4.7.2.1 Coffin Bay and the Great Australian Bight

The Great Australian Bight is part of the world's longest southern facing coastline. It is recognised as of global conservation significance due to high levels of biodiversity and endemism, and the presence of critical habitats and migratory pathways for many keystone

species (Baghurst et al. 2017). This region appears to be highly important to SABD gene flow and metapopulation connectivity. With a strong signal of IBD and high contemporary migration estimates to the west of Eyre Peninsula, the CBO dolphins are likely acting as a link between CBI and the Great Australian Bight region (including Albany, Esperance and SFI communities). The CBO dolphins appear to have a more transient lifestyle than is typical of most inshore bottlenose dolphins (i.e. Wiszniewski et al. 2010; Fruet et al. 2014), and the CBI dolphins in particular (Passadore et al. 2018a). This is supported by lower density and encounter rates of SABD in this area than in CBI (Passadore et al. 2018a), and a much lower proportion of non-migrants found in this study. This may be driven by instability in productivity and resources and the unprotected nature of CBO (Passadore et al. 2018b), forcing these dolphins to adopt transient behaviours more typical of open-coast or pelagic bottlenose dolphins (see Wells et al. 1999; Möller 2012). Indeed, the CBO population may be facilitating a stepping-stone pattern of gene flow across approximately 1,800kms, between the western side of the Great Australian Bight and CBI, potentially supported by the Leeuwin and coastal Currents (Dimmlich et al. 2004). This region may also harbour potential unsampled SABD populations that aid in connecting dolphins across the Great Australian Bight. Due to difficult oceanic conditions across the Bight and its remoteness, further sampling in the region was not possible. A similar pattern of large-scale dispersal was found in the broadcast spawning snail, *Nerita atramentosa*, with coastal currents likely affecting population connectivity across the Great Australian Bight, and genetic structure reflecting a pattern of IBD across the broader southern Australian coast (Teske et al. 2015). Contemporary migration of SABD was however, higher moving westward than eastward, with movement in both directions across the Great Australian Bight perhaps facilitated by seasonal change in the strength and influence of the Leeuwin Current (Feng et al. 2009). This finding therefore, challenges our current knowledge about range patterns and gene flow in inshore dolphins and demonstrates the broader impact of habitat and environmental features in shaping genomic differentiation.

Environmental conditions within the CB embayment are vastly different to those seen across the Great Australian Bight. The inner bays are protected from the open ocean by a long spit of land (Point Longnose), creating a semi-enclosed inverse estuary with restricted water exchange (Kämpf and Ellis 2015). Different habitat types and associated variation in fish assemblages has been suggested to be driving social division of SABD among the interconnected bays in CB (Passadore et al. 2018a; Diaz-Aguirre et al. 2019; Whitmarsh et al. in review). While there was no evidence of genomic subdivision within CBI, the influence of habitat and resource specialisation on these dolphins has likely resulted in strong philopatry and social structure. It may also be having an impact on the genomic separation observed between CBI and the more transient CBO dolphins, as well as the wider SABD metapopulation. GEA analysis did however, indicate that differentiation between these two communities could also be related to minimum current velocity and potentially to salinity variation. Indeed, due to the shallow depths throughout CBI (mean = ~2.6m, max = 5m), the embayment has strong variation in salinity. Evaporation during summer creates hypersaline conditions up to 50psu, while freshwater input over winter causes the psu to drop to ~36.5 (Kämpf and Ellis 2015). The presence of Point Longnose significantly affects current velocity in CBI. As a result, current velocity is much lower than in CBO and connectivity with the shelf waters is restricted to the upwelling season (Kämpf and Ellis 2015). This suggests that dolphins residing in CBI could be locally adapted to the variation in salinity and low current velocity, causing them to adaptively diverge from the open-coast SABD populations. Genetic differentiation associated with semi-enclosed estuaries is found in several species of small cetacean, including common bottlenose dolphins (*T. truncatus*; Fruet et al. 2014), Indo-Pacific finless porpoises (*Neophocaena phocaenoides*; Jia et al. 2014) and franciscana dolphins (*Pontoporia blainvillei*; Costa-Urrutia et al. 2012). Estuaries are therefore, likely to be an important habitat for the differentiation of cetaceans, potentially through the provision of niche space that is underutilised by other apex predators. This is also expected to apply to the two South Australian gulfs, which are both classified as large inverse estuaries (IMCRA Technical Group 1998).

4.7.2.2 Gulf Localities

The two South Australian gulfs provide important habitat for SABD, with significantly higher densities and overall abundance within the two gulfs than in outer-gulf and shelf waters in South Australia (Bilgmann et al. 2019). Genomic differentiation was identified between gulf and open-coast SABD dolphins, consistent with previous findings based on microsatellites (Bilgmann et al. 2007b; Pratt et al. 2018). Isolation of the SG and GSV populations from the outer-gulf dolphins was highlighted by contemporary migration estimates being high between the two gulfs, but negligible to low between gulf and outer-gulf populations. In part, this appears to be related to a strong gradient in minimum SST across the southern Australian coastline, with considerably higher minimum SST in the west than in the gulfs to the east. This is likely as a result of the Leeuwin Current drawing warm tropical waters down the west coast and into the temperate waters of southern Australia (Rochford 1986). The formation of summer/autumn thermal and saline fronts, and the presence of several islands at the entrances to both SG and GSV, effectively shelter the gulfs and separate gulf and shelf waters (Petrusevics 1993; Harvey 2006; O'Connell et al. 2016). These oceanographic discontinuities have already been suggested to be responsible for species-level distributional separation between the coastal SABD (referred to as *T. aduncus*) and the offshore-type *T. truncatus* (common bottlenose dolphins) (Gibbs and Kemper 2014). The influence on bottlenose dolphin gene flow could be occurring indirectly through the impact on prey populations and/or may be having a direct impact on the local adaptation of the dolphins to their respective environments.

GEA analysis identified several loci in the SABD genome as candidates for being under selection in relation to the gradient in minimum SST. Subsequently, several GO terms related to thermogenesis, and in particular cold-induced thermogenesis, were found to be significantly over-enriched in the putatively adaptive dataset compared to the full dataset. One candidate locus of particular interest was found in an exonic region of a homolog of the *CMKLR1* gene.

The stark increase in the frequency of homozygotes and occurrence of the major allele moving east across southern Australia suggests that directional selection could be acting on this gene, bringing the major allele closer to fixation in the cooler, eastern parts of the study area (see Hedrick 2007). *CMKLR1* is closely associated with the process of adipogenesis (Muruganandan et al. 2010; Ernst et al. 2012; Audet et al. 2016). Specifically, this gene is required for adipocyte (fat cell) differentiation from bone marrow precursor cells (Muruganandan et al. 2010). An inhibition of *CMKLR1* supresses appetite and reduces weight gain and fat storage in mice (Ernst et al. 2012). Changes in the expression of this gene in the hypothalamus of mice also affects individual core temperature and susceptibility to hypothermia (Audet et al. 2016). It is hypothesised that *CMKLR1,* and potentially several other genes, vary in SABD in response to the minimum SST gradient in southern Australia creating differing fat storage and thermogenic requirements in the dolphins of different areas. For example, dolphins to the west of the study region are subject to warmer minimum temperatures and thus would require less fat storage compared to those residing in the two South Australian gulfs, which experience the lowest minimum temperatures. Adipogenesis and volume of the resulting brown adipose tissue historically differ among human populations living in different climate zones, likely a result of genomic adaptation to differing thermogenic requirements between the climates (Sazzini et al. 2014). Genes associated with adipogenesis have also been found to differ between brown bears and polar bears and between killer whale (*Orcinus orca*) ecotypes residing in regions characterised by starkly different temperature profiles (Liu et al. 2014; Foote et al. 2016). These studies not only linked the variation in adipogenesis-related genes to SST, but also to differing diets. This has also been explored recently at the macroevolutionary scale, with lipid and glucose metabolism pathways undergoing major adaptation in the transition of cetaceans into the aquatic environment (Nery et al. 2013b; Derous et al. 2019). This may be linked to a subsequent change in diet to one high in fat and protein, and further influenced by energy requirements during diving (Derous et al. 2019). The present study also found over-enrichment in GO terms associated with the cellular response to carbohydrate, monosaccharide, hexose and glucose stimuli. It could thus

be speculated that in southern Australia, variation in the *CMKLR1* gene and others in SABD may be associated with both SST and perhaps a change in prey choice across the seascape. Validating studies are needed to evaluate this further. It is however, clear that selection on genes related to adipogenesis is potentially a crucial factor in the adaptation of mammals to climatic temperature differences and possibly the associated differences in diet. Variation in genes associated with adipogenesis, including *CMKLR1,* may become increasingly important in the adaptation of species to warmer conditions predicted under current human-induced climate change models.

While broad-scale temperature gradients across the southern Australian coastline may be causing SABD differentiation of the gulfs from open-coast populations, salinity is also likely to be having an impact, particularly on within-gulf adaptive differentiation.

4.7.2.2.1 Spencer Gulf

Isolation of SG from shelf waters has resulted in a very unique ecosystem, with further subdivision of both abiotic and biotic aspects also present within the gulf. Adaptive genomic differentiation between dolphins in NSG and those in southern SG (PL and SESG) was detected. Pratt et al. (2018) proposed a historic basis to this differentiation, finding that three of the four mtDNA haplotypes in NSG dolphins were unique to the area. These two communities reside in different bioregions (IMCRA Technical Group 1998) characterised by distinct flushing regimes that physically separate the northern and southern gulf waters (Bullock 1975; Kämpf et al. 2010). NSG is much warmer than southern SG (Petrusevics 1993) and while the gulf in general is subject to high salinity, the northern reaches are particularly hypersaline (past estimates show winter salinities of >42psu in NSG compared to <36psu in southern SG) (Vaz et al. 1990; Kämpf et al. 2009; Kämpf et al. 2010). This is due to high levels of evaporation in the shallow NSG waters and little freshwater input (Bullock 1975). GEA analysis suggested that SG communities and NSG in particular, were differentiated from other SABD based on an association with salinity. It should be noted that while salinity range was

chosen for inclusion and was significant in the final model, this variable was highly correlated with the excluded minimum, maximum and mean salinity variables and therefore, the effects of each individual parameter cannot be dismissed. The general salinity gradient across the seascape will thus be referred to, rather than focusing on the effect of salinity range specifically. The environmental conditions within SG have substantial impacts on the local biota (Currie and Sorokin 2010). Specifically, population division in giant cuttlefish (*Sepia apama*; Gillanders et al. 2016), and western king prawns (*Penaeus latisulcatus*; Roberts et al. 2012) has been documented between northern and southern regions. Accordingly, Gibbs et al. (2011) revealed that major differences existed in the diets of SABD in NSG and southern SG. In particular, giant cuttlefish and crustaceans (i.e. prawns) were important prey items for SABD in NSG but were largely absent from the diet of southern SG dolphins. Furthermore, the NSG dolphins have developed a unique strategy for the manipulation of cuttlefish prey prior to eating, with suggestions of cultural transmission and/or learning of this technique among members of the population (Finn et al. 2009). Prey specialisation has been suggested to drive patterns of population genetic structure in bottlenose dolphins through its association with natal philopatry and sociality (Wiszniewski et al. 2010; Möller 2012; Chapter 3). Evidence to support this hypothesis is provided here, demonstrating a potential connection between adaptive differentiation and prey handling techniques and associated diet. This has implications for other species that engage in social learning (e.g. killer whales), suggesting that prey specialisation may have a direct impact on population divergence and genomic adaptation. Many marine mammal species are currently experiencing declining prey abundances potentially forcing them to diversify their diet in the coming years, as has already been seen in polar bears (Rode et al. 2015). Overfishing and human-induced climate change affecting prey stocks may therefore, have direct consequences for the adaptation and population structure of marine mammals.

While prey specialisation is likely influencing the genomic differentiation of NSG dolphins, the hypersaline conditions may also be driving direct physiological adaptations. Functional enrichment analysis revealed that GO terms associated with kidney development and ion channel activity were significantly enriched in the putative adaptive loci. GEA analysis also uncovered moderate to high correlations between salinity and a number of candidate loci associated with genes involved in ion transport. This includes the *RYR2* gene, which codes for a cardiac ryanodine receptor and was also found to be potentially involved in the early-stage evolution of the inshore bottlenose dolphin ecotype in the SWAO (Chapter 2). This gene has a crucial role in the regulation of heart beat rhythm and is affected by intracellular sodium levels (Toischer et al. 2013). In addition, increased consumption of omega-3 fatty acids can have inhibitory effects on *RYR2*, reducing the rate of heart failure in humans (Ismail 2005). Major differences in the frequency of occurrence of several fish species with high levels of fatty acids (e.g. mackerel, tuna, herring and sardines) was observed in the stomach contents of NSG and southern SG dolphins (Gibbs et al. 2011). The variation in an exonic region of this gene across the southern Australian coast may therefore, be a response to both the strong salinity gradient and prey choice differences in relation to healthy heart function. Prey choice is typically thought to be associated with population genomic differentiation of cetaceans through its close relationship with sociality. This finding however, suggests that there may also be a role for prey choice in the physiological adaptations of these animals.

Allele frequency changes in *KCNT2* and *SLC22A18* on the other hand, may be related to both salinity and SST variation. *KCNT2* codes for a sodium-dependent potassium channel (Tomasello 2017). These channels are normally activated by sodium elevation (Thomson et al. 2015), but in the absence of a working *KCNT2* channel (which can be caused by low intracellular sodium concentrations), mice are more susceptible to hypothermia and pain responses (Tomasello 2017). *SLC22A18* however, which encodes the solute carrier family 22 member 18 protein, controls transport of compounds in the kidney (Reece et al. 1998) and

regulates fat accumulation, potentially relating to thermoregulation (Yamamoto et al. 2013). With further SST and salinity changes predicted under human-induced climate change, variations in homologs of these genes may therefore, be important in the adaptation of marine species into the future. Previous studies of macro- and microevolution in cetaceans have documented members of the *KCN* gene family (potassium channels) to be important in a wide range of processes, including in cardiovascular development and functioning, and response to hypoxia (Chapter 2; McGowen et al. 2012; Zhou et al. 2015; Zhou et al. 2018b; Huelsmann et al. 2019). Members of the *SLC* (solute carrier) gene family have also been previously found to be critical in the evolutionary adaptations of marine mammals, and cetaceans specifically, to the aquatic lifestyle (Nery et al. 2013b; Zhou et al. 2013; Yim et al. 2014; Foote et al. 2015; Huelsmann et al. 2019). Most notably, Zhou et al. (2018) found that a member of this family (*SLC14A2*) was involved in adaptations related to renal water homeostasis and urea transport between marine and freshwater finless porpoises (*Neophocaena* spp.).

For *KCNT2, SLC22A18* and *RYR2,* the minor allele substantially increases in frequency in SG. Minor allele homozygotes for each of these genes are only present in this gulf and in particularly high frequency in NSG. For each gene, homozygosity at the minor allele was however, still low, indicating that selection on this variant of the genes has perhaps not yet had time to reach fixation. Although physiological adaptation of cetaceans to salinity gradients is well documented at the macroevolutionary scale (Ortiz 2001; Yim et al. 2014), very few studies have investigated the genomic basis of this adaptation at a species or population level. However, Ruan et al. (2015) revealed that selection on genes associated with ion transport pathways and kidney development and functioning have been an important factor in the evolution of freshwater adaptation in finless porpoises (*N. asiaeorientalis* spp.). Selection on genes involved in these pathways have also been implicated in the development of freshwater and marine populations of rainwater killifish (*Lucania parva*; Kozak et al. 2014), and alewife (*Alosa pseudoharengus*; Czesny et al. 2012), and in differences in freshwater tolerance levels in blue mussels (*Mytilus* sp.; Lockwood and Somero 2011). With several similar GO terms

found to be over-enriched in the candidate genes found here, genomic changes affecting ion transport and kidney development, among many others, are likely to be vital to osmoregulatory adaptation in not just cetaceans, but potentially many marine species.

4.7.2.2.2 Gulf St. Vincent

GSV dolphins are differentiated from other SABD at both putatively adaptive and neutral loci. At the neutral level, genomic differentiation of GSV from other SABD may reflect its colonisation history. This gulf was likely founded by a single maternal lineage (Pratt et al. 2018) as the gulf flooded relatively rapidly around 7,000 years ago (Harvey 2006). This dolphin community was likely isolated and became quickly differentiated from other dolphins, which is supported by the generally low genomic diversity detected at neutral loci in GSV. A high level of contemporary gene flow out of (and a moderate level into) GSV may be facilitated through the slightly more admixed outer-GSV community of Cape Jervis (CJ), diluting the signal of founder events and protecting the population against inbreeding. The stronger signal of gene flow out of GSV as opposed to into the gulf, coupled with adaptation to its local environmental conditions however, appears to reinforce the strong division between GSV dolphins and other SABD. GEA analysis suggested GSV dolphins to be differentiated from other SABD based on an association with minimum SST, minimum chloA and variation in current velocity. These factors are largely influenced by the presence of Kangaroo Island in the GSV mouth, affecting both the oceanography and geomorphology of this gulf (Edyvane 2008). Thermal and saline fronts effectively separate gulf and shelf waters during the summer months, with water temperatures in the GSV entrance being several degrees warmer than on the adjacent shelf (Petrusevics 1993). This pattern is reversed in winter, with cooler waters found within GSV than on the shelf (Petrusevics 1993). These fronts also reduce the inflow of highly productive waters created by summer upwelling off southwestern Kangaroo Island (Kämpf et al. 2004). Circulation within GSV is also heavily dependent on the local temperature gradient and thus changes seasonally with the heating and cooling of gulf waters (de Silva Samarasinghe et al.

2003). The presence of Kangaroo Island and its impacts on GSV oceanographic conditions, particularly local SST, could therefore, be having a direct influence on the local adaptation and genomic differentiation of SABD. Indeed, homozygosity in the temperature-associated candidate gene, *CMKLR1*, increases as one moves east across the study region, with 94% of the inner-GSV dolphins being homozygotes for the major allele. Cooler minimum temperatures in this region than elsewhere in southern Australia, particularly compared to shelf waters, are thus likely to be driving strong directional selection for this genotype in GSV dolphins. This further highlights the potential importance of this gene and others associated with adipogenesis in adapting to SST changes in marine species. While GSV dolphins are collectively differentiated from other SABD, there is a high level of gene flow among dolphin communities within the gulf. This is in stark contrast to SG and may be facilitated by the relatively homogenous GSV seascape, with only one bioregion recognised in this gulf (IMCRA Technical Group 1998). There is however, some variation present. The upper reaches of the gulf are slightly warmer and more saline than the southern region (de Silva Samarasinghe 1998), and there is a mosaic of different habitat types throughout (IMCRA Technical Group 1998). SABD communities in GSV show preference for different habitat types (Cribb et al. 2013), which is linked to fine-scale social division (Zanardo et al. 2018). With different fish assemblages documented for each habitat type within GSV (Whitmarsh 2018; Whitmarsh et al. in review), it is likely that prey specialisation is closely associated with strong social structure, as found for other Australian coastal bottlenose dolphins (Wiszniewski et al. 2009; Diaz-Aguirre et al. 2019). Social structure in turn can influence population genomic structure and specialisation (Hoelzel 2009; Möller 2012), and may be having a role in driving the low-level, putative adaptive divergence of PW dolphins suggested by the SNP dataset. The level of differentiation found here is a testament to the power of genome-wide SNPs to detect fine-scale adaptive population genomic differentiation. With a narrow environmental gradient potentially impacting on the adaptive differentiation of these dolphins, it also highlights that even small-scale alterations to environmental features may have impacts for the adaptation of

cetaceans and potentially other marine species. This could drive further population divergence and fragmentation within marine metapopulations.

4.7.3 Conservation Implications

Despite evidence indicating negligible gene flow between SABD and Indo-Pacific bottlenose dolphins (*T. aduncus*) (Möller et al. 2008; Charlton-Robb et al. 2011; Gray et al. 2018; Chapter 2), *T.* cf. *australis* is as of yet not formally recognised as a separate species or subspecies (Committee on Taxonomy of the Society for Marine Mammalogy 2019). It is thus very important to improve our understanding of this putative taxon to ensure that management of these dolphins is specific to its lineage and regional conditions. This study provides evidence for the presence of a SABD hierarchical metapopulation in southern Australia, with differing levels of subdivision among populations both at neutral and adaptive levels. Specifically, it is clear that SG and GSV dolphins are distinct from each other and from dolphins to the west of Eyre Peninsula. In the Great Australian Bight region on the other hand, gene flow over a much larger spatial scale is evident, facilitating connectivity with the isolated CBI dolphins. Potential soft barriers to the dispersal of SABD exist in the waters around Eyre Peninsula, at the narrow entrance to inner CB and through the summer formation of thermal and saline fronts at the SG and GSV entrances. Adaptive differentiation potentially driven by a strong salinity gradient, was discovered between NSG and the southern SG dolphins. We can therefore, not only define management units based on neutral differentiation as done in the past, but also based on adaptive differences between social communities within populations. This is an important step forward in elucidating cryptic population differentiation that could be crucial to the adaptive capacity of the species over the long-term (Hoelzel et al. 2019). It is therefore, recommended that management strategies recognise a) the strong genomic differentiation of GSV dolphins, taking into account a potentially emerging genomic divergence within this gulf (i.e. between PW and other GSV dolphins); b) the genomic separation of the SG population and clear adaptive divergence of NSG dolphins; c) the relative isolation and adaptive

divergence of the CBI community; and d) the importance of the Great Australian Bight for SABD gene flow and that impacts to this region are likely to also affect CBI residents and the wider SABD metapopulation in southern Australia. In addition, conservation and management plans also need to consider the influence of environmental variables on population genomic structure and how this is likely to be altered with ongoing climate change. The five environmental variables that are putatively influencing SABD genomic differentiation most significantly, based on this genomic dataset, are minimum SST, salinity, minimum chloA and current velocity range and minimum. This may be through indirect impacts on prey assemblages or directly on the dolphins themselves. These environmental variables are particularly susceptible to climate change and thus could have serious implications for bottlenose dolphin populations. Indeed, climate change and specifically increases in SST and salinity, were found to be posing the most significant threat to the long–term viability of marine mammals in SG (Robbins et al. 2017). Without modelling future climatic conditions however, it is difficult to foresee what impact this may have on SABD population genomic differentiation. Two possible opposing scenarios are that changes in these conditions will cause further fragmentation of the metapopulation, or that climate change may homogenise environmental gradients and facilitate gene flow across the region. In either scenario, climate change will result in strong selective pressure for traits involved in adaptation to the changing conditions (Gienapp et al. 2008). It is therefore, important to safeguard SABD's standing genomic variation and gene flow across southern Australia, to enable these dolphins to retain potentially crucial genomic variation that will assist as they attempt to adapt to environmental changes.

Due to the nature of ddRADseq the entire SABD genome cannot be investigated. It is therefore, acknowledged that there are likely many other genes influencing the adaptive differentiation and population genomic structure of coastal bottlenose dolphins in southern Australia. Future research with the use of whole-genome sequencing will enable the genome to be studied at high resolution and will likely enable the detection of several other candidate genes associated with adaptation to environmental conditions. With a dataset of over 8,000

SNPs however, population genomic structure at both fine and broad scales, as well as potential drivers of these patterns, were detected. Over-enrichment in GO terms and selection on candidate genes associated with processes that may be crucial to the adaptation of cetaceans and potentially other marine species were identified. This highlights the importance of maintaining gene flow and genomic diversity in marine populations to ensure that they are well equipped to adapt to ongoing human-induced climate change.

4.8 Conclusion

A high-quality SNP dataset was used to test the relationship between population genomic structure of SABD and regional environmental heterogeneity. Both adaptive and neutral SNP datasets supported the presence of a hierarchical metapopulation of SABD in southern Australia. Fine-scale adaptive differentiation potentially driven by strong environmental gradients was also evident. Hundreds of candidate loci putatively involved in this differentiation were identified. Analysis of candidate loci revealed potential selection on genes associated with several processes that may be involved in the adaptation of these dolphins to their local environments, particularly gradients in SST and salinity. This included over-enrichment in GO terms and putative selection on genes associated with adipogenesis, thermogenesis, ion channel activity and kidney development. Changes in these physiological processes have been demonstrated to be important to the adaptation of both marine and terrestrial species to different climates. This study therefore, highlights several candidate genes and associated physiological processes that may be vital in the adaptation of marine species to ongoing human-induced climate change. While the capacity of these species to adapt to climate change is largely unknown, it is important that gene flow and genomic diversity are maintained in marine populations to allow adaptive resilience to changing environments.

Chapter 5 : General Discussion

5.1 Introduction

Climate change and other anthropogenic threats are significantly altering habitats worldwide. This is particularly pronounced in the marine environment, with further increases in SST, ocean acidification and sea level predicted, as well as changes to the strength and distribution of currents and upwelling systems (Suthers et al. 2011; Poloczanska et al. 2013; Scheffers et al. 2016). Additional human disturbances are created through habitat degradation and destruction, tourism, shipping, pollution and overfishing (Gales et al. 2003; Pelletier and Coltman 2018). Many marine organisms are heavily influenced by the oceanographic and environmental features of their habitats, and thus anthropogenic stressors are likely to have major consequences for these marine species. This justifies the importance of understanding interactions between marine organisms and their environment and how this can lead to adaptation and genomic divergence. Scientists can then evaluate adaptive capacity of species in the face of ongoing habitat and climate change and propose appropriate conservation and management strategies. To address these issues, this study investigated the genomic basis of environmental adaptation in bottlenose dolphins (genus *Tursiops*) using ddRADseq SNP datasets. Specifically, potential spatial and environmental drivers of genomic divergence at species, subspecies and population levels were explored and tested. This was addressed by first investigating phylogenomic relationships among currently proposed and recognised bottlenose dolphin species and subspecies in the Southern Hemisphere – *T. t. truncatus*, *T. t. gephyreus*, *T. aduncus* and SABD. Within the two Australian inshore bottlenose dolphin lineages, *T. aduncus* and SABD, population-level genomic structure was also assessed. Environmental features were then considered for how they may drive species, subspecies and population-level genomic differentiation in bottlenose dolphins. Genomic datasets of ~8,000 – 18,000 high-quality SNPs were used for each individual data chapter. This ensured adequate power to detect fine-scale genomic differentiation and enabled individual candidate loci to be investigated, identifying potential pathways driving adaptation and divergence in these dolphins.

5.2 *Tursiops* Genomic Divergence

The genus *Tursiops* has a long history of confusing taxonomy, with as many as 20 species previously described (Hershkovitz 1966), but only two currently recognised (Committee on Taxonomy of the Society for Marine Mammalogy 2019). Conflicting patterns of divergence have been found by different DNA markers due to incomplete lineage sorting and potential hybridisation among *Tursiops* species and with other delphinids (Amaral et al. 2012b; Moura et al. 2013). Repeated evolution of inshore and offshore bottlenose dolphin ecotypes worldwide (e.g. Hoelzel et al. 1998; Wang et al. 1999; Fruet et al. 2014; Oudejans et al. 2015; Allen et al. 2016) has further complicated efforts to resolve *Tursiops* taxonomy. In the Southern Hemisphere several *Tursiops* lineages exist with mixed levels of taxonomic acceptance. Genomic divergence within the genus *Tursiops* in the Southern Hemisphere was therefore, investigated for one offshore (*T. t. truncatus*) and three inshore (*T.* cf. *australis* (SABD), *T. aduncus* (IPBD) and *T. t. gephyreus*) lineages spanning over three ocean basins. Strong genomic differentiation was found among all lineages, with proposed subspecies divergence within the two currently recognised species, IPBD and *T. truncatus.* Specifically, support was given for the current subspecies classification of the inshore SWAO dolphins, *T. t. gephyreus*, within *T. truncatus*. This is conservatively suggested based on significant reproductive isolation and divergence found in genetic/genomic, phenotypic and osteological studies (Chapter 2; Costa et al. 2016; Fruet et al. 2017), whilst also taking into account previous reports of gene flow between the two lineages (de Oliveira et al. 2019). Significant reproductive isolation between these lineages is supported by the finding that the offshore SWAO *T. t. truncatus* dolphins are more genomically similar to other offshore dolphins across three ocean basins, than to the geographic neighbour *T. t. gephyreus*. This is despite being regularly sighted in mixed groups with these dolphins (Fruet et al. 2017). It is also proposed here that the previously described *T. australis* (SABD) should instead be recognised as a subspecies of *T. aduncus*. This is in support of Moura et al. (in review), and is based on several lines of evidence including morphology and genetics/genomics (Chapter 2; Charlton et al. 2006;

Möller et al. 2008; Charlton-Robb et al. 2011). This is a conservative classification while still acknowledging the clear divergence between the two lineages. No previous studies of this genus, phylogenomic or otherwise, have included both *T. t. gephyreus* and SABD. This project provides novel insights into the genomic relationships within the genus *Tursiops* and into how these lineages have evolved in the opposing inshore and offshore environments.

With ecotypic differentiation and subsequent reproductive isolation likely driven by adaptation to their respective habitats, it is suggested that environmental heterogeneity has a strong impact on genomic differentiation in bottlenose dolphins. This is suggested to affect these dolphins not just at the species and subspecies levels but also creating population-level genomic differentiation (Natoli et al. 2005; Bilgmann et al. 2007b; Möller et al. 2007; Wiszniewski et al. 2010; Amaral et al. 2017). Population genomic structure and the association with environmental and oceanographic patterns were therefore, investigated in the inshore Australian lineages, IPBD and SABD. This was investigated using separate neutral and putatively adaptive datasets, allowing the effects of neutral processes, such as genetic drift, migration and mutation, to be untangled from the impact of natural selection (Nosil and Feder 2012; Raeymaekers et al. 2017). At the neutral level population genomic structure in both SABD and IPBD appears to be influenced by IBD and social behaviour, which in turn may be impacted by prey and habitat specialisation. Therefore, similar to results found at the species and subspecies level (Chapter 2), albeit at finer scales, habitat type changes along southern and eastern Australian coasts appears to be impacting bottlenose dolphin population genomic divergence. For example, there was strong neutral differentiation of embayment communities, such as GSV (Chapter 4) and PS (Chapter 3), from neighbouring populations. Neutral genomic differentiation was however, also associated with less obvious habitat changes such as between east and west Eyre Peninsula (Chapter 4), and northern, central and southern NSW populations (Chapter 3). These examples may be impacted by the southern Eyre Peninsula upwelling system and water mass discontinuity in the EAC, respectively. It is hypothesised

that habitat and environmental heterogeneity in these regions impacts on prey distribution and abundance and subsequently on bottlenose dolphin resource specialisation and social structure. This in turn, indirectly affects neutral population differentiation in these dolphins, as has been suggested previously for other populations (Hoelzel 2009; Möller 2012). While the putatively adaptive loci also revealed structure corresponding to the major habitat type changes specified above (e.g. GSV and PS), more subtle differentiation was found associated with environmental and oceanographic gradients in the regions. Previous studies of cetacean population genetic structure only had the power to infer these relationships, without empirically testing for associations. Through the implementation of seascape genomics frameworks in Chapters 3 and 4, this study statistically evaluated relationships between bottlenose dolphin genomic variation and a large suite of environmental and oceanographic variables. This revealed previously undisclosed adaptive differentiation (e.g. between NSG and southern SG populations (Chapter 4)) associated with gradients in ecological features such as SST, depth, salinity and productivity. Previous seascape genetic/genomic studies have revealed that these variables appear to be influencing population structure in other marine species, including common dolphins (Amaral et al. 2012a), sandgobies (*Psammogobius knysnaensis*; Teske et al. 2019), and eastern oysters (*Crassostrea virginica*; Bernatchez et al. 2019), among many others. With climate change already impacting on these variables worldwide (Böning et al. 2008; Domingues et al. 2008; Van Gennip et al. 2017), it is likely that population structure of bottlenose dolphins, and potentially many other marine species, will be strongly affected. While formal modelling is needed to predict how these populations may respond to climate change, further population subdivision and fragmentation is likely and could be detrimental to population and lineage persistence.

With environmental discontinuities implicated in species, subspecies and population-level divergence in bottlenose dolphins, it appears that there is a close relationship between the two. This suggests that the environment can affect genomic differentiation over scales of just tens of kilometres up to thousands. Next-generation sequencing techniques now allow

investigations of the genomic basis to this differentiation and can provide insights into how these parameters are potentially impacting on bottlenose dolphin genomic adaptation, and that of other species.

5.3 Genomic Basis of Environmental Adaptation

The genomic basis of the macroevolutionary transition of cetaceans into the aquatic environment has been well documented (Sun et al. 2012; Nery et al. 2013b; McGowen et al. 2014; Yim et al. 2014; Foote et al. 2015; Huelsmann et al. 2019). At the microevolutionary level (i.e. within lineages) however, the gene regions potentially associated with local adaptation in cetaceans have not been extensively explored (but see Ruan et al. 2015; Foote et al. 2016; Zhou et al. 2018b). Genomic datasets of between ~8,000 and 18,000 SNPs were therefore, used to investigate the genomic basis of adaptation within the genus *Tursiops*. Hundreds of candidate loci were identified to be potentially involved in the evolution of inshore and offshore bottlenose dolphin ecotypes and in local environmental adaptation in populations of the two Australian inshore lineages. Functional enrichment analysis and individual annotation of these loci then revealed the major biological pathways that are putatively under selection to allow adaptation to particular habitats and/or environmental variables in bottlenose dolphins in the Southern Hemisphere. Modification of several important physiological systems and processes were found to be potentially involved in the successful colonisation of the inshore habitat, many of which were also identified to be putatively under some degree of selection at the population level within the Australian inshore lineages. Adaptation of many of these same bodily systems, gene families and indeed, specific genes have been implicated in the macroevolution of cetaceans (e.g. McGowen et al. 2012; Nery et al. 2013b; Foote et al. 2015; Zhou et al. 2015; Foote et al. 2016; Zhou et al. 2018b; see Table 5.1). This suggests that these modifications have been occurring over millions of years to allow cetaceans to colonise the oceans, and subsequently diversify and specialise in specific regions and habitats. This study documents microevolutionary adaptations occurring within the bottlenose

dolphin lineages. Variation in exonic (coding) regions of several candidate genes was discovered, particularly involved in the evolution of inshore and offshore bottlenose dolphin ecotypes (Chapter 2), and population-level differentiation of SABD (Chapter 4). Of particular note is differentiation in exonic regions of the genes *CMKLR1* and *RYR2* that vary in SABD in correlation with the SST and salinity gradients in southern Australia, respectively. These genes may have critical roles in the adaptation of SABD to the strong environmental gradients in the region by regulating processes around thermoregulation and osmoregulation. Many of the loci identified as candidates for selection were however, not found in exonic regions of their respective genes. Very little is known about the importance of non-coding regions, but there is growing evidence that adaptive changes in these sections of the genome play an important part in the adaptation and evolution of species (Andolfatto 2005; Jones et al. 2012). Variation in non-coding regions was found to be putatively involved in adaptation of almost all major bodily systems in bottlenose dolphins in the Southern Hemisphere, providing support to this hypothesis. Further work is needed to understand the roles of coding versus non-coding genes in the adaptive response of marine species.

Table 5.1 Candidate adaptive genes found to be under selection in multiple data chapters of this thesis and/or previously published literature (including members of the same gene family)

Chapter 2 (Genomic Divergence and Ecotype Formation)		Chapter 3 (Seascape Genomics of *T. aduncus*)	Chapter 4 (Seascape Genomics of *T.* cf. *australis*)	Candidates from the Literature	Reference
Parallel Evolution Candidate	**Early-stage Evolution Candidate**				
AGR2				*AGRP*	McGowen et al. 2012
AP3M2				*AP3S1*	Foote et al. 2015
ARHGAP12				*ARHGAP5*	Foote et al. 2015
				ARHGAP8	Huelsmann et al. 2019
			ARRB2	*ARR3*	Zhou et al. 2013
			BHLHE40	*BHLHB9*	Moura et al. 2014
CACNA1B			*CACNA1A*	*CACNB4*	Zhou et al. 2015
				CACNG7	Foote et al. 2015
CALHM6				*CALHM1*	Chikina et al. 2016
CDH23				*CDH23*	Shen et al. 2012a; Zhou et al. 2013; McGowen et al. 2014
				CDH24	Sun et al. 2012
			DDX31	*DDX54*	Foote et al. 2015
				DDX58	Huelsmann et al. 2019
DNAJB6				*DNAH1, DNAH3, DNAH7*	Sun et al. 2012
				DNAJC22	Huelsmann et al. 2019
DUSP23 (exonic)				*DUSP28*	Foote et al. 2015
		ELOVL2		*ELOVL3*	Huelsmann et al. 2019; Lopes-Marques et al. 2019
FREM2				*FREM1, FREM2*	Sun et al. 2012
			GRIK3	*GRIA1, GRIA2*	Zhou et al. 2015
				GRIN2C, GRIN3B	Chikina et al. 2016
				GRIN3A	Foote et al. 2015
GTF2IRD1 (exonic)				*GTF2B*	DeFaveri et al. 2011
IFNGR				*IFNGR1*	Zhou et al. 2018a
IL2RB, IL22RA2				*IL1F10, IL31RA*	Huelsmann et al. 2019

				IL20	McGowen et al. 2012
	KCNH5		*KCNT2*	*KCNA5, KCNQ3*	Foote et al. 2015
				KCNJ2, KCNK18	McGowen et al. 2012
				KCNG4	Zhou et al. 2015
				KCNMA1	Zhou et al. 2018b
				KCNMB3	Huelsmann et al. 2019
			KIAA0556	*KIAA1468, KIAA1683*	Foote et al. 2015
KIF13A				*KIF27, KIF2C*	Foote et al. 2015
LRRC1				*LRRC29*	Huelsmann et al. 2019
				LRRC66	Foote et al. 2015
MAPKAPK3				*MAP3K14*	Foote et al. 2015
				MAP3K19	Huelsmann et al. 2019
				MAPK10	Zhou et al. 2015
				MAPK8	Batley et al. 2018
MYH11				*MYH7B*	Foote et al. 2015
NKX2-2, NKX2-3				*NKX2*	Nery et al. 2013b
OTUD5				*OTUD6A*	Foote et al. 2015
PDE1C, PDE9A				*PDE3B*	Wang et al. 2015
PIK3R1				*PIK3CB, PIK3R1*	Zhou et al. 2018a
				PIK3R5	Chikina et al. 2016
		PIP5K1B		*PIP4K2B*	Foote et al. 2015
PRKG1	*PRKAG2*			*PRKAG3*	Foote et al. 2015
				PRKCA	Zhou et al. 2015
PRRT4				*PRRT3*	Foote et al. 2015
RBM20				*RBM23*	Nery et al. 2013b
				RBM41	Foote et al. 2015
			RNF130	*RNF182*	McGowen et al. 2012
				RNF222	Chikina et al. 2016
				RNF24	Foote et al. 2015
	RYR2		*RYR2*		
SEC14L1 (exonic)			*SEC14L1*	*SEC14L3*	Huelsmann et al. 2019

SEMA3E				*SEM141*	Sun et al. 2012
SHROOM4				*SHROOM4*	Foote et al. 2015
			SLC22A18	*SLC12A1*	Ruan et al. 2015
				SLC14A2 (UT2)	Ruan et al. 2015; Zhou et al. 2018b
				SLC16A1	Yim et al. 2014
				SLC26A5 (Prestin)	Li et al. 2010; Liu et al. 2010; Shen et al. 2012a; Zhou et al. 2013; McGowen et al. 2014
				SLC27A2	Nery et al. 2013b
				SLC35A1, SLC39A12, SLC9B2	Foote et al. 2015
				SLC5A1	Wang et al. 2016
				SLC9A3	Ruan et al. 2015
				SLC28A1, SLC47A1, SLC4A9, SLC6A18, SLC25A41	Huelsmann et al. 2019
THOC3 (exonic)				*THOC6*	Nery et al. 2013b
		TRIM68		*TRIM14, TRIM34*	Huelsmann et al. 2019
				TRIM29	Chikina et al. 2016
				TRIM63	Nery et al. 2013b
USP10				*USP26*	McGowen et al. 2012
				USPL1	Foote et al. 2015
			VPS13D	*VPS8*	Foote et al. 2015
	ZNF345			*ZNF106, ZNF12, ZNF136, ZNF23, ZNF304, ZNF350, ZNF582, ZNF791*	Foote et al. 2015
				ZNF597	McGowen et al. 2012

5.3.1 Osmoregulation and Salinity Tolerance

Variation was found in both coding and non-coding regions of several genes potentially associated with adaptation of the osmoregulatory and salinity tolerance systems in the inshore bottlenose dolphin ecotype (Chapter 2) and on a finer scale within the SABD inshore lineage correlated with a stark salinity gradient in SG (Chapter 4). Potassium channel genes (*KCN* genes) were identified as being putatively involved in the successful colonisation of the opposing environments of the hypersaline NSG waters (Chapter 4) and the estuarine conditions in SWAO by *T. t. gephyreus* (Chapter 2; Table 5.1). These genes are associated with neural development and functioning (Ju and Wray 2002; Tomasello et al. 2015), and are activated by high sodium levels (Tomasello et al. 2017). This family of genes was also important in the macroevolution of cetaceans and their subsequent diversification into both freshwater and marine environments (McGowen et al. 2012; Foote et al. 2015; Zhou et al. 2015; Zhou et al. 2018b; Huelsmann et al. 2019). Several other candidate genes and over-enriched GO terms were also found to be associated with kidney development and functioning in Chapters 2 and 4, suggesting that salinity gradients are likely driving local osmoregulatory adaptation in inshore bottlenose dolphins, potentially similar to that found for freshwater finless porpoises (*Neophocaena asiaeorientalis asiaeorientalis*; Ruan et al. 2015; Yuan et al. 2018).

5.3.2 Cardiovascular System

Salinity gradients not only drive changes in osmoregulatory pathways. Differentiation was found in a homolog of *RYR2* correlated with the salinity gradient in SG in southern Australia (Chapter 4). This gene was also found as a candidate for early-stage evolution of inshore SWAO dolphins (Chapter 2; Table 5.1). With *RYR2* being identified as a candidate for selection at both the subspecies and population divergence levels, it is clear that variation in this gene is important for the adaptation of bottlenose dolphins to environmental heterogeneity. *RYR2* is involved in cardiac muscle contraction and could aid in hypoxia tolerance (Toischer et al. 2013). It is therefore, proposed that this gene may be helping in the adaptation of the

bottlenose dolphin cardiovascular system to different salinity levels and hypoxic conditions. In this case hypoxia may be as a result of high salinity waters (e.g. in NSG) or due to deep, prolonged dives (e.g. in the offshore ecotype). Several other candidate genes related to heart and blood vessel development were also found to be implicated in adaptation of bottlenose dolphins to the inshore environment (Chapter 2), to the PS embayment (Chapter 3) and to the SST gradient across southern Australia (Chapter 4). While cardiovascular adaptations are likely to, at least in part, be associated with salinity and hypoxia tolerance, changes in SST as seen in southern Australia may also drive changes in this system through increased stress on the body. These kinds of adaptations have been previously reported in northern elephant seals (*Mirounga angustirostris*; Meir et al. 2009), and Atlantic salmon (*Salmo salar*; Anttila et al. 2013), among other species. This suggests that changes to the cardiovascular system may be important in the adaptation of marine species to environmental heterogeneity. It should also be noted that *RYR2* is inhibited by high levels of the essential omega-3 fatty acids EPA and DHA, found in excess in oily fish such as herring, tuna and sardines (Ismail 2005). With these fish all being potential prey items for bottlenose dolphins (see Gibbs et al. 2011), this suggests that adaptations of the heart in SABD and inshore SWAO dolphins could also be in relation to a change in diet between populations and ecotypes.

5.3.3 Energy Production and Thermoregulation

Prey choice has been suggested to differ between inshore and offshore bottlenose dolphin ecotypes worldwide based on observations, stable isotope analysis and osteological differentiation (Walker 1981; Mead and Potter 1995; Wang et al. 2000; Gibbs et al. 2011; Costa et al. 2016). Different prey items inevitably means different intake levels of fat and essential nutrients, such as DHA and EPA (Ismail 2005; Ishikawa et al. 2019). This in turn may create differing selective pressures on the dolphins for energy production, fat storage and the synthesis of EPA and DHA. Accordingly, the inshore ecotype was found to have over-enrichment in GO terms, and significant differentiation in several genes, related to energy

production and adipogenesis pathways impacted by diet (Chapter 2). This included the *PRDM16* gene, potentially important in early-stage evolution of inshore SWAO dolphins. This gene has a central role in brown adipose tissue deposition for thermoregulation and metabolism and is known to be affected by DHA and EPA levels (Zhuang et al. 2019). Subsequently, variation in the candidate gene *ELOVL2* was discovered to be correlated with the productivity gradient across the NSW coast (Chapter 3). This gene is involved in the synthesis of EPA and DHA, likely moderated by the amount of these essential nutrients gained through the diet (Morais et al. 2009; Monroig et al. 2016). Several other genes related to the development and functioning of the gastrointestinal system were found to be potentially under selection in the inshore ecotype (Chapter 2). Differentiation in fish assemblages associated with changes in habitat type, salinity and productivity, may therefore, be driving adaptations to handle and metabolise different prey items in these dolphins. With repeated selection on genes influenced by EPA and DHA levels across each data chapter (including *RYR2* in Chapters 2 and 4), it appears that these elements of the diet are particularly important in this relationship. Indeed, Ishikawa et al. (2019) suggested that genetic adaptation in freshwater threespine sticklebacks (*Gasterosteus aculeatus*) and potentially several other teleost species, was occurring in response to a lack of DHA in the freshwater ecosystem food web. While diet may be impacting thermoregulatory pathways, they are also likely affected by SST gradients. This could be in part driving selection on *PRDM16* in inshore SWAO dolphins corresponding to a potential change in temperature profile from the offshore environment. This was also seen in southern Australia, where the strong east-west thermal gradient was correlated with variation in the candidate gene *CMKLR1* (Chapter 4), a gene with functions in adipogenesis and thermoregulation (Ernst et al. 2012; Audet et al. 2016). There was also significant over-enrichment in associated GO terms, such as temperature homeostasis, adaptive thermogenesis and cold-induced thermogenesis, in Chapter 4. The SST differentiation between inshore and offshore habitats, as well as within the inshore environment may therefore, impact on the adaptation of energy production and adipogenesis pathways through differing thermogenic pressures. Adaptation of these processes to differing

thermal profiles and diets has also been reported in killer whales (*Orcinus orca*) and polar bears (*Ursus maritimus*) (Liu et al. 2014; Foote et al. 2016), as well as in human populations (Clemente et al. 2014; Sazzini et al. 2014; Fumagalli et al. 2015). Modification of these pathways is therefore, potentially an important step in the adaptation of mammals to different climates and diets and may become increasingly important with ongoing ocean warming and changes to prey abundance and distribution.

5.3.4 Musculoskeletal System

Macroevolutionary changes to the musculoskeletal system are perhaps the most well studied examples of adaptation in the cetacean transition from land to water (e.g. Thewissen and Bajpai 2001; Thewissen et al. 2007; Thewissen et al. 2009). Musculoskeletal differences have also been extensively documented at the microevolutionary scale, between bottlenose dolphin species, subspecies and ecotypes (Mead and Potter 1995; Wang et al. 2000; Kemper 2004; Costa et al. 2016). They are thus potentially important to the successful colonisation of new aquatic niche spaces. The genomic basis to these differences however, had not been studied previously. Several genes related to skin, muscle and bone development were found to be potentially associated with adaptation to the major habitat changes from offshore to inshore (Chapter 2) and from the open coast to the PS embayment (Chapter 3). Although not discussed in detail in Chapter 4, variation in an additional four genes with GO terms related to skeletal development were found to be significantly correlated with salinity and SST gradients in southern Australia. This was further corroborated by significant over-enrichment in GO terms such as positive regulation of muscle tissue development and skeletal muscle cell differentiation, among others (Chapter 4). In each case, adaptations of the musculoskeletal system are likely related to changes in selective pressures associated with diving and locomotion, as well as potential sensory adaptations to allow successful colonisation of different inshore habitats. Discovery of variation in genes associated with this system among populations of IPBD and SABD poses the question of whether there are detectable

morphological and skeletal differences within these bottlenose dolphin lineages, which warrants further study.

5.3.5 Sensory System

Odontocete cetaceans developed the ability to echolocate in the Late Eocene (36-34 MYA) (Steeman et al. 2009). Due to reduced visibility in the water, echolocation proved to be a more efficient method for hunting and communication for the raptorial, social lifestyle of odontocetes (see McGowen et al. 2014). Development of this ability was enabled by positive selection in a number of genes associated with high frequency hearing and echolocation. This was coupled with inactivation of genes related to the aquatic eye and olfactory bulb as odontocetes came to rely less on sight and smell (McGowen et al. 2014). Echolocation likely allowed these cetaceans to exploit new niche spaces and food resources (see Fordyce 1980; Steeman et al. 2009), and has subsequently differentiated among species and ecotypes (e.g. *T. aduncus* and *T. truncatus*, Wahlberg et al. 2011; and sympatric fish- and mammal-eating killer whale ecotypes, Barrett-Lennard et al. 1996). Differentiation was found in several genes associated with eye and inner ear development between inshore and offshore bottlenose dolphin ecotypes, potentially implicated in parallel evolution in inshore dolphins (Chapter 2). Genes with similar GO terms were discovered to vary with the SST and salinity gradients in southern Australia (Chapter 4). This included a homolog of the gene *SLC22A18.* While this specific gene is likely associated with kidney functioning and fat accumulation, it is of the same family as the gene *SLC26A5* (i.e. the protein Prestin), which is strongly implicated in the evolution of echolocation in both bats and odontocete cetaceans (Li et al. 2010; Liu et al. 2010; Shen et al. 2012a; Zhou et al. 2013; McGowen et al. 2014). Several other members of this gene family have been discovered to be under selection in both the macro- and microevolution of cetaceans. *SLC* genes should therefore, be studied in more detail due to their potential importance for adaptation of cetaceans to different aquatic environments. Modifications to the visual and auditory sensory systems at the ecotype and population levels could be associated

with habitat changes (i.e. habitat complexity, depth, noise pollution), and/or hunting specialisations for different prey items as has been suggested by previous studies of echolocation activity in odontocetes (Barrett-Lennard et al. 1996; Wahlberg et al. 2011; Gutstein et al. 2014; Samarra and Miller 2015; La Manna et al. 2019). While population-level differences in echolocation have not been previously documented in bottlenose dolphins, future study in this area could help to identify potential changes in clicking behaviour related to prey and habitat type.

5.3.6 Brain and Nervous System

Inshore bottlenose dolphins generally demonstrate higher social and behavioural complexity than their offshore counterparts (Möller 2012), with demonstrated region-specific prey handling and tool use techniques (e.g. Krützen et al. 2005; Finn et al. 2009). These factors have previously been suggested to influence brain size (Marino 2005). Large brain size could be an important factor in the successful colonisation of new habitats by improving innovation and adaptability, as has been shown in birds (Sol et al. 2005). This could be an important factor associated with differentiation found in several genes related to development and functioning of the brain and nervous system between inshore and offshore bottlenose dolphins, and among inshore populations. Although not discussed in detail in Chapter 4, differentiation in five genes was found to be correlated with salinity and SST gradients in southern Australia, including genes with functions in memory and learning. Several brain and nervous system genes were also found to be repeatedly selected for across the inshore lineages and also implicated in the early stages of *T. t. gephyreus* evolution in SWAO (Chapter 2). Adaptation of these systems may be important for successful colonisation of inshore environments, including the development of intricate social systems typically seen in inshore bottlenose dolphin populations (see Möller 2012). This may also enable subsequent local adaptation to prey items and habitat features, working to reinforce the disclosed population genomic structure in the inshore bottlenose dolphin lineages.

5.3.7 *Differences and Parallelism in Adaptive Responses of Bottlenose Dolphin Inshore Lineages*

This study investigated the potential for parallel adaptive responses in inshore bottlenose dolphins in the Southern Hemisphere. This was addressed in Chapter 2 by comparing genes putatively under selection among inshore-offshore ecotype pairs, and in Chapters 3 and 4 by conducting similar seascape genomics analyses on IPBD in eastern Australia and SABD in southern Australia, respectively. This allows comparison of how environmental and oceanographic heterogeneity impact the two Australian inshore lineages to identify similarities and/or differences in the way that they respond to these selective pressures. While it is clear that both Australian inshore lineages are affected by environmental and oceanographic features, differences in the strength of these gradients between the east and south coasts likely influences the strength of the adaptive response in the dolphins. For example, SST and salinity were identified as potentially being the most influential environmental variables along the southern coast (Chapter 4), while primary productivity and chloA concentration were significant along the eastern coast (Chapter 3). The different environmental gradients are likely impacting prey abundance and distribution in the area and are subsequently creating differing adaptive responses in the two lineages. Specifically, it is hypothesised that for both IPBD on the eastern coast and SABD on the southern coast, difference in prey choice along heterogeneous environments is driving physiological adaptations, but in contrasting ways. This was exemplified in IPBD by variation in the gene *ELOVL2* in association with the primary productivity gradient in NSW (Chapter 3) and in SABD by variation in *CMKLR1* and *RYR2*, likely also influenced by the region's SST and salinity gradients (Chapter 4). Potential adaptive response to differences in prey choice was also a major finding of Chapter 2. Parallel adaptation was suggested for genes associated with the gastrointestinal system and mouth, and adipogenic and energy production pathways among the inshore bottlenose dolphin lineages. This demonstrates that not all inshore environments create the same selective pressures, although there is some overlap that may create similarity in the adaptive responses

of inshore bottlenose dolphin lineages across the Southern Hemisphere. This thesis demonstrates that bottlenose dolphins have a vast capacity to adapt to a wide range of environmental features, including both subtle and steep environmental and habitat gradients. These adaptations have however, likely taken place over thousands of years. For example, *T. t. gephyreus*, representing perhaps the most recent divergence, colonised the inshore environment several thousand years ago, likely after the LGM (see Fruet et al. 2017). It may be that anthropogenically accelerated climate change and habitat modification are occurring at a pace that is too rapid for these dolphins, and indeed many other long-lived species, to adapt to.

5.4 Contribution to Science

Genomic techniques are quickly emerging as a powerful tool to understand how organisms interact with their local environment and how this can drive patterns of divergence and speciation. This has been developed in terrestrial systems and is now being introduced into the marine environment and applied to a wide variety of non-model organisms. While genomic methods have been utilised extensively in recent years to study the macroevolution of cetaceans (e.g. Nery et al. 2013b; McGowen et al. 2014; Foote et al. 2015; Zhou et al. 2015), the microevolutionary adaptations are just starting to be explored (finless porpoises, Ruan et al. 2015; Zhou et al. 2018b; killer whales, Moura et al. 2014; Foote et al. 2016). This thesis investigated the genomic basis of ecotype formation and population structure in bottlenose dolphins (genus *Tursiops*) of the Southern Hemisphere. Evidence is provided for subspecies classification of SABD within *T. aduncus*, showing clear genomic divergence between the two lineages. As SABD are currently classified as IPBD, they may not be being managed appropriately with strategies that are tailored to the lineage and regional conditions. Reassessment of conservation plans for SABD in southern Australia are therefore, needed, considering the hierarchical metapopulation and adaptive differentiation disclosed here. Using the unique bottlenose dolphin system where inshore-offshore ecotype pairs have repeatedly

evolved, the potential for parallel evolution in the inshore lineage is explored, a concept that has not been previously reported in cetaceans. While several studies have inferred environmental associations in cetaceans before, no study has empirically tested for such relationships at a genomic level. As such, Chapters 3 and 4 are the first seascape genomics studies of any marine mammal. The seascape genomics framework allowed specific environmental features to be identified as potentially driving patterns of genomic differentiation and local adaptation. Loci that are putatively under selection could also then be further investigated to identify physiological systems and processes that are potentially undergoing adaptation. With marine ecosystems being particularly threatened by anthropogenic activities, it is important to understand the factors affecting genomic divergence and the adaptive capacity of species to properly inform management and conservation strategies now and into the future. This study found that a complex interplay of IBD, social factors and ecological adaptation are likely driving the genomic divergence of bottlenose dolphins at the species, subspecies and population levels. This information must be considered when reassessing conservation management strategies. In particular, that climate change and anthropogenic disturbances will likely impact on bottlenose dolphin adaptive differentiation and may further fragment these populations. The specific ecological variables that impact bottlenose dolphin adaptation and divergence are assessed at a genomic level for the first time in this thesis, providing new insights into how climate change may impact this dolphin genus. By identifying the bottlenose dolphin bodily systems that are undergoing adaptation to the inshore and offshore habitats and specific environmental features, and comparing this to previous studies in other species, common patterns were found and can be extrapolated for understudied organisms. This study therefore, not only informs in regard to how bottlenose dolphins are likely to be impacted by further climate change and habitat modifications but suggests the adaptive pathways that will be important for cetaceans, and perhaps to other marine species and their wider communities.

5.5 Limitations and Future Directions

It is recommended that this study be expanded to include the other putative bottlenose dolphin lineages (e.g. *T. aduncus* from the African coast and Arabian Sea). Ideally all proposed and recognised lineages would be included in a worldwide investigation, but with the origin of this genus proposed for coastal Indo-Pacific waters a study covering the Indian and Pacific Oceans would provide important insights into the early pattern of bottlenose dolphin radiation and ecological divergence. As it was not the aim of this thesis to resolve taxonomic confusion in the genus *Tursiops*, only basic phylogenomic analyses were carried out to determine the degree of genomic differentiation among the putative lineages. As such, it is suggested that further phylogenomic work is needed for *Tursiops* (and the wider delphinid lineage) and that this should include all known bottlenose dolphin lineages.

Having demonstrated the utility of the seascape genomics framework in the inshore Australian bottlenose dolphin lineages, GEA analysis between the inshore and offshore ecotypes could help to identify environmental and habitat variables that are driving ecotypic divergence. More comprehensive sampling of *T. truncatus* across the Southern Hemisphere (particularly in Australia) would however, be required to ensure adequate representation and sufficient statistical power. Despite not conducting formal GEA tests between the inshore-offshore bottlenose dolphin ecotypes, this study was able to identify hundreds of genes putatively under selection for adaptation to particular environments and habitats. Although many of these genes were discussed in detail, further investigation of many more is warranted. This includes the aforementioned candidate genes discovered in Chapter 4, and several additional genes associated with immune response and reproduction that were found to be under parallel evolution in the inshore ecotype but were not researched in detail (Chapter 2). This included the immunogene *SEC14L1*, which was also detected to exhibit allele frequency variation in SABD in correlation with the salinity gradient in southern Australia (Chapter 4; Table 5.1). This

highlights the potential importance of immune system genes to local adaptation in bottlenose dolphins, suggesting that they should be investigated further.

In the rapidly changing field of molecular ecology, whole-genome techniques are quickly becoming more feasible for use in non-model species, such as cetaceans. Future research should therefore, consider addressing the questions raised here using whole genomes with appropriate sequence coverage. While ddRADseq data cover vastly more of the genome than traditional genetic datasets (e.g. microsatellite makers), this still only represents a small fraction of the entire genome (see Davey et al. 2011). Whole-genome data provide high-resolution records of variants across the genome and information about causative genes, rather than information about markers as provided by ddRADseq and related methods. Despite almost all major bodily systems being found here to be putatively under some form of adaptation, further genomic changes are likely to have been missed. For example, no candidate genes were found to be associated with adaptation of the respiratory system. Using a whole-genome approach would allow the genomic basis of environmental adaptation in bottlenose dolphins to be more fully explored, potentially revealing many additional adaptations that were not found here.

Conclusion

This thesis investigated the genomic basis of environmental adaptation in bottlenose dolphins (genus *Tursiops*) using ddRADseq datasets of ~8,000 to 18,000 SNPs. This was addressed by first establishing the phylogenomic relationships among the recognised and proposed bottlenose dolphin species and subspecies in the Southern Hemisphere and then establishing how adaptation to the inshore and offshore habitats may have impacted this divergence. Within the two inshore Australian lineages, population genomic structure was elucidated and particular ecological variables that may be driving population-level genomic differentiation were identified. While patterns of neutral population structure appear to be influenced by both spatial and environmental variables, adaptive divergence is significantly associated with changes in habitat and environmental conditions (e.g. salinity, SST and productivity). The genomic basis of species, subspecies and population divergence was then investigated. While many differences in adaptive response were discovered among the inshore lineages, there was also substantial parallelism in the ways that these dolphins have adapted to environmental and habitat changes. Indeed, many of the same bodily systems, gene families and specific genes were putatively under selection at both the species/subspecies and population levels. Of particular note is the adaptation of thermogenic, energy production, digestive and metabolic pathways likely in response to differing diets in opposing regions, driven by changes in habitat and environmental features. This thesis presents the first seascape genomics studies of any marine mammal and provides novel insights into the microevolutionary adaptations of cetaceans, at the species, subspecies and population levels. The complex relationship between bottlenose dolphins and their local environmental conditions will likely be affected by climate change. It is therefore, important to understand how these changes may impact them and assess the adaptive resilience of these dolphins and other marine organisms. While evidence of the vast capacity of bottlenose dolphins to adapt to environmental heterogeneity was found, this has likely occurred over several thousands of years. The current pace of climate change may be too rapid for dolphins, and

species with similar life history traits, to adapt to. This work thus, provides important information that should be used in management and conservation plans to protect bottlenose dolphins and marine ecosystems. In the face of ongoing anthropogenic climate change and habitat modification, it is crucial that modern science is integrated into these strategies to ensure that they are targeted and effective in securing the future health of marine ecosystems.

Reference List

Adams LD, Rosel PE 2006, 'Population differentiation of the Atlantic spotted dolphin (*Stenella frontalis*) in the western North Atlantic, including the Gulf of Mexico', Marine Biology, vol. 148, no. 3, pp. 671-681.

Agashe D, Bolnick DI 2010, 'Intraspecific genetic variation and competition interact to influence niche expansion', Proceedings of the Royal Society B-Biological Sciences, vol. 277, no. 1696, pp. 2915-2924.

Ahmad F, Arad M, Musi N, He H, Wolf C, Branco D, Perez-Atayde AR, Stapleton D, Bali D, Xing Y 2005, 'Increased 2 subunit-associated AMPK activity and PRKAG2 cardiomyopathy', Circulation, vol. 112, no. 20, pp. 3140-3148.

Ahonen H, Lowther AD, Harcourt RG, Goldsworthy SD, Charrier I, Stow AJ 2016, 'The limits of dispersal: fine scale spatial genetic structure in Australian sea lions', Frontiers in Marine Science, vol. 3, pp. 65.

Albertson GR, Baird RW, Oremus M, Poole MM, Martien KK, Baker CS 2017, 'Staying close to home? Genetic differentiation of rough-toothed dolphins near oceanic islands in the central Pacific Ocean', Conservation Genetics, vol. 18, no. 1, pp. 33-51.

Alexander DH, Novembre J, Lange K 2009, 'Fast model-based estimation of ancestry in unrelated individuals', Genome Research, vol. 19, no. 9, pp. 1655-1664.

Allendorf FW 1986, 'Genetic drift and the loss of alleles versus heterozygosity', Zoo Biology, vol. 5, no. 2, pp. 181-190.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990, 'Basic local alignment search tool', Journal of Molecular Biology, vol. 215, no. 3, pp. 403-410.

Amaral AR, Beheregaray LB, Bilgmann K, Boutov D, Freitas L, Robertson KM, Sequeira M, Stockin KA, Coelho MM, Möller LM 2012a, 'Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (genus *Delphinus*)', PLoS One, vol. 7, no. 2, pp. e31482.

Amaral AR, Jackson JA, Möller LM, Beheregaray LB, Coelho MM 2012b, 'Species tree of a recent radiation: the subfamily Delphininae (Cetacea, Mammalia)', Molecular Phylogenetics and Evolution, vol. 64, no. 1, pp. 243-253.

Amaral AR, Smith BD, Mansur RM, Brownell RL, Rosenbaum HC 2017, 'Oceanographic drivers of population differentiation in Indo-Pacific bottlenose (*Tursiops aduncus*) and humpback (*Sousa* spp.) dolphins of the northern Bay of Bengal', Conservation Genetics, vol. 18, no. 2, pp. 371-381.

Anand-Apte B, Ebrahem Q, Cutler A, Farage E, Sugimoto M, Hollyfield J, Folkman J 2010, 'Betacellulin induces increased retinal vascular permeability in mice', PLoS One, vol. 5, no. 10, pp. e13444.

Andolfatto P 2005, 'Adaptive evolution of non-coding DNA in *Drosophila*', Nature, vol. 437, no. 7062, pp. 1149.

Ansmann IC, Parra GJ, Lanyon JM, Seddon JM 2012, 'Fine-scale genetic population structure in a mobile marine mammal: Inshore bottlenose dolphins in Moreton Bay, Australia', Molecular Ecology, vol. 21, no. 18, pp. 4472-4485.

Anttila K, Dhillon RS, Boulding EG, Farrell AP, Glebe BD, Elliott JA, Wolters WR, Schulte PM 2013, 'Variation in temperature tolerance among families of Atlantic salmon (*Salmo salar*) is associated with hypoxia tolerance, ventricle size and myoglobin level', Journal of Experimental Biology, vol. 216, no. 7, pp. 1183-1190.

Anushiravani A, Faghihi MA, Dastsooz H, Lankarani KB 2017, 'Glycogen storage disease IIIa: A private homozygous splice site mutation in AGL gene', Gene Reports, vol. 9, pp. 61-64.

Archer E 2016, 'rfPermute: estimate permutation p-values for Random Forest importance metrics', R Package, vol. Version 2.1.5, no. 1.

Armanet N, Metay C, Brisset S, Deschenes G, Pineau D, Petit FM, Di Rocco F, Goossens M, Tachdjian G, Labrune P 2015, 'Double Xp11. 22 deletion including SHROOM4 and

CLCN5 associated with severe psychomotor retardation and Dent disease', Molecular Cytogenetics, vol. 8, no. 1, pp. 8.
Attard CR, Beheregaray LB, Sandoval-Castillo J, Jenner KCS, Gill PC, Jenner M-NM, Morrice MG, Möller LM 2018, 'From conservation genetics to conservation genomics: A genome-wide assessment of blue whales (*Balaenoptera musculus*) in Australian feeding aggregations', Royal Society Open Science, vol. 5, no. 1, pp. 1-11.
Au W 2000, 'Hearing in whales and dolphins: an overview'. In: Au, W, Popper, A & Fay, R (eds) *Hearing by Whales and Dolphins*. Springer, New York, USA, pp. 1-42.
Au WW 1997, 'Echolocation in dolphins with a dolphin-bat comparison', Bioacoustics, vol. 8, no. 1-2, pp. 137-162.
Audet GN, Dineen SM, Quinn CM, Leon LR 2016, 'Altered hypothalamic inflammatory gene expression correlates with heat stroke severity in a conscious rodent model', Brain Research, vol. 1637, pp. 81-90.
Baghurst B, Lukatelich R, Smith D, Begg G, Lewis R, Smith R 2017, 'Findings from the Great Australian Bight Research Program – an integrated study of environmental, economic and social values', Journal of the Australian Petroleum Production & Exploration Association, vol. 57, no. 2, pp. 388-392.
Bakun A, Black BA, Bograd SJ, Garcia-Reyes M, Miller AJ, Rykaczewski RR, Sydeman WJ 2015, 'Anticipated effects of climate change on coastal upwelling ecosystems', Current Climate Change Reports, vol. 1, no. 2, pp. 85-93.
Baldwin J 1988, 'Predicting the swimming and diving behaviour of penguins from muscle biochemistry', Hydrobiologia, vol. 165, no. 1, pp. 255-261.
Baltazar-Soares M, Biastoch A, Harrod C, Hanel R, Marohn L, Prigge E, Evans D, Bodles K, Behrens E, Böning CW 2014, 'Recruitment collapse and population structure of the European eel shaped by local ocean current dynamics', Current Biology, vol. 24, no. 1, pp. 104-108.
Banks SC, Ling SD, Johnson CR, Piggott MP, Williamson JE, Beheregaray LB 2010, 'Genetic structure of a recent climate change-driven range extension', Molecular Ecology, vol. 19, no. 10, pp. 2011-2024.
Banks SC, Piggott MP, Williamson JE, Bové U, Holbrook NJ, Beheregaray LB 2007, 'Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin', Ecology, vol. 88, no. 12, pp. 3055-3064.
Barahona M, Broitman BR, Faugeron S, Jaugeon L, Ospina-Alvarez A, Veliz D, Navarrete SA 2019, 'Environmental and demographic factors influence the spatial genetic structure of an intertidal barnacle in central-northern Chile', Marine Ecology Progress Series, vol. 612, pp. 151-165.
Bardarson H, McAdam BJ, Thorsteinsson V, Hjorleifsson E, Marteinsdottir G 2017, 'Otolith shape differences between ecotypes of Icelandic cod (*Gadus morhua*) with known migratory behaviour inferred from data storage tags', Canadian Journal of Fisheries and Aquatic Sciences, vol. 74, no. 12, pp. 2122-2130.
Barford K, Yap CC, Dwyer ND, Winckler B 2017, 'The related neuronal endosomal proteins NEEP21 (Nsg1) and P19 (Nsg2) have divergent expression profiles in vivo', Journal of Comparative Neurology, vol. 525, no. 8, pp. 1861-1878.
Barnes TC, Junge C, Myers SA, Taylor MD, Rogers PJ, Ferguson GJ, Lieschke JA, Donnellan SC, Gillanders BM 2016, 'Population structure in a wide-ranging coastal teleost (*Argyrosomus japonicus,* Sciaenidae) reflects marine biogeography across southern Australia', Marine and Freshwater Research, vol. 67, no. 8, pp. 1103-1113.
Barragán-Barrera DC, May-Collado LJ, Tezanos-Pinto G, Islas-Villanueva V, Correa-Cárdenas CA, Caballero S 2017, 'High genetic structure and low mitochondrial diversity in bottlenose dolphins of the Archipelago of Bocas del Toro, Panama: A population at risk?', PLoS One, vol. 12, no. 12, pp. e0189370.
Barrett-Lennard LG, Ford JK, Heise KA 1996, 'The mixed blessing of echolocation: differences in sonar use by fish-eating and mammal-eating killer whales', Animal Behaviour, vol. 51, no. 3, pp. 553-565.

Barrett RD, Schluter D 2008, 'Adaptation from standing genetic variation', Trends in Ecology & Evolution, vol. 23, no. 1, pp. 38-44.

Batley KC, Sandoval-Castillo J, Kemper CM, Attard CRM, Zanardo N, Tomo I, Beheregaray LB, Moller LM 2019, 'Genome-wide association study of an unusual dolphin mortality event reveals candidate genes for susceptibility and resistance to cetacean morbillivirus', Evolutionary Applications, vol. 12, no. 4, pp. 718-732.

Bayas-Rea RD, Felix F, Montufar R 2018, 'Genetic divergence and fine scale population structure of the common bottlenose dolphin (*Tursiops truncatus*, Montagu) found in the Gulf of Guayaquil, Ecuador', PeerJ.

Beaumont MA, Nichols RA 1996, 'Evaluating loci for use in the genetic analysis of population structure', Proccedings of the Royal Society of London B, vol. 263, no. 1377, pp. 1619-1626.

Belperio A, Harvey N, Bourman R 2002, 'Spatial and temporal variability in the Holocene sea-level record of the South Australian coastline', Sedimentary Geology, vol. 150, no. 1, pp. 153-169.

Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, Rochette R, Bernatchez L 2016, 'Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*)', Molecular Ecology, vol. 25, no. 20, pp. 5073-5092.

Benjamini Y, Yekutieli D 2001, 'The control of the false discovery rate in multiple testing under dependency', Annals of Statistics, vol. 29, no. 4, pp. 1165-1188.

Berger W 2007, 'Cenozoic cooling, Antarctic nutrient pump, and the evolution of whales', Deep Sea Research Part II: Topical Studies in Oceanography, vol. 54, no. 21, pp. 2399-2421.

Bernal-Ramírez J, Adcock G, Hauser L, Carvalho G, Smith P 2003, 'Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents', Marine Biology, vol. 142, no. 3, pp. 567-574.

Bernatchez L 2016, 'On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes', Journal of Fish Biology, vol. 89, no. 6, pp. 2519-2556.

Bernatchez S, Xuereb A, Laporte M, Benestan L, Steeves R, Laflamme M, Bernatchez L, Mallet MA 2019, 'Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada', Evolutionary Applications, vol. 12, no. 3, pp. 587-609.

Bilgmann K, Griffiths O, Allen S, Möller L 2007a, 'A biopsy pole system for bow-riding dolphins: sampling success, behavioral responses, and test for sampling bias', Marine Mammal Science, vol. 23, no. 1, pp. 218-225.

Bilgmann K, Möller L, Harcourt R, Gibbs S, Beheregaray L 2007b, 'Genetic differentiation in bottlenose dolphins from South Australia: Association with local oceanography and coastal geography', Marine Ecology Progress Series, vol. 341, no. 1, pp. 265–276.

Bilgmann K, Möller LM, Harcourt RG, Gales R, Beheregaray LB 2008, 'Common dolphins subject to fisheries impacts in southern Australia are genetically differentiated: Implications for conservation', Animal Conservation, vol. 11, no. 6, pp. 518-528.

Bilgmann K, Parra G, Zanardo N, Beheregaray L, Möller L 2014, 'Multiple management units of short-beaked common dolphins subject to fisheries bycatch off southern and southeastern Australia', Marine Ecology Progress Series, vol. 500, no. 1, pp. 265-279.

Bilgmann K, Parra GJ, Holmes L, Peters KJ, Jonsen ID, Möller LM 2019, 'Abundance estimates and habitat preferences of bottlenose dolphins reveal the importance of two gulfs in South Australia', Scientific Reports, vol. 9, no. 1, pp. 8044.

Bilgmann K, Parra GJ, Möller LM 2018, 'Occurrence, distribution and abundance of cetaceans off the western Eyre Peninsula in the Great Australian Bight', Deep Sea Research Part II: Topical Studies in Oceanography, vol. 157, pp. 134-145.

Bjørnstad ON, Grenfell BT 2001, 'Noisy clockwork: time series analysis of population fluctuations in animals', Science, vol. 293, no. 5530, pp. 638-643.

Bolger AM, Lohse M, Usadel B 2014, 'Trimmomatic: a flexible trimmer for Illumina sequence data', Bioinformatics, vol. 30, no. 15, pp. 2114-2120.

Bolz H, von Brederlow B, Ramírez A, Bryda EC, Kutsche K, Nothwang HG, Seeliger M, Cabrera MdC-S, Vila MC, Molina OP 2001, 'Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D', Nature Genetics, vol. 27, no. 1, pp. 108.

Böning CW, Dispert A, Visbeck M, Rintoul S, Schwarzkopf FU 2008, 'The response of the Antarctic Circumpolar Current to recent climate change', Nature Geoscience, vol. 1, no. 12, pp. 864.

Bossley MI, Steiner A, Rankin RW, Bejder L 2017, 'A long-term study of bottlenose dolphins (*Tursiops aduncus*) in an Australian industrial estuary: increased sightings associated with environmental improvements', Marine Mammal Science, vol. 33, no. 1, pp. 277-290.

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A 2007, 'UniProtKB/Swiss-Prot', Methods in Molecular Biology, vol. 406, pp. 89-112.

Braasch I, Brunet F, Volff J-N, Schartl M 2009, 'Pigmentation pathway evolution after whole-genome duplication in fish', Genome Biology and Evolution, vol. 1, pp. 479-493.

Brandt SB 1983, 'Pelagic squid associations with a warm-core eddy of the East Australian Current', Marine and Freshwater Research, vol. 34, no. 4, pp. 573-585.

Brauer CJ, Hammer MP, Beheregaray LB 2016, 'Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin', Molecular Ecology, vol. 25, no. 20, pp. 5093-5113.

Breiman L 2001, 'Random Forests', Machine Learning, vol. 45, no. 1, pp. 5-32.

Brieuc MS, Waters CD, Drinan DP, Naish KA 2018, 'A practical introduction to Random Forest for genetic association studies in ecology and evolution', Molecular Ecology Resources, vol. 18, pp. 755-766.

Briggs J (1974) Marine Biogeography. McGraw-Hill, New York.

Briscoe J, Sussel L, Serup P, Hartigan-O'Connor D, Jessell T, Rubenstein J, Ericson J 1999, 'Homeobox gene Nkx2. 2 and specification of neuronal identity by graded Sonic hedgehog signalling', Nature, vol. 398, no. 6728, pp. 622.

Browne M, Tyrrell A, Chapman C, Carey J, Glover D, Grosveld F, Dodd I, Robinson J 1985, 'Isolation of a human tissue-type plasminogen-activator genomic DNA clone and its expression in mouse L cells', Gene, vol. 33, no. 3, pp. 279-284.

Bullock D 1975, 'The general water circulation of Spencer Gulf, South Australia, in the period February to May', Transactions of the Royal Society of South Australia, vol. 99, no. 1, pp. 43-53.

Caballero S, Duchêne S, Garavito MF, Slikas B, Baker CS 2015, 'Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes', PLoS One, vol. 10, no. 5, pp. e0123543.

Caballero S, Trujillo F, Vianna J, Barrios-Garrido H, Montiel M, Beltrán-Pedreros S, Marmontel M, Santos M, Rossi-Santos M, Santos F 2007, 'Taxonomic status of the genus Sotalia: species level ranking for “tucuxi”(*Sotalia fluviatilis*) and “costero”(*Sotalia guianensis*) dolphins', Marine Mammal Science, vol. 23, no. 2, pp. 358-386.

Cai W, Shi G, Cowan T, Bi D, Ribbe J 2005, 'The response of the Southern Annular Mode, the East Australian Current, and the southern mid-latitude ocean circulation to global warming', Geophysical Research Letters, vol. 32, no. 23.

Caldas H, Herman GE 2003, 'NSDHL, an enzyme involved in cholesterol biosynthesis, traffics through the Golgi and accumulates on ER membranes and on the surface of lipid droplets', Human Molecular Genetics, vol. 12, no. 22, pp. 2981-2991.

Cammen K, Andrews K, Carroll E, Foote A, Humble E, Khudyakov J, Louis M, McGowen M, Olsen M, Van Cise A 2016, 'Genomic methods take the plunge: Recent advances in high-throughput sequencing of marine mammals', Journal of Heredity, vol. 107, no. 6, pp. 481-495.

Cammen K, Schultz T, Rosel P, Wells R, Read A 2015a, 'Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*)', Molecular Ecology, vol. 24, no. 18, pp. 4697-4710.

Cammen KM, Wilcox LA, Rosel PE, Wells RS, Read AJ 2015b, 'From genome-wide to candidate gene: An investigation of variation at the major histocompatibility complex in common bottlenose dolphins exposed to harmful algal blooms', Immunogenetics, vol. 67, no. 2, pp. 125-133.
Cannon B, Nedergaard J 2004, 'Brown adipose tissue: function and physiological significance', Physiological Reviews, vol. 84, no. 1, pp. 277-359.
Carling D 2004, 'The AMP-activated protein kinase cascade–a unifying system for energy control', Trends in Biochemical Sciences, vol. 29, no. 1, pp. 18-24.
Carr M-E, Kearns EJ 2003, 'Production regimes in four Eastern Boundary Current systems', Deep Sea Research Part II: Topical Studies in Oceanography, vol. 50, no. 22-26, pp. 3199-3221.
Cassens I, Vicario S, Waddell VG, Balchowsky H, Van Belle D, Ding W, Fan C, Mohan RL, Simões-Lopes PC, Bastida R 2000, 'Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages', Proceedings of the National Academy of Sciences, vol. 97, no. 21, pp. 11343-11347.
Catón J, Luder H-U, Zoupa M, Bradman M, Bluteau G, Tucker AS, Klein O, Mitsiadis TA 2009, 'Enamel-free teeth: Tbx1 deletion affects amelogenesis in rodent incisors', Developmental Biology, vol. 328, no. 2, pp. 493-505.
Chantranupong L, Wolfson RL, Sabatini DM 2015, 'Nutrient-sensing mechanisms across evolution', Cell, vol. 161, no. 1, pp. 67-83.
Charlesworth B, Lande R, Slatkin M 1982, 'A neo-Darwinian commentary on macroevolution', Evolution, vol. 36, no. 3, pp. 474-498.
Charlesworth D 2006, 'Balancing selection and its effects on sequences in nearby genome regions', PLoS Genetics, vol. 2, no. 4, pp. e64.
Charlton-Robb K, Gershwin L, Thompson R, Austin J, Owen K, McKechnie S 2011, 'A new dolphin species, the Burrunan dolphin *Tursiops australis* sp. nov., endemic to southern Australian coastal waters', PLoS One, vol. 6, no. 9, pp. e24047.
Charlton-Robb K, Taylor A, McKechnie S 2015, 'Population genetic structure of the Burrunan dolphin (*Tursiops australis*) in coastal waters of south-eastern Australia: Conservation implications', Conservation Genetics, vol. 16, no. 1, pp. 195-207.
Charlton K, Taylor A, McKechnie SW 2006, 'A note on divergent mtDNA lineages of bottlenose dolphins from coastal waters of southern Australia', Journal of Cetacean Research and Management, vol. 8, no. 2, pp. 173.
Chen X, An Y, Gao Y, Guo L, Rui L, Xie H, Sun M, Lam Hung S, Sheng X, Zou J 2017, 'Rare Deleterious PARD3 Variants in the aPKC-Binding Region are Implicated in the Pathogenesis of Human Cranial Neural Tube Defects Via Disrupting Apical Tight Junction Formation', Human Mutation, vol. 38, no. 4, pp. 378-389.
Cherniack NS, Longobardo G 1970, 'Oxygen and carbon dioxide gas stores of the body', Physiological Reviews, vol. 50, no. 2, pp. 196-243.
Chikina M, Robinson JD, Clark NL 2016, 'Hundreds of genes experienced convergent shifts in selective pressure in marine mammals', Molecular Biology and Evolution, vol. 33, no. 9, pp. 2182-2192.
Choe SS, Huh JY, Hwang IJ, Kim JI, Kim JB 2016, 'Adipose tissue remodeling: its role in energy metabolism and metabolic disorders', Frontiers in Endocrinology, vol. 7, no. 30, pp. 1-16.
Chong Z, Ruan J, Wu C-I 2012, 'Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads', Bioinformatics, vol. 28, no. 21, pp. 2732-2737.
Christiansen F, Rasmussen M, Lusseau D 2013, 'Whale watching disrupts feeding activities of minke whales on a feeding ground', Marine Ecology Progress Series, vol. 478, pp. 239-251.
Christin P-A, Weinreich DM, Besnard G 2010, 'Causes and evolutionary significance of genetic convergence', Trends in Genetics, vol. 26, no. 9, pp. 400-405.
Claesson-Welsh L, Welsh M 2013, 'VEGFA and tumour angiogenesis', Journal of Internal Medicine, vol. 273, no. 2, pp. 114-127.

Clemente FJ, Cardona A, Inchley CE, Peter BM, Jacobs G, Pagani L, Lawson DJ, Antão T, Vicente M, Mitt M 2014, 'A selective sweep on a deleterious mutation in CPT1A in Arctic populations', The American Journal of Human Genetics, vol. 95, no. 5, pp. 584-589.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM 2005, 'Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles', Science, vol. 307, no. 5717, pp. 1928-1933.

Committee on Taxonomy of the Society for Marine Mammalogy (2019) List of marine mammal species and subspecies [Online]. Society for Marine Mammalogy. Available: www.marinemammalscience.org [Accessed 15/01/2020].

Condamine FL, Rolland J, Morlon H 2013, 'Macroevolutionary perspectives to environmental change', Ecology Letters, vol. 16, pp. 72-85.

Condie S, Mansbridge J, Cahill M 2011, 'Contrasting local retention and cross-shore transports of the East Australian Current and the Leeuwin Current and their relative influences on the life histories of small pelagic fishes', Deep Sea Research Part II: Topical Studies in Oceanography, vol. 58, no. 5, pp. 606-615.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M 2005, 'Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research', Bioinformatics, vol. 21, no. 18, pp. 3674-3676.

Conner JK, Hartl DL (2004) A primer of ecological genetics. Sinauer Associates Incorporated.

Cooke GM, Chao NL, Beheregaray LB 2012, 'Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*', Molecular Ecology, vol. 21, no. 10, pp. 2410-2427.

Coolen MW, Van Loo KM, Van Bakel NN, Pulford DJ, Serneels L, De Strooper B, Ellenbroek BA, Cools AR, Martens GJ 2005, 'Gene dosage effect on γ-secretase component Aph-1b in a rat model for neurodevelopmental disorders', Neuron, vol. 45, no. 4, pp. 497-503.

Costa-Urrutia P, Abud C, Secchi ER, Lessa EP 2012, 'Population genetic structure and social kin associations of franciscana dolphin, *Pontoporia blainvillei*', Journal of Heredity, vol. 103, no. 1, pp. 92-102.

Costa AP, Fruet PF, Secchi ER, Daura-Jorge FG, Simões-Lopes PC, Di Tullio JC, Rosel P 2019, 'Ecological Divergence and Speciation in Common Bottlenose Dolphins in the Western South Atlantic', Journal of Evolutionary Biology.

Costa AP, Rosel PE, Daura-Jorge FG, Simões-Lopes PC 2016, 'Offshore and coastal common bottlenose dolphins of the western South Atlantic face-to-face: What the skull and the spine can tell us', Marine Mammal Science, vol. 32, no. 4, pp. 1433-1457.

Costa APB, Fruet P, Daura-Jorge FG, Simões-Lopes PC, Ott PH, Valiati VH, de Oliveira LR 2015, 'Bottlenose dolphin communities from the southern Brazilian coast: do they exchange genes or are they just neighbours?', Marine and Freshwater Research, vol. 66, no. 12, pp. 1201-1210.

Cowen RK, Sponaugle S 2009, 'Larval dispersal and marine population connectivity', Annual Review of Marine Science, vol. 1, pp. 443-466.

Cribb N, Miller C, Seuront L 2013, 'Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) habitat preference in a heterogeneous, urban, coastal environment', Aquatic Biosystems, vol. 9, no. 1, pp. 3.

Cunha H, Da Silva V, Lailson-Brito J, Santos MdO, Flores P, Martin A, Azevedo A, Fragoso A, Zanelatto R, Solé-Cava A 2005, 'Riverine and marine ecotypes of *Sotalia* dolphins are different species', Marine Biology, vol. 148, no. 2, pp. 449.

Currie DR, Sorokin SJ 2010, 'The distribution and trophodynamics of demersal fish from Spencer Gulf', Transactions of the Royal Society of South Australia, vol. 134, no. 2, pp. 198-227.

Curry B, Smith J 1997, 'Phylogeographic structure of the bottlenose dolphin (*Tursiops truncatus*): Stock identification and implications for management', Molecular Genetics of Marine Mammals. Special Publication, vol. 3, no. 1, pp. 327-247.

Czesny S, Epifanio J, Michalak P 2012, 'Genetic divergence between freshwater and marine morphs of alewife (*Alosa pseudoharengus*): a 'next-generation'sequencing analysis', PLoS One, vol. 7, no. 3, pp. e31803.

Daura-Jorge FG, Simões-Lopes PC 2011, 'Lobomycosis-like disease in wild bottlenose dolphins *Tursiops truncatus* of Laguna, southern Brazil: monitoring of a progressive case', Diseases of Aquatic Organisms, vol. 93, no. 2, pp. 163-170.

Daura-Jorge FG, Ingram SN, Simões-Lopes PC 2013, 'Seasonal abundance and adult survival of bottlenose dolphins (*Tursiops truncatus*) in a community that cooperatively forages with fishermen in southern Brazil', Marine Mammal Science, vol. 29, no. 2, pp. 293-311.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML 2011, 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', Nature Reviews Genetics, vol. 12, no. 7, pp. 499-510.

Davies K, Cotton J, Kirwan J, Teeling E, Rossiter S 2012, 'Parallel signatures of sequence evolution among hearing genes in echolocating mammals: An emerging model of genetic convergence', Heredity, vol. 108, no. 5, pp. 480-489.

Day V, Paxinos R, Emmett J, Wright A, Goecker M 2008, 'The Marine Planning Framework for South Australia: A new ecosystem-based zoning policy for marine management', Marine Policy, vol. 32, no. 4, pp. 535-543.

de Buffrénil V, Canoville A, D'Anastasio R, Domning DP 2010, 'Evolution of sirenian pachyosteosclerosis, a model-case for the study of bone structure in aquatic tetrapods', Journal of Mammalian Evolution, vol. 17, no. 2, pp. 101-120.

de Oliveira LR, Fraga LD, Ott PH, Siciliano S, Lopes F, Almeida R, Wickert JC, Milmann L, Danilewicz D, Emin-Lima NR, Meirelles AC, Luz V, do Nascimento LF, de Thoisy B, Tavares M, Zerbini AN, Baumgarten M, Valiati VH, Bonatto SL 2019, 'Population structure, phylogeography, and genetic diversity of the common bottlenose dolphin in the tropical and subtropical southwestern Atlantic Ocean', Journal of Mammalogy, vol. 100, no. 2, pp. 564-577.

de Silva Samarasinghe J 1998, 'Revisiting Upper Gulf St Vincent in South Australia: the salt balance and its implications', Estuarine, Coastal and Shelf Science, vol. 46, no. 1, pp. 51-63.

de Silva Samarasinghe J, Bode L, Mason L 2003, 'Modelled response of Gulf St Vincent (South Australia) to evaporation, heating and winds', Continental Shelf Research, vol. 23, no. 14-15, pp. 1285-1313.

Debey LB, Pyenson ND 2013, 'Osteological correlates and phylogenetic analysis of deep diving in living and extinct pinnipeds: What good are big eyes?', Marine Mammal Science, vol. 29, no. 1, pp. 48-83.

DeFaveri J, Shikano T, Shimada Y, Goto A, Merilä J 2011, 'Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*)', Evolution: International Journal of Organic Evolution, vol. 65, no. 6, pp. 1800-1807.

DEH (Department of Environment and Heritage) 2006, *A guide to the integrated marine and coastal regionalisation of Australia version 4.0,* no. Australian Government Data Sources: Canberra, Australia.

Derous D, Sahu J, Douglas AB, Lusseau D, Wenzel M 2019, 'Adaptations of energy metabolism in cetaceans have consequence for their response to foraging disruption', BioRxiv, vol. preprint.

Diaz-Aguirre F, Parra GJ, Passadore C, Möller L 2018, 'Kinship influences social bonds among male southern Australian bottlenose dolphins (*Tursiops* cf. *australis*)', Behavioral Ecology and Sociobiology, vol. 72, no. 12, pp. 190.

Diaz-Aguirre F, Parra GJ, Passadore C, Möller L 2019, 'Genetic relatedness delineates the social structure of southern Australian bottlenose dolphins', Behavioral Ecology, vol. 30, no. 4, pp. 948-959.

Diaz-Gamboa RE, Gendron D, Busquets-Vass G 2018, 'Isotopic niche width differentiation between common bottlenose dolphin ecotypes and sperm whales in the Gulf of California', Marine Mammal Science, vol. 34, no. 2, pp. 440-457.

Dimmlich W, Breed W, Geddes M, Ward T 2004, 'Relative importance of gulf and shelf waters for spawning and recruitment of Australian anchovy, *Engraulis australis*, in South Australia', Fisheries Oceanography, vol. 13, no. 5, pp. 310-323.

Diopere E, Vandamme SG, Hablützel PI, Cariani A, Van Houdt J, Rijnsdorp A, Tinti F, Consortium F, Volckaert FA, Maes GE 2017, 'Seascape genetics of a flatfish reveals local selection under high levels of gene flow', ICES Journal of Marine Science, vol. 75, no. 2, pp. 675-689.

do Prado FD, Vera M, Hermida M, Bouza C, Pardo BG, Vilas R, Blanco A, Fernández C, Maroso F, Maes GE 2018, 'Parallel evolution and adaptation to environmental factors in a marine flatfish: Implications for fisheries and aquaculture management of the turbot (*Scophthalmus maximus*)', Evolutionary Applications, vol. 11, no. 8, pp. 1322-1341.

Domingues CM, Church JA, White NJ, Gleckler PJ, Wijffels SE, Barker PM, Dunn JR 2008, 'Improved estimates of upper-ocean warming and multi-decadal sea-level rise', Nature, vol. 453, no. 7198, pp. 1090.

Doney SC, Ruckelshaus M, Duffy JE, Barry JP, Chan F, English CA, Galindo HM, Grebmeier JM, Hollowed AB, Knowlton N 2011, 'Climate change impacts on marine ecosystems', Annual Reviews of Marine Science, vol. 4, pp. 11-37.

Drabek CM, Burns JM 2002, 'Heart and aorta morphology of the deep-diving hooded seal (*Cystophora cristata*)', Canadian Journal of Zoology, vol. 80, no. 11, pp. 2030-2036.

DSE (Department of Sustainability and Environment) 2013, *Advisory list of threatened vertebrate fauna in Victoria,* no. State Government of Victoria: Melbourne, Australia.

Du SJ, Tan X, Zhang J 2014, 'SMYD proteins: key regulators in skeletal and cardiac muscle development and function', The Anatomical Record, vol. 297, no. 9, pp. 1650-1662.

Duffield DA, Ridgway SH, Cornell LH 1983, 'Hematology distinguishes coastal and offshore forms of dolphins (*Tursiops*)', Canadian Journal of Zoology, vol. 61, no. 4, pp. 930-933.

Dwivedi PP, Grose RH, Filmus J, Hii CS, Xian CJ, Anderson PJ, Powell BC 2013, 'Regulation of bone morphogenetic protein signalling and cranial osteogenesis by Gpc1 and Gpc3', Bone, vol. 55, no. 2, pp. 367-376.

Edyvane K 1999, 'Conserving marine biodiversity in South Australia - Part 2 - Identification of areas of high conservation value in South Australia'. *South Australian Research and Development (SARDI) Research Report Series*. SARDI, Adelaide, SA,

Edyvane K 2000, 'The Great Australian Bight'. In: Sheppard, C (ed.) *Seas at the Millennium: An Environmental Evaluation*. Elsevier Science, Amsterdam, Netherlands, pp. 673-690.

Edyvane K 2008, 'Macroalgal biogeography and assemblages of Gulf St Vincent', Field Experiments, vol. 19, pp. 248-263.

Ellegren H, Galtier N 2016, 'Determinants of genetic diversity', Nature Reviews Genetics, vol. 17, no. 7, pp. 422.

Ellstrand NC, Elam DR 1993, 'Population genetic consequences of small population size: Implications for plant conservation', Annual Review of Ecology and Systematics, vol. 24, no. 1, pp. 217-242.

Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton, New Jersey.

Endo Y, Kamei Ki, Inoue-Murayama M 2018, 'Genetic signatures of lipid metabolism evolution in Cetacea since the divergence from terrestrial ancestor', Journal of Evolutionary Biology, vol. 31, no. 11, pp. 1655-1665.

Ernst MC, Haidl ID, Zúñiga LA, Dranse HJ, Rourke JL, Zabel BA, Butcher EC, Sinal CJ 2012, 'Disruption of the chemokine-like receptor-1 (CMKLR1) gene is associated with reduced adiposity and glucose intolerance', Endocrinology, vol. 153, no. 2, pp. 672-682.

Erwin DH 2000, 'Macroevolution is more than repeated rounds of microevolution', Evolution & Development, vol. 2, no. 2, pp. 78-84.

Erwin DH 2010, 'Microevolution and macroevolution are not governed by the same processes'. In: Ayala, F & Arp, R (eds) *Contemporary debates in the philosophy of biology*. Wiley-Blackwell, United Kingdom, pp. 180-193.

Everett JD, Baird ME, Roughan M, Suthers IM, Doblin MA 2014, 'Relative impact of seasonal and oceanographic drivers on surface chlorophyll a along a Western Boundary Current', Progress in Oceanography, vol. 120, pp. 340-351.

Excoffier L, Hofer T, Foll M 2009, 'Detecting loci under selection in a hierarchically structured population', Heredity, vol. 103, no. 4, pp. 285.

Excoffier L, Lischer HE 2010, 'Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows', Molecular Ecology Resources, vol. 10, no. 3, pp. 564-567.

Falconer DS, Mackay TF, Frankham R 1996, 'Introduction to quantitative genetics (4th edn)', Trends in Genetics, vol. 12, no. 7, pp. 280.

Fekete DM, Campero AM 2007, 'Axon guidance in the inner ear', International Journal of Developmental Biology, vol. 51, no. 6-7, pp. 549-556.

Félix F, Van Waerebeek K, Sanino GP, Castro C, Van Bressem M-F, Santillán L 2018, 'Variation in dorsal fin morphology in common bottlenose dolphin *Tursiops truncatus* (Cetacea: Delphinidae) populations from the southeast Pacific Ocean', Pacific Science, vol. 72, no. 3, pp. 307-320.

Feng M, Waite A, Thompson P 2009, 'Climate variability and ocean production in the Leeuwin Current system off the west coast of Western Australia', Journal of the Royal Society of Western Australia, vol. 92, pp. 67.

Finn J, Tregenza T, Norman M 2009, 'Preparing the perfect cuttlefish meal: complex prey handling by dolphins', PLoS One, vol. 4, no. 1, pp. e4217.

Fiol DF, Mak SK, Kültz D 2007, 'Specific TSC22 domain transcripts are hypertonically induced and alternatively spliced to protect mouse kidney cells during osmotic stress', The FEBS journal, vol. 274, no. 1, pp. 109-124.

Foll M, Gaggiotti O 2008, 'A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective', Genetics, vol. 180, no. 2, pp. 977-993.

Foote A, Morin P 2016, 'Genome-wide SNP data suggest complex ancestry of sympatric North Pacific killer whale ecotypes', Heredity, vol. 117, no. 5, pp. 316-325.

Foote AD, Liu Y, Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V 2015, 'Convergent evolution of the genomes of marine mammals', Nature Genetics, vol. 47, no. 3, pp. 272-275.

Foote AD, Newton J, Piertney SB, Willerslev E, Gilbert MTP 2009, 'Ecological, morphological and genetic divergence of sympatric North Atlantic killer whale populations', Molecular Ecology, vol. 18, no. 24, pp. 5207-5217.

Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD 2016, 'Genome-culture coevolution promotes rapid divergence of killer whale ecotypes', Nature Communications, vol. 7, no. 11693, pp. 1-12.

Ford J 2009, 'Killer Whale'. In: Perrin, W, Wursig, B & Thewissen, J (eds) *Encyclopedia of Marine Mammals*. Academic Press, San Diego, pp. 650-657.

Ford JK, Ellis GM, Barrett-Lennard LG, Morton AB, Palm RS, Balcomb III KC 1998, 'Dietary specialization in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters', Canadian Journal of Zoology, vol. 76, no. 8, pp. 1456-1471.

Fordyce R 2009, 'Cetacean Evolution'. In: Perrin, W, Wursig, B & Thewissen, J (eds) *Encyclopedia of Marine Mammals*. Academic Press, San Diego, pp. 201-207.

Fordyce RE 1980, 'Whale evolution and Oligocene Southern Ocean environments', Palaeogeography, Palaeoclimatology, Palaeoecology, vol. 31, no. 1, pp. 319-336.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR 2016, 'Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes', Molecular Ecology, vol. 25, no. 1, pp. 104-120.

Forester BR, Lasky JR, Wagner HH, Urban DL 2018, 'Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations', Molecular Ecology, vol. 27, no. 9, pp. 2215-2233.

Francois H, Julie J, Sebastien L, Jeremy M 2015. Facto Mine R: Multivariate Exploratory Data Analysis and Data Mining. R package version 129.

Fruet P, Secchi E, Di Tullio J, Kinas P 2011, 'Abundance of bottlenose dolphins, *Tursiops truncatus*, inhabiting the Patos Lagoon estuary, southern Brazil: implications for conservation', Zoologia, vol. 28, pp. 23-30.

Fruet P, Secchi E, Di Tullio J, Simoes-Lopes P, Daura-Jorge F, Paula Costa A, Vermeulen E, Flores P, Genoves RC, Laporta P, Beheregaray L, Möller L 2014a, 'Strong genetic differentiation between coastal and offshore ecotypes of common bottlenose dolphins (*Tursiops truncatus*) in the Southwest Atlantic Ocean', International Whaling Commission Report, pp. 1-18.

Fruet PF, Kinas PG, da Silva KG, Di Tullio JC, Monteiro DS, Dalla Rosa L, Estima SC, Secchi ER 2012, 'Temporal trends in mortality and effects of by-catch on common bottlenose dolphins, *Tursiops truncatus*, in southern Brazil', Journal of the Marine Biological Association of the United Kingdom, vol. 92, no. 8, pp. 1865-1876.

Fruet PF, Secchi ER, Daura-Jorge F, Vermeulen E, Flores PA, Simoes-Lopes PC, Genoves RC, Laporta P, Di Tullio JC, Freitas TRO 2014b, 'Remarkably low genetic diversity and strong population structure in common bottlenose dolphins (*Tursiops truncatus*) from coastal waters of the Southwestern Atlantic Ocean', Conservation Genetics, vol. 15, no. 4, pp. 879-895.

Fruet PF, Secchi ER, Di Tullio JC, Simões-Lopes PC, Daura-Jorge F, Costa APB, Vermeulen E, Flores PAC, Genoves RC, Laporta P, Beheregaray LB, Möller LM 2017, 'Genetic divergence between two phenotypically distinct bottlenose dolphin ecotypes suggests separate evolutionary trajectories', Ecology and Evolution, vol. 7, no. 21, pp. 9131-9143.

Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A 2015, 'Greenlandic Inuit show genetic signatures of diet and climate adaptation', Science, vol. 349, no. 6254, pp. 1343-1347.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW 2012, 'Harnessing genomics for delineating conservation units', Trends in Ecology & Evolution, vol. 27, no. 9, pp. 489-496.

Gabler MK, Gay DM, Westgate AJ, Koopman HN 2018, 'Microvascular characteristics of the acoustic fats: Novel data suggesting taxonomic differences between deep and shallow-diving odontocetes', Journal of Morphology, vol. 279, no. 4, pp. 458-471.

Gales N, Hindell M, Kirkwood R (2003) Marine mammals: fisheries, tourism and management issues. CSIRO Publishing, Victoria, Australia.

Galpern P, Peres-Neto PR, Polfus J, Manseau M 2014, 'MEMGENE: Spatial pattern detection in genetic distance data', Methods in Ecology and Evolution, vol. 5, no. 10, pp. 1116-1120.

Garrison E, Marth G 2012, 'Haplotype-based variant detection from short-read sequencing', arXiv preprint arXiv:1207.3907, pp. 1-9.

Gaspari S, Scheinin A, Holcer D, Fortuna C, Natali C, Genov T, Frantzis A, Chelazzi G, Moura AE 2015, 'Drivers of population structure of the bottlenose dolphin (*Tursiops truncatus*) in the eastern Mediterranean Sea', Evolutionary Biology, vol. 42, no. 2, pp. 177-190.

Gatesy J, Geisler JH, Chang J, Buell C, Berta A, Meredith RW, Springer MS, McGowen MR 2013, 'A phylogenetic blueprint for a modern whale', Molecular Phylogenetics and Evolution, vol. 66, no. 2, pp. 479-506.

Gaughran SJ, Quinzin MC, Miller JM, Garrick RC, Edwards DL, Russello MA, Poulakakis N, Ciofi C, Beheregaray LB, Caccone A 2018, 'Theory, practice, and conservation in the age of genomics: The Galápagos giant tortoise as a case study', Evolutionary Applications, vol. 11, no. 7, pp. 1084-1093.

Gibbs S, Harcourt R, Kemper C 2011, 'Niche differentiation of bottlenose dolphin species in South Australia revealed by stable isotopes and stomach contents', Wildlife Research, vol. 38, no. 4, pp. 261-270.

Gibbs S, Kemper CM 2014, 'Whales and dolphins in Spencer Gulf'. In: Shepard, S, Madigan, S, Gillanders, B, Murray-Jones, S & Wiltshire, D (eds) *Natural History of Spencer Gulf*. Royal Society of South Australia, Adelaide, pp. 242-253.

Gienapp P, Teplitsky C, Alho J, Mills J, Merilä J 2008, 'Climate change and evolution: disentangling environmental and genetic responses', Molecular Ecology, vol. 17, no. 1, pp. 167-178.

Gillanders B, Donnellan SC, Prowse TA, Fordham DA, Izzo C, Myers SA, Rowling KP, Steer MA, Woodcock SH (2016) Giant Australian cuttlefish in South Australian waters. Fisheries Research and Development Corporation. University of Adelaide, South Australia.

Gillanders B, Payne N 2014, 'Giant Australian cuttlefish'. In: Shepard, S, Madigan, S, Gillanders, B, Murray-Jones, S & Wiltshire, D (eds) *Natural History of Spencer Gulf*. Royal Society of South Australia, Adelaide, pp. 288-301.

Gingerich PD, von Koenigswald W, Sanders WJ, Smith BH, Zalmout IS 2009, 'New protocetid whale from the middle Eocene of Pakistan: birth on land, precocial development, and sexual dimorphism', PLoS One, vol. 4, no. 2, pp. e4366.

Goldbogen JA, Hazen EL, Friedlaender AS, Calambokidis J, DeRuiter SL, Stimpert AK, Southall BL 2015, 'Prey density and distribution drive the three-dimensional foraging strategies of the largest filter feeder', Functional Ecology, vol. 29, no. 7, pp. 951-961.

Gray H, Nishida S, Welch A, Moura A, Tanabe S, Kiani M, Culloch R, Möller L, Natoli A, Ponnampalam L 2018, 'Cryptic lineage differentiation among Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in the northwest Indian Ocean', Molecular Phylogenetics and Evolution, vol. 122, pp. 1-14.

Gregory J, Dixon K, Stouffer R, Weaver A, Driesschaert E, Eby M, Fichefet T, Hasumi H, Hu A, Jungclaus J 2005, 'A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO2 concentration', Geophysical Research Letters, vol. 32, no. 12.

Grisaru S, Cano-Gauci D, Tee J, Filmus J, Rosenblum ND 2001, 'Glypican-3 modulates BMP- and FGF-mediated effects during renal branching morphogenesis', Developmental Biology, vol. 231, no. 1, pp. 31-46.

Groen JL, Andrade A, Ritz K, Jalalzadeh H, Haagmans M, Bradley TE, Jongejan A, Verbeek DS, Nürnberg P, Denome S 2014, 'CACNA1B mutation is linked to unique myoclonus-dystonia syndrome', Human Molecular Genetics, vol. 24, no. 4, pp. 987-993.

Grummer J, Beheregaray L, Bernatchez L, Hand B, Luikart G, Narum S, Taylor A 2019, 'Aquatic landscape genomics and environmental effects on genetic variation', Trends in Ecology & Evolution, vol. 34, no. 7, pp. 641-654.

Guard SE, Poss ZC, Ebmeier CC, Pagratis M, Simpson H, Taatjes DJ, Old WM 2019, 'The nuclear interactome of DYRK1A reveals a functional role in DNA damage repair', Scientific Reports, vol. 9, no. 1, pp. 1-12.

Gui D, Jia KT, Xia J, Yang LL, Chen JL, Wu YP, Yi MS 2013, '*De novo* assembly of the Indo-Pacific humpback dolphin leucocyte transcriptome to identify putative genes involved in the aquatic adaptation and immune response', PLoS One, vol. 8, no. 8, pp. 9.

Guo D-c, Regalado E, Casteel DE, Santos-Cortez RL, Gong L, Kim JJ, Dyack S, Horne SG, Chang G, Jondeau G 2013, 'Recurrent gain-of-function mutation in PRKG1 causes thoracic aortic aneurysms and acute aortic dissections', The American Journal of Human Genetics, vol. 93, no. 2, pp. 398-404.

Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, Maatz H, Schulz H, Li S, Parrish AM 2012, 'RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing', Nature Medicine, vol. 18, no. 5, pp. 766.

Gutstein CS, Figueroa-Bravo CP, Pyenson ND, Yury-Yañez RE, Cozzuol MA, Canals M 2014, 'High frequency echolocation, ear morphology, and the marine–freshwater transition:

A comparative study of extant and extinct toothed whales', Palaeogeography, Palaeoclimatology, Palaeoecology, vol. 400, pp. 62-74.
Hagen W, Kattner G, Friedrich C 2000, 'The lipid compositions of high-Antarctic notothenioid fish species with different life strategies', Polar Biology, vol. 23, no. 11, pp. 785-791.
Hale P, Barreto A, Ross G 2000, 'Comparative morphology and distribution of the *aduncus* and *truncatus* forms of bottlenose dolphin *Tursiops* in the Indian and Western Pacific Oceans', Aquatic Mammals, vol. 26, no. 2, pp. 101-110.
Hallegraeff G, Jeffrey S 1993, 'Annually recurrent diatom blooms in spring along the New South Wales coast of Australia', Marine and Freshwater Research, vol. 44, no. 2, pp. 325-334.
Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE 2008, 'A global map of human impact on marine ecosystems', Science, vol. 319, no. 5865, pp. 948-952.
Han P, Werber J, Surana M, Fleischer N, Michaeli T 1999, 'The calcium/calmodulin-dependent phosphodiesterase PDE1C down-regulates glucose-induced insulin secretion', Journal of Biological Chemistry, vol. 274, no. 32, pp. 22337-22344.
Han X, Yoshizaki K, Miyazaki K, Arai C, Funada K, Yuta T, Tian T, Chiba Y, Saito K, Iwamoto T 2018, 'The transcription factor NKX2-3 mediates p21 expression and ectodysplasin-A signaling in the enamel knot for cusp formation in tooth development', Journal of Biological Chemistry, vol. 293, no. 38, pp. 14572-14584.
Hansen TF, Martins EP 1996, 'Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data', Evolution, vol. 50, no. 4, pp. 1404-1417.
Harvey N 2006, 'Holocene coastal evolution: Barriers, beach ridges, and tidal flats of South Australia', Journal of Coastal Research, vol. 22, no. 1, pp. 90-99.
Hastings A, Harrison S 1994, 'Metapopulation dynamics and genetics', Annual Review of Ecology and Systematics, vol. 25, no. 1, pp. 167-188.
Heard-Costa NL, Zillikens MC, Monda KL, Johansson Å, Harris TB, Fu M, Haritunians T, Feitosa MF, Aspelund T, Eiriksdottir G 2009, 'NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium', PLoS Genetics, vol. 5, no. 6, pp. e1000539.
Hedrick PW 2007, 'Balancing selection', Current Biology, vol. 17, no. 7, pp. R230-R231.
Heger J, Bornbaum J, Würfel A, Hill C, Brockmann N, Gáspár R, Pálóczi J, Varga ZV, Sárközy M, Bencsik P 2018, 'JDP2 overexpression provokes cardiac dysfunction in mice', Scientific Reports, vol. 8, no. 1, pp. 7647.
Heimlich-Boran JR 1988, 'Behavioral ecology of killer whales (*Orcinus orca*) in the Pacific Northwest', Canadian Journal of Zoology, vol. 66, no. 3, pp. 565-578.
Hein J, Schierup M, Wiuf C (2004) Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, USA.
Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V 2007, 'Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.)', Molecular Ecology, vol. 16, no. 15, pp. 3104-3118.
Hersh SL, Duffield DA 1990, 'Distinction between northwest Atlantic offshore and coastal bottlenose dolphins based on hemoglobin profile and morphometry'. In: Leatherwood, S & Reeves, RR (eds) *The Bottlenose Dolphin*. Academic Press, San Diego, California, pp. 129-139.
Hershkovitz P 1966, 'A catalogue of living whales', Bulletin of the United States National Museum, vol. 246, pp. 1-259.
Hindell MA, Slip DJ, Burton HR, Bryden MM 1992, 'Physiological implications of continuous, prolonged, and deep dives of the southern elephant seal (*Mirounga leonina*)', Canadian Journal of Zoology, vol. 70, no. 2, pp. 370-379.
Hoelzel A 2009, 'Evolution of population genetic structure in marine mammal species'. In: Bertorelle, G, Bruford, M, Hauffe, H, Rizzoli, A & Vernesi, C (eds) *Population genetics for animal conservation*. Cambridge University Press, Cambridge, UK, pp. 294-318.

Hoelzel A, Moura A 2015, 'Resource specialisation and the divergence of killer whale populations', Heredity, vol. 115, no. 1, pp. 93.

Hoelzel A, Potter C, Best P 1998, 'Genetic differentiation between parapatric 'nearshore' and 'offshore' populations of the bottlenose dolphins', Proceedings of the Royal Society B, vol. 265, no. 1402, pp. 1177-1183.

Hoelzel AR 1998, 'Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: implications for conservation policy', Journal of Heredity, vol. 89, pp. 451-458.

Hoelzel AR, Bruford MW, Fleischer RC 2019, 'Conservation of adaptive potential and functional diversity', Conservation Genetics, vol. 20, pp. 1-5.

Hoffman JI, Clarke A, Clark MS, Fretwell P, Peck LS 2012, 'Unexpected fine-scale population structure in a broadcast-spawning Antarctic marine mollusc', PLoS One, vol. 7, no. 3, pp. e32415.

Hoffmann AA, Sgro CM 2011, 'Climate change and evolutionary adaptation', Nature, vol. 470, no. 7335, pp. 479-485.

Holderegger R, Kamm U, Gugerli F 2006, 'Adaptive vs. neutral genetic diversity: Implications for landscape genetics', Landscape Ecology, vol. 21, no. 6, pp. 797-807.

Holderegger R, Wagner HH 2006, 'A brief guide to landscape genetics', Landscape Ecology, vol. 21, no. 6, pp. 793-796.

Holme RH, Steel KP 2004, 'Progressive hearing loss and increased susceptibility to noise-induced hearing loss in mice carrying a Cdh23 but not a Myo7a mutation', Journal of the Association for Research in Otolaryngology, vol. 5, no. 1, pp. 66-79.

Horrocks LA, Yeo YK 1999, 'Health benefits of docosahexaenoic acid (DHA)', Pharmacological Research, vol. 40, no. 3, pp. 211-225.

Hoskin M 2000, 'Effects of the East Australian Current on the genetic structure of a direct developing muricid snail (*Bedeva hanleyi*, Angas): variability within and among local populations', Biological Journal of the Linnean Society, vol. 69, no. 2, pp. 245-262.

Howarth GS, Bastian SE, Dunbar AJ, Goddard C 2003, 'Betacellulin promotes growth of the gastrointestinal organs and effects a diuresis in normal rats', Growth Factors, vol. 21, no. 2, pp. 79-86.

Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TF 2016, 'Spontaneous mutations and the origin and maintenance of quantitative genetic variation', Elife, vol. 5, no. e14625, pp. 1-25.

Huelsmann M, Hecker N, Springer MS, Gatesy J, Sharma V, Hiller M 2019, 'Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations', Science Advances, vol. 5, no. 9, pp. 12.

Hume ID 2002, 'Digestive strategies of mammals', Acta Zoologica Sinica, vol. 48, no. 1, pp. 1-19.

Imaseki I, Wakabayashi M, Hara Y, Watanabe T, Takabe S, Kakumura K, Honda Y, Ueda K, Murakumo K, Matsumoto R 2019, 'Comprehensive analysis of genes contributing to euryhalinity in the bull shark, *Carcharhinus leucas*; Na+-Cl− co-transporter is one of the key renal factors upregulated in acclimation to low-salinity environment', Journal of Experimental Biology, vol. 222, no. 12, pp. jeb201780.

IMCRA Technical Group (Interim Marine and Coastal Regionalisation for Australia Technical Group) 1998, *Interim Marine and Coastal Regionalisation for Australia: An ecosystem-based classification for marine and coastal environments version 3.3,* no. Environment Australia, Department of the Environment: Canberra, Australia.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A 2019, 'A key metabolic gene for recurrent freshwater colonization and radiation in fishes', Science, vol. 364, no. 6443, pp. 886-889.

Isler K, Van Schaik CP 2006, 'Metabolic costs of brain size evolution', Biology Letters, vol. 2, no. 4, pp. 557-560.

Ismail HM 2005, 'The role of omega-3 fatty acids in cardiac protection: an overview', Frontiers Bioscience, vol. 10, pp. 1079-88.

Jablonski D 2000, 'Micro-and macroevolution: Scale and hierarchy in evolutionary biology and paleobiology', Paleobiology, vol. 26, no. sp4, pp. 15-52.
Jamniczky HA, Le A, Barry TN, Rogers SM 2018, 'Freshwater influence is associated with differences in bone mineral density and armour configuration in threespine stickleback (*Gasterosteus aculeatus*)', FACETS, vol. 3, no. 1, pp. 665-681.
Jayasankar P, Anoop B, Vivekanandan, Rajagopalan M, Yousuf KMM, Reynold P, Krishnakumar PK, Kumaran PL, Afsal VV, Krishnan AA 2008, 'Molecular identification of delphinids and finless porpoise (Cetacea) from the Arabian Sea and Bay of Bengal', Zootaxa, no. 1853, pp. 57-67.
Jefferson TA, Wang JY 2011, 'Revision of the taxonomy of finless porpoises (genus Neophocaena): the existence of two species', Journal of Marine Animals and Their Ecology, vol. 4, no. 1, pp. 3-16.
Jefferson TA, Webber MA, Pitman RL (2011) Marine mammals of the world: A comprehensive guide to their identification. First edn. Elsevier, London, UK.
Jia KT, Lin WZ, Gui D, Karczmarski L, Wu YP 2014, 'Molecular evidence reveals the distinctiveness of Indo-Pacific finless porpoises (*Neophocaena phocaenoides*) in the Pearl River Estuary and insights into genus *Neophocaena*'s origin', Marine Biology, vol. 161, no. 8, pp. 1919-1930.
Johnson P, Elsner R, Zenteno-Savín T 2004, 'Hypoxia-inducible factor in ringed seal (*Phoca hispida*) tissues', Free Radical Research, vol. 38, no. 8, pp. 847-854.
Jombart T 2008, 'adegenet: a R package for the multivariate analysis of genetic markers', Bioinformatics, vol. 24, no. 11, pp. 1403-1405.
Jombart T, Ahmed I 2011, 'Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data', Bioinformatics, vol. 27, no. 21, pp. 3070-3071.
Jombart T, Devillard S, Balloux F 2010, 'Discriminant analysis of principal components: A new method for the analysis of genetically structured populations', BMC Genetics, vol. 11, no. 1, pp. 94.
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S 2012, 'The genomic basis of adaptive evolution in threespine sticklebacks', Nature, vol. 484, no. 7392, pp. 55-61.
Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, Taberlet P 2007, 'A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation', Molecular Ecology, vol. 16, no. 18, pp. 3955-3969.
Ju M, Wray D 2002, 'Molecular identification and characterisation of the human eag2 potassium channel', FEBS Letters, vol. 524, no. 1-3, pp. 204-210.
Kämpf J, Brokensha C, Bolton T 2009, 'Hindcasts of the fate of desalination brine in large inverse estuaries: Spencer Gulf and Gulf St. Vincent, South Australia', Desalination and Water Treatment, vol. 2, no. 1-3, pp. 335-344.
Kämpf J, Doubell M, Griffin D, Matthews RL, Ward TM 2004, 'Evidence of a large seasonal coastal upwelling system along the southern shelf of Australia', Geophysical Research Letters, vol. 31, no. 9.
Kämpf J, Ellis H 2015, 'Hydrodynamics and flushing of Coffin Bay, South Australia: a small tidal inverse estuary of interconnected bays', Journal of Coastal Research, vol. 31, no. 2, pp. 447-456.
Kämpf J, Payne N, Malthouse P 2010, 'Marine connectivity in a large inverse estuary', Journal of Coastal Research, vol. 26, no. 6, pp. 1047-1056.
Keane JP, Neira FJ 2008, 'Larval fish assemblages along the south-eastern Australian shelf: Linking mesoscale non-depth-discriminate structure and water masses', Fisheries Oceanography, vol. 17, no. 4, pp. 263-280.
Kelley JL, Brown AP, Therkildsen NO, Foote AD 2016, 'The life aquatic: Advances in marine vertebrate genomics', Nature Reviews Genetics, vol. 17, no. 9, pp. 523-534.
Kelly RP, Palumbi SR 2010, 'Genetic structure among 50 species of the northeastern Pacific rocky intertidal community', PLoS One, vol. 5, no. 1, pp. e8594.

Kemper CM 2004, 'Osteological variation and taxonomic affinities of bottlenose dolphins, *Tursiops* spp., from South Australia', Australian Journal of Zoology, vol. 52, no. 1, pp. 29-48.

Kienle SS, Berta A 2016, 'The better to eat you with: the comparative feeding morphology of phocid seals (Pinnipedia, Phocidae)', Journal of Anatomy, vol. 228, no. 3, pp. 396-413.

Kishida T, Thewissen J, Hayakawa T, Imai H, Agata K 2015, 'Aquatic adaptation and the evolution of smell and taste in whales', Zoological Letters.

Kooyman G 2009, 'Diving Physiology'. In: Perrin, W, Wursig, B & Thewissen, J (eds) *Encyclopedia of Marine Mammals*. Academic Press, San Diego, pp. 327-332.

Kooyman G, Ponganis P 1998, 'The physiological basis of diving to depth: Birds and mammals', Annual Review of Physiology, vol. 60, no. 1, pp. 19-32.

Kopp D, Lefebvre S, Cachera M, Villanueva MC, Ernande B 2015, 'Reorganization of a marine trophic network along an inshore–offshore gradient due to stronger pelagic–benthic coupling in coastal areas', Progress in Oceanography, vol. 130, pp. 157-171.

Kozak GM, Brennan RS, Berdan EL, Fuller RC, Whitehead A 2014, 'Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches', Evolution, vol. 68, no. 1, pp. 63-80.

Kroker KS, Mathis C, Marti A, Cassel J-C, Rosenbrock H, Dorner-Ciossek C 2014, 'PDE9A inhibition rescues amyloid beta-induced deficits in synaptic plasticity and cognition', Neurobiology of Aging, vol. 35, no. 9, pp. 2072-2078.

Krützen M, Barre L, Möller L, Heithaus M, Simms C, Sherwin W 2002, 'A biopsy system for small cetaceans: Darting success and wound healing in *Tursiops* spp.', Marine Mammal Science, vol. 18, no. 4, pp. 863-878.

Krützen M, Mann J, Heithaus MR, Connor RC, Bejder L, Sherwin WB 2005, 'Cultural transmission of tool use in bottlenose dolphins', Proceedings of the National Academy of Sciences, vol. 102, no. 25, pp. 8939-8943.

Kuang S-Q, Kwartler CS, Byanova KL, Pham J, Gong L, Prakash SK, Huang J, Kamm KE, Stull JT, Sweeney HL 2012, 'Rare, nonsynonymous variant in the smooth muscle-specific isoform of myosin heavy chain, MYH11, R247C, alters force generation in the aorta and phenotype of smooth muscle cells', Circulation Research, vol. 110, no. 11, pp. 1411-1422.

La Manna G, Rako-Gòspic N, Manghi M, Ceccherelli G 2019, 'Influence of environmental, social and behavioural variables on the whistling of the common bottlenose dolphin (*Tursiops truncatus*)', Behavioral Ecology and Sociobiology, vol. 73, no. 9, pp. 121.

Lacy RC 1987, 'Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision', Conservation Biology, vol. 1, no. 2, pp. 143-158.

Lande R 1995, 'Mutation and Conservation', Conservation Biology, vol. 9, no. 4, pp. 782-791.

Lande R, Shannon S 1996, 'The role of genetic variation in adaptation and population persistence in a changing environment', Evolution, vol. 50, no. 1, pp. 434-437.

Latta RG 2004, 'Gene flow, adaptive population divergence and comparative population structure across loci', New Phytologist, vol. 161, no. 1, pp. 51-58.

Lau P, Fitzsimmons RL, Raichur S, Wang S-CM, Lechtken A, Muscat GEO 2008, 'The orphan nuclear receptor, RORA, regulates gene expression that controls lipid metabolism', Journal of Biological Chemistry, vol. 283, no. 26, pp. 18411-18421.

Le Moan A, Gagnaire PA, Bonhomme F 2016, 'Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact', Molecular Ecology, vol. 25, no. 13, pp. 3187-3202.

LeDuc R 2009, 'Delphinids, Overview'. In: Perrin, W, Wursig, B & Thewissen, J (eds) *Encyclopedia of Marine Mammals*. Academic Press, San Diego, pp. 298-302.

LeDuc R, Perrin W, Dizon A 1999, 'Phylogenetic relationships among the delphinid cetaceans based on full cytochrome b sequences', Marine Mammal Science, vol. 15, no. 3, pp. 619-648.

LeDuc RG, Robertson KM, Pitman RL 2008, 'Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species', Biology Letters, vol. 4, no. 4, pp. 426-429.
Lee DI, Zhu G, Sasaki T, Cho G-S, Hamdani N, Holewinski R, Jo S-H, Danner T, Zhang M, Rainer PP 2015, 'Phosphodiesterase 9A controls nitric-oxide-independent cGMP and hypertrophic heart disease', Nature, vol. 519, no. 7544, pp. 472.
Lee R, Pritchard T, Ajani P, Black K 2007, 'The influence of the East Australian Current eddy field on phytoplankton dynamics in the coastal zone', Journal of Coastal Research, no. 50, pp. 576-584.
Leon F, Galvan-Tirado C, Velasco LS, Silva-Segundo CA, Hernandez-Guzman R, Barriga-Sosa ID, Jaimes PD, Canino M, Cruz-Hernandez P 2018, 'Role of oceanography in shaping the genetic structure in the North Pacific hake *Merluccius productus*', PLoS One, vol. 13, no. 3, pp. 26.
Levenson DH, Schusterman RJ 1999, 'Dark adaptation and visual sensitivity in shallow and deep-diving pinnipeds', Marine Mammal Science, vol. 15, no. 4, pp. 1303-1313.
Li H 2013, 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', Oxford University Press.
Li H, Durbin R 2010, 'Fast and accurate long-read alignment with Burrows–Wheeler transform', Bioinformatics, vol. 26, no. 5, pp. 589-595.
Li JC, Foighil DO, Park JK 2013, 'Triton's trident: cryptic Neogene divergences in a marine clam (Lasaea australis) correspond to Australia's three temperate biogeographic provinces', Molecular Ecology, vol. 22, no. 7, pp. 1933-1946.
Li W, Jaroszewski L, Godzik A 2001, 'Clustering of highly homologous sequences to reduce the size of large protein databases', Bioinformatics, vol. 17, no. 3, pp. 282-283.
Li Y-C, Yang C-S, Zhou W-L, Li H-S, Han Y-J, Wang Q-S, Wu H-B 2018, 'Low glucose metabolism in hepatocellular carcinoma with GPC3 expression', World Journal of Gastroenterology, vol. 24, no. 4, pp. 494.
Li Y, Liu Z, Shi P, Zhang J 2010, 'The hearing gene Prestin unites echolocating bats and whales', Current Biology, vol. 20, no. 2, pp. R55-R56.
Li YG, Lasar D, Fromme T, Klingenspor M 2014, 'White, brite, and brown adipocytes: the evolution and function of a heater organ in mammals', Canadian Journal of Zoology, vol. 92, no. 7, pp. 615-626.
Lin YC, Roffler SR, Yan YT, Yang RB 2015, 'Disruption of Scube2 impairs endochondral bone formation', Journal of Bone and Mineral Research, vol. 30, no. 7, pp. 1255-1267.
Lin Z, Hegarty JP, Floros J, Franke A 2016, 'NKX2-3 (NK2 homeobox 3)', Atlas of Genetics and Cytogenetics in Oncology and Haematology.
Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C 2014, 'Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears', Cell, vol. 157, no. 4, pp. 785-794.
Liu Y, Rossiter SJ, Han X, Cotton JA, Zhang S 2010, 'Cetaceans on a molecular fast track to ultrasonic hearing', Current Biology, vol. 20, no. 20, pp. 1834-1839.
Lockwood BL, Somero GN 2011, 'Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*)', Molecular Ecology, vol. 20, no. 3, pp. 517-529.
Lopes-Marques M, Machado AM, Alves LQ, Fonseca MM, Barbosa S, Sinding M-HS, Rasmussen MH, Iversen MR, Frost Bertelsen M, Campos PF 2019, 'Complete inactivation of sebum-producing genes parallels the loss of sebaceous glands in Cetacea', Molecular Biology and Evolution, vol. 36, no. 6, pp. 1270-1280.
Louis M, Fontaine MC, Spitz J, Schlund E, Dabin W, Deaville R, Caurant F, Cherel Y, Guinet C, Simon-Bouhet B 2014a, 'Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator', Proceedings of the Royal Society of London B: Biological Sciences, vol. 281, no. 1795, pp. 20141558.
Louis M, Viricel A, Lucas T, Peltier H, Alfonsi E, Berrow S, Brownlow A, Covelo P, Dabin W, Deaville R 2014b, 'Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in the North-East Atlantic', Molecular Ecology, vol. 23, no. 4, pp. 857-874.

Lowther-Thieleking J, Archer F, Lang A, Weller D 2015, 'Genetic differentiation among coastal and offshore common bottlenose dolphins, *Tursiops truncatus*, in the eastern North Pacific Ocean', Marine Mammal Science, vol. 31, no. 1, pp. 1-20.

Lowther AD, Goldsworthy S 2011, 'Detecting alternate foraging ecotypes in Australian sea lion (*Neophoca cinerea*) colonies using stable isotope analysis', Marine Mammal Science, vol. 27, no. 3, pp. 567-586.

Lowther AD, Harcourt RG, Goldsworthy SD, Stow A 2012, 'Population structure of adult female Australian sea lions is driven by fine-scale foraging site fidelity', Animal Behaviour, vol. 83, no. 3, pp. 691-701.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P 2003, 'The power and promise of population genomics: from genotyping to genome typing', Nature Reviews Genetics, vol. 4, no. 12, pp. 981-994.

Luque VJ 2016, 'Drift and Evolutionary Forces: Scrutinizing the Newtonian Analogy', Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia, vol. 31, no. 3, pp. 397-410.

Macdonald H, Roughan M, Baird M, Wilkin J 2013, 'A numerical modeling study of the East Australian Current encircling and overwashing a warm-core eddy', Journal of Geophysical Research: Oceans, vol. 118, no. 1, pp. 301-315.

Mah S, Nelson M, Delisi L, Reneland R, Markward N, James M, Nyholt D, Hayward N, Handoko H, Mowry B 2006, 'Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia', Molecular Psychiatry, vol. 11, no. 5, pp. 471.

Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, Aurelle D 2016, 'Genomic resources and their influence on the detection of the signal of positive selection in genome scans', Molecular Ecology, vol. 25, no. 1, pp. 170-184.

Marino L 2005, 'Big brains do matter in new environments', Proceedings of the National Academy of Sciences, vol. 102, no. 15, pp. 5306-5307.

Marino L 2007, 'Cetacean brains: how aquatic are they?', The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, vol. 290, no. 6, pp. 694-700.

Martien KK, Chivers SJ, Baird RW, Archer FI, Gorgone AM, Hancock-Hanser BL, Mattila D, McSweeney DJ, Oleson EM, Palmer C 2014, 'Nuclear and mitochondrial patterns of population structure in North Pacific false killer whales (*Pseudorca crassidens*)', Journal of Heredity, vol. 105, no. 5, pp. 611-626.

Marx FG, Uhen MD 2010, 'Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales', Science, vol. 327, no. 5968, pp. 993-996.

Matsuzawa Y, Kamezaki N, Ishihara T, Omuta K, Takeshita H, Goto K, Arata T, Honda H, Kameda K, Kashima Y 2016, 'Fine-scale genetic population structure of loggerhead turtles in the Northwest Pacific', Endangered Species Research, vol. 30, pp. 83-93.

Matthews JM, Sunde M 2002, 'Zinc Fingers--Folds for Many Occasions', IUBMB Life, vol. 54, no. 6, pp. 351-355.

McGowen MR, Gatesy J, Wildman DE 2014, 'Molecular evolution tracks macroevolutionary transitions in Cetacea', Trends in Ecology & Evolution, vol. 29, no. 6, pp. 336-346.

McGowen MR, Grossman LI, Wildman DE 2012, 'Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown', Proceedings of the Royal Society B, vol. 279, no. 1743, pp. 3643–3651.

McKinnon JS, Rundle HD 2002, 'Speciation in nature: the threespine stickleback model systems', Trends in Ecology & Evolution, vol. 17, no. 10, pp. 480-488.

Mead J, Potter C 1995, 'Recognizing two populations of the bottlenose dolphin (*Tursiops truncatus*) off the coast of North America: Morphologic and ecological considerations', Index of Biological Integrity Report, vol. 5, pp. 3144.

Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A 2007, 'Global climate projections', Climate Change, vol. 3495, no. 1, pp. 747-845.

Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C, Bordino P 2010, 'Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range', Molecular Ecology, vol. 19, no. 11, pp. 2212-2228.

Mendez M, Subramaniam A, Collins T, Minton G, Baldwin R, Berggren P, Särnblad A, Amir O, Peddemors V, Karczmarski L 2011, 'Molecular ecology meets remote sensing: environmental drivers to population structure of humpback dolphins in the Western Indian Ocean', Heredity, vol. 107, no. 4, pp. 349.

Meredith RW, Gatesy J, Emerling CA, York VM, Springer MS 2013, 'Rod monochromacy and the coevolution of cetacean retinal opsins', PLoS Genetics, vol. 9, no. 4, pp. e1003432.

Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TL, Stadler T 2011, 'Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification', Science, vol. 334, no. 6055, pp. 521-524.

Middleton JF, Bye JAT 2007, 'A review of the shelf-slope circulation along Australia's southern shelves: Cape Leeuwin to Portland', Progress in Oceanography, vol. 75, no. 1, pp. 1-41.

Mikkola ML, Thesleff I 2003, 'Ectodysplasin signaling in development', Cytokine & Growth Factor Reviews, vol. 14, no. 3-4, pp. 211-224.

Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC 2013, 'Ocean currents influence the genetic structure of an intertidal mollusc in southeastern A ustralia– implications for predicting the movement of passive dispersers across a marine biogeographic barrier', Ecology and Evolution, vol. 3, no. 5, pp. 1248-1261.

Mink JW, Blumenschine RJ, Adams DB 1981, 'Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis', American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, vol. 241, no. 3, pp. 203-212.

Mirceta S, Signore AV, Burns JM, Cossins AR, Campbell KL, Berenbrink M 2013, 'Evolution of mammalian diving capacity traced by myoglobin net surface charge', Science, vol. 340, no. 6138, pp. 1303-1312.

Mirimin L, Miller R, Dillane E, Berrow SD, Ingram S, Cross TF, Rogan E 2011, 'Fine-scale population genetic structuring of bottlenose dolphins in Irish coastal waters', Animal Conservation, vol. 14, no. 4, pp. 342-353.

Mironova GD, Shigaeva MI, Gritsenko EN, Murzaeva SV, Gorbacheva OS, Germanova EL, Lukyanova LD 2010, 'Functioning of the mitochondrial ATP-dependent potassium channel in rats varying in their resistance to hypoxia. Involvement of the channel in the process of animal's adaptation to hypoxia', Journal of Bioenergetics and Biomembranes, vol. 42, no. 6, pp. 473-481.

Mizukami S, Watanabe Y, Saegusa Y, Nakajima K, Ito Y, Masubuchi Y, Yoshida T, Shibutani M 2017, 'Downregulation of UBE2E2 in rat liver cells after hepatocarcinogen treatment facilitates cell proliferation and slowing down of DNA damage response in GST-P-expressing preneoplastic lesions', Toxicology and Applied Pharmacology, vol. 334, pp. 207-216.

Möller L 2012, 'Sociogenetic structure, kin associations and bonding in delphinids', Molecular Ecology, vol. 21, no. 3, pp. 745-764.

Möller L, Allen S, Harcourt R 2002, 'Group characteristics, site fidelity and seasonal abundance of bottlenose dolphins (*Tursiops aduncus*) in Jervis Bay and Port Stephens, south-eastern Australia', Australian Mammalogy, vol. 24, no. 1, pp. 11-22.

Möller L, Beheregaray L, Allen S, Harcourt R 2006, 'Association patterns and kinship in female Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) of southeastern Australia', Behavioural Ecology and Sociobiology, vol. 61, no. 1, pp. 109-117.

Möller L, Bilgmann K, Charlton-Robb K, Beheregaray L 2008, 'Multi-gene evidence for a new bottlenose dolphin species in southern Australia', Molecular Phylogenetics and Evolution, vol. 49, no. 1, pp. 674-681.

Möller L, Valdez F, Allen S, Bilgmann K, Corrigan S, Beheregaray L 2011, 'Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current', Marine Biology, vol. 158, no. 1, pp. 113-126.

Möller L, Wiszniewski J, Allen S, Beheregaray L 2007, 'Habitat type promotes rapid and extremely localised genetic differentiation in dolphins', Marine and Freshwater Research, vol. 58, no. 7, pp. 640-648.

Möller LM, Beheregaray LB 2001, 'Coastal bottlenose dolphins from southeastern Australia are *Tursiops aduncus* according to sequences of the mitochondrial DNA control region', Marine Mammal Science, vol. 17, no. 2, pp. 249-263.

Möller LM, Beheregaray LB 2004, 'Genetic evidence for sex-biased dispersal in resident bottlenose dolphins (*Tursiops aduncus*)', Molecular Ecology, vol. 13, no. 6, pp. 1607-1612.

Möller LM, Beheregaray LB, Harcourt RG, Krützen M 2001, 'Alliance membership and kinship in wild male bottlenose dolphins (*Tursiops aduncus*) of southeastern Australia', Proceedings of the Royal Society of London B: Biological Sciences, vol. 268, no. 1479, pp. 1941-1947.

Monroig Ó, Lopes-Marques M, Navarro JC, Hontoria F, Ruivo R, Santos MM, Venkatesh B, Tocher DR, Castro LFC 2016, 'Evolutionary functional elaboration of the Elovl2/5 gene family in chordates', Scientific Reports, vol. 6.

Montes I, Zarraonaindia I, Iriondo M, Grant WS, Manzano C, Cotano U, Conklin D, Irigoien X, Estonba A 2016, 'Trancriptome analysis deciphers evolutionary mechanisms underlying genetic differentiation between coastal and offshore anchovy populations in the Bay of Biscay', Marine Biology, vol. 163, no. 10, pp. 205.

Morais S, Monroig O, Zheng X, Leaver MJ, Tocher DR 2009, 'Highly unsaturated fatty acid synthesis in Atlantic salmon: characterization of ELOVL5-and ELOVL2-like elongases', Marine Biotechnology, vol. 11, no. 5, pp. 627-639.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L 2010, 'Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species', Genome Research, vol. 20, no. 7, pp. 908-916.

Motta PJ, Huber DR 2004, 'Prey capture behavior and feeding mechanics of elasmobranchs'. In: Carrier, J, Musick, J & Heithaus, MR (eds) *Biology of Sharks and Their Relatives*. CRC Press LLC, USA, pp. 165-202.

Moura A, Kenny J, Chaudhuri R, Hughes M, Reisinger R, De Bruyn P, Dahlheim M, Hall N, Hoelzel A 2015, 'Phylogenomics of the killer whale indicates ecotype divergence in sympatry', Heredity, vol. 114, no. 1, pp. 48-55.

Moura A, Nielsen S, Vilstrup J, Moreno-Mayar J, Gilbert M, Gray H, Natoli A, Möller L, Hoelzel A 2013, 'Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change', Systematic Biology, vol. 62, no. 6, pp. 865-877.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, J Welch A, Reisinger RR, Bruyn P, Dahlheim ME, Hall N, Hoelzel AR 2014, 'Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift', Molecular Ecology, vol. 23, no. 21, pp. 5179-5192.

Moura AE, Shreves K, Pilot M, Andrews KR, Moore DM, Kishida T, Möller L, Natoli A, Gaspari S, McGowen M 2020, 'Phylogenomics of the genus *Tursiops* and closely related Delphininae reveals extensive reticulation among lineages and provides inference about eco-evolutionary drivers', Molecular Phylogenetics and Evolution, vol. 146, pp. 106756.

Muruganandan S, Roman AA, Sinal CJ 2010, 'Role of chemerin/CMKLR1 signaling in adipogenesis and osteoblastogenesis of bone marrow stem cells', Journal of Bone and Mineral Research, vol. 25, no. 2, pp. 222-234.

Nagaishi M, Kim Y-H, Mittelbronn M, Giangaspero F, Paulus W, Brokinkel B, Vital A, Tanaka Y, Nakazato Y, Legras-Lachuer C 2012, 'Amplification of the STOML3, FREM2, and LHFP genes is associated with mesenchymal differentiation in gliosarcoma', The American Journal of Pathology, vol. 180, no. 5, pp. 1816-1823.

Nagy JA, Dvorak AM, Dvorak HF 2007, 'VEGF-A and the induction of pathological angiogenesis', Annual Review of Pathology: Mechanisms of Disease, vol. 2, pp. 251-275.

Nakade K, Pan J, Yoshiki A, Ugai H, Kimura M, Liu B, Li H, Obata Y, Iwama M, Itohara S 2007, 'JDP2 suppresses adipocyte differentiation by regulating histone acetylation', Cell Death and Differentiation, vol. 14, no. 8, pp. 1398.

Nanninga GB, Saenz-Agudelo P, Manica A, Berumen ML 2014, 'Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea', Molecular Ecology, vol. 23, no. 3, pp. 591-602.

Narum SR 2006, 'Beyond Bonferroni: less conservative analyses for conservation genetics', Conservation Genetics, vol. 7, no. 5, pp. 783-787.

Natoli A, Birkun A, Lopez A, Hoelzel A 2005, 'Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*)', The Royal Society, vol. 272, no. 1569, pp. 1217–1226.

Natoli A, Peddemors VM, Rus Hoelzel A 2004, 'Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses', Journal of Evolutionary Biology, vol. 17, no. 2, pp. 363-375.

Nery MF, Arroyo JI, Opazo JC 2013a, 'Accelerated evolutionary rate of the myoglobin gene in long-diving whales', Journal of Molecular Evolution, vol. 76, no. 6, pp. 380-387.

Nery MF, González DJ, Opazo JC 2013b, 'How to make a dolphin: Molecular signature of positive selection in cetacean genome', PLoS One, vol. 8, no. 6, pp. e65491.

Nevo E 2001, 'Evolution of genome–phenome diversity under environmental stress', Proceedings of the National Academy of Sciences, vol. 98, no. 11, pp. 6233-6240.

Ni J, Zhou K 1988, 'Rencular structural indices and urinary concentrating capacity of *Neophocaena phocaenoides*', Acta Zoologica Sinica, vol. 3, pp. 243-250.

Norris R, Turner SK, Hull P, Ridgwell A 2013, 'Marine ecosystem responses to Cenozoic global change', Science, vol. 341, no. 6145, pp. 492-498.

Nosil P, Feder JL 2012, 'Genomic divergence during speciation: Causes and consequences', Philosophical Transactions of The Royal Society, vol. 367, pp. 332-342.

Nykänen M, Kaschner K, Dabin W, Brownlow A, Davison NJ, Deaville R, Garilao C, Kesner-Reyes K, Gilbert TP, Penrose R 2019, 'Post-glacial colonization of northern coastal habitat by bottlenose dolphins: A marine leading-edge expansion?', Journal of Heredity, vol. 110, no. 6, pp. 1-13.

O'Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM 2015, 'A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA', Elife, vol. 4, pp. e05290.

O'Connell LG, James NP, Doubell M, Middleton JF, Luick J, Currie DR, Bone Y 2016, 'Oceanographic controls on shallow-water temperate carbonate sedimentation: Spencer Gulf, South Australia', Sedimentology, vol. 63, no. 1, pp. 105-135.

O’Brien RM 2007, 'A caution regarding rules of thumb for variance inflation factors', Quality & Quantity, vol. 41, no. 5, pp. 673-690.

Oelschläger H, Ridgway S, Knauth M 2010, 'Cetacean brain evolution: Dwarf sperm whale (*Kogia sima*) and common dolphin (*Delphinus delphis*) – an investigation with high-resolution 3D MRI', Brain, Behavior and Evolution, vol. 75, no. 1, pp. 33-62.

Ohazama A, Sharpe PT 2007, 'TFII-I gene family during tooth development: Candidate genes for tooth anomalies in Williams syndrome', Developmental Dynamics: An Official Publication of the American Association of Anatomists, vol. 236, no. 10, pp. 2884-2888.

Oke PR, Middleton JH 2001, 'Nutrient enrichment off Port Stephens: The role of the East Australian Current', Continental Shelf Research, vol. 21, no. 6-7, pp. 587-606.

Oke PR, Roughan M, Cetina-Heredia P, Pilo GS, Ridgway KR, Rykova T, Archer MR, Coleman RC, Kerry CG, Rocha C 2019, 'Revisiting the circulation of the East Australian Current: Its path, separation, and eddy field', Progress in Oceanography, vol. 176, pp. 102139.

Oksanen J (2011) Vegan: R functions for vegetation ecologists [Online]. Available: https://cran.r-project.org/web/packages/vegan/index.html].

Ooi A, Chong V 2011, 'Larval fish assemblages in a tropical mangrove estuary and adjacent coastal waters: Offshore–inshore flux of marine and estuarine species', Continental Shelf Research, vol. 31, no. 15, pp. 1599-1610.

Ortiz RM 2001, 'Osmoregulation in marine mammals', Journal of Experimental Biology, vol. 204, no. 11, pp. 1831-1844.
Oudejans MG, Visser F, Englund A, Rogan E, Ingram SN 2015, 'Evidence for distinct coastal and offshore communities of bottlenose dolphins in the North East Atlantic', PLoS One, vol. 10, no. 4, pp. 15.
Paquette S 2011. PopGenKit v.1.0. CRAN Repository: https://CRAN.R-project.org/package=PopGenKit.
Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS 2005, 'The effect of genetic drift in a young genetically isolated population', Annals of Human Genetics, vol. 69, no. 3, pp. 288-295.
Park S-W, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, Killeen N, Erle DJ 2009, 'The protein disulfide isomerase AGR2 is essential for production of intestinal mucus', Proceedings of the National Academy of Sciences, vol. 106, no. 17, pp. 6950-6955.
Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ 2013, 'Genome-wide signatures of convergent evolution in echolocating mammals', Nature, vol. 502, no. 7470, pp. 228-231.
Passadore C, Möller L, Diaz-Aguirre F, Parra GJ 2017, 'Demography of southern Australian bottlenose dolphins living in a protected inverse estuary', Aquatic Conservation: Marine and Freshwater Ecosystems, vol. 27, no. 6, pp. 1186-1197.
Passadore C, Möller L, Diaz-Aguirre F, Parra GJ 2018a, 'High site fidelity and restricted ranging patterns in southern Australian bottlenose dolphins', Ecology and Evolution, vol. 8, no. 1, pp. 242-256.
Passadore C, Möller LM, Diaz-Aguirre F, Parra GJ 2018b, 'Modelling dolphin distribution to inform future spatial conservation decisions in a marine protected area', Scientific Reports, vol. 8, no. 1, pp. 15659.
Payne NL, Snelling EP, Semmens JM, Gillanders BM 2013, 'Mechanisms of population structuring in giant Australian cuttlefish *Sepia apama*', PLoS One, vol. 8, no. 3, pp. 7.
Peakall R, Smouse PE 2006, 'GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research', Molecular Ecology Resources, vol. 6, no. 1, pp. 288-295.
Peakall R, Smouse PE 2012, 'GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update', Bioinformatics, vol. 28, pp. 2537-2539.
Peck JR 1994, 'A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex', Genetics, vol. 137, no. 2, pp. 597-606.
Pelletier F, Coltman DW 2018, 'Will human influences on evolutionary dynamics in the wild pervade the Anthropocene?', BMC Biology, vol. 16, no. 1, pp. 7.
Pérez-Alvarez MJ, Vásquez RA, Moraga R, Santos-Carvallo M, Kraft S, Sabaj V, Capella J, Gibbons J, Vilina Y, Poulin E 2018, 'Home sweet home: Social dynamics and genetic variation of a long-term resident bottlenose dolphin population off the Chilean coast', Animal Behaviour, vol. 139, pp. 81-89.
Perreault-Payette A, Muir AM, Goetz F, Perrier C, Normandeau E, Sirois P, Bernatchez L 2017, 'Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior', Molecular Ecology, vol. 26, no. 6, pp. 1477-1497.
Perrin WF, Robertson KM, Van Bree PJ, Mead JG 2007, 'Cranial description and genetic identity of the holotype specimen of *Tursiops aduncus* (Ehrenberg, 1832)', Marine Mammal Science, vol. 23, no. 2, pp. 343-357.
Perrin WF, Thieleking JL, Walker WA, Archer FI, Robertson KM 2011, 'Common bottlenose dolphins (*Tursiops truncatus*) in California waters: Cranial differentiation of coastal and offshore ecotypes', Marine Mammal Science, vol. 27, no. 4, pp. 769-792.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE 2012, 'Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species', PLoS One, vol. 7, no. 5, pp. e37135.

Peterson D, Hanazaki N, Simoes-Lopes PC 2008, 'Natural resource appropriation in cooperative artisanal fishing between fishermen and dolphins (*Tursiops truncatus*) in Laguna, Brazil', Ocean & Coastal Management, vol. 51, no. 6, pp. 469-475.

Petrusevics P 1993, 'SST fronts in inverse estuaries, South Australia - Indicators of reduced gulf-shelf exchange', Marine and Freshwater Research, vol. 44, no. 2, pp. 305-323.

Piggott MP, Banks SC, Tung P, Beheregaray LB 2008, 'Genetic evidence for different scales of connectivity in a marine mollusc', Marine Ecology Progress Series, vol. 365, pp. 127-136.

Pitman RL, Durban JW 2012, 'Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters', Marine Mammal Science, vol. 28, no. 1, pp. 16-36.

Pitman RL, Durban JW, Greenfelder M, Guinet C, Jorgensen M, Olson PA, Plana J, Tixier P, Towers JR 2011, 'Observations of a distinctive morphotype of killer whale (*Orcinus orca*), type D, from subantarctic waters', Polar Biology, vol. 34, no. 2, pp. 303-306.

Poloczanska ES, Brown CJ, Sydeman WJ, Kiessling W, Schoeman DS, Moore PJ, Brander K, Bruno JF, Buckley LB, Burrows MT 2013, 'Global imprint of climate change on marine life', Nature Climate Change, vol. 3, no. 10, pp. 919.

Portnoy DS, Hollenbeck CM, Belcher CN, Driggers WB, Frazier BS, Gelsleichter J, Grubbsk RD, Gold JR 2014, 'Contemporary population structure and post-glacial genetic demography in a migratory marine species, the blacknose shark, *Carcharhinus acronotus*', Molecular Ecology, vol. 23, no. 22, pp. 5480-5495.

Pratt EAL, Beheregaray LB, Bilgmann K, Zanardo N, Diaz-Aguirre F, Möller LM 2018, 'Hierarchical metapopulation structure in a highly mobile marine predator: the southern Australian coastal bottlenose dolphin (*Tursiops* cf. *australis*)', Conservation Genetics, vol. 19, no. 3, pp. 637-654.

Puritz JB, Hollenbeck CM, Gold JR 2014, 'dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms', PeerJ, vol. 2.

Pyenson ND, Kelley NP, Parham JF 2014, 'Marine tetrapod macroevolution: Physical and biological drivers on 250Ma of invasions and evolution in ocean ecosystems', Palaeogeography, Palaeoclimatology, Palaeoecology, vol. 400, no. 1, pp. 1-8.

Queller DC, Goodnight KF 1989, 'Estimating relatedness using genetic markers', Evolution, vol. 43, no. 2, pp. 258-275.

Querouil S, Silva MA, Freitas L, Prieto R, Magalhaes S, Dinis A, Alves F, Matos JA, Mendonca D, Hammond PS, Santos RS 2007, 'High gene flow in oceanic bottlenose dolphins (*Tursiops truncatus*) of the North Atlantic', Conservation Genetics, vol. 8, no. 6, pp. 1405-1419.

Raeymaekers JA, Chaturvedi A, Hablützel PI, Verdonck I, Hellemans B, Maes GE, De Meester L, Volckaert FA 2017, 'Adaptive and non-adaptive divergence in a common landscape', Nature Communications, vol. 8, no. 1, pp. 267.

Rambaut A (2014) FigTree v.1.4.3: tree figure drawing tool [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/].

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA 2018, 'Posterior summarisation in Bayesian phylogenetics using Tracer 1.7', Systematic Biology, vol. 67, no. 5, pp. 901-904.

Rannala B 2007, 'BayesAss edition 3.0 user's manual', Davis: University of California at Davis, vol. Updated 2015.

Reece M, Prawitt D, Landers J, Kast C, Gros P, Housman D, Zabel BU, Pelletier J 1998, 'Functional characterization of ORCTL2-an organic cation transporter expressed in the renal proximal tubules', FEBS Letters, vol. 433, no. 3, pp. 245-250.

Reeves RR, Smith BD, Crespo EA, Notarbartolo di Sciara G (2003) Dolphins, whales and porpoises: 2002-2010 conservation action plan for the world's cetaceans. IUCN/SSC Cetacean Specialist Group, IUCN, Gland Switzwerland and Cambridge, UK.

Reidenberg JS 2007, 'Anatomical adaptations of aquatic mammals', The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, vol. 290, no. 6, pp. 507-513.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R 2015, 'A practical guide to environmental association analysis in landscape genomics', Molecular Ecology, vol. 24, no. 17, pp. 4348-4370.

Reznick DN, Ricklefs RE 2009, 'Darwin's bridge between microevolution and macroevolution', Nature, vol. 457, no. 7231, pp. 837-842.

Richard G, Cox SL, Picard B, Vacquié-Garcia J, Guinet C 2016, 'Southern elephant seals replenish their lipid reserves at different rates according to foraging habitat', PLoS One, vol. 11, no. 11, pp. e0166747.

Ridgway K 2007, 'Long-term trend and decadal variability of the southward penetration of the East Australian Current', Geophysical Research Letters, vol. 34, no. 13.

Ridgway SH, Carlin KP, Van Alstyne KR 2018, 'Delphinid brain development from neonate to adulthood with comparisons to other cetaceans and artiodactyls', Marine Mammal Science, vol. 34, no. 2, pp. 420-439.

Ridgway SH, Carlin KP, Van Alstyne KR, Hanson AC, Tarpley RJ 2016, 'Comparison of dolphins' body and brain measurements with four other groups of cetaceans reveals great diversity', Brain, Behavior and Evolution, vol. 88, no. 3-4, pp. 235-257.

Rifas L, Weitzmann MN 2009, 'A novel T cell cytokine, secreted osteoclastogenic factor of activated T cells, induces osteoclast formation in a RANKL-independent manner', Arthritis & Rheumatism: Official Journal of the American College of Rheumatology, vol. 60, no. 11, pp. 3324-3335.

Robbins WD, Huveneers C, Parra GJ, Möller L, Gillanders BM 2017, 'Anthropogenic threat assessment of marine-associated fauna in Spencer Gulf, South Australia', Marine Policy, vol. 81, pp. 392-400.

Roberts S, Dixon C, Andreacchio L 2012, 'Temperature dependent larval duration and survival of the western king prawn, *Penaeus (Melicertus) latisulcatus* Kishinouye, from Spencer Gulf, South Australia', Journal of Experimental Marine Biology and Ecology, vol. 411, pp. 14-22.

Rochford D 1986, 'Seasonal changes in the distribution of Leeuwin Current waters of Southern Australia', Marine and Freshwater Research, vol. 37, no. 1, pp. 1-10.

Rode KD, Robbins CT, Nelson L, Amstrup SC 2015, 'Can polar bears use terrestrial foods to offset lost ice-based hunting opportunities?', Frontiers in Ecology and the Environment, vol. 13, no. 3, pp. 138-145.

Rodríguez-Zárate CJ, Sandoval-Castillo J, van Sebille E, Keane RG, Rocha-Olivares A, Urteaga J, Beheregaray LB 2018, 'Isolation by environment in the highly mobile olive ridley turtle (*Lepidochelys olivacea*) in the eastern Pacific', Proceedings of the Royal Society B: Biological Sciences, vol. 285, no. 1878, pp. 20180264.

Rogers S, Bernatchez L 2005, 'Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*)', Molecular Ecology, vol. 14, no. 2, pp. 351-361.

Rogers S, Bernatchez L 2007, 'The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs', Molecular Biology and Evolution, vol. 24, no. 6, pp. 1423-1438.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N 2014, 'Comparative population genomics in animals uncovers the determinants of genetic diversity', Nature, vol. 515, no. 7526, pp. 261.

Ross G, Cockcroft V 1990, 'Comments on Australian bottlenose dolphins and the taxonomic status of *Tursiops aduncus* (Ehrenberg, 1832)'. In: Leatherwood, S & Reeves, R (eds) *The Bottlenose Dolphin*. Academic Press, San Diego, pp. 101-128.

Ruan R, Guo A-H, Hao Y-J, Zheng J-S, Wang D 2015, 'De novo assembly and characterization of narrow-ridged finless porpoise renal transcriptome and identification of candidate genes involved in osmoregulation', International Journal of Molecular Sciences, vol. 16, no. 1, pp. 2220-2238.

Rybczynski N, Dawson MR, Tedford RH 2009, 'A semi-aquatic Arctic mammalian carnivore from the Miocene epoch and origin of Pinnipedia', Nature, vol. 458, no. 7241, pp. 1021-1024.

Saisawat P, Tasic V, Vega-Warner V, Kehinde EO, Günther B, Airik R, Innis JW, Hoskins BE, Hoefele J, Otto EA 2012, 'Identification of two novel CAKUT-causing genes by massively parallel exon resequencing of candidate genes in patients with unilateral renal agenesis', Kidney International, vol. 81, no. 2, pp. 196-200.

Sakakibara S-i, Nakamura Y, Yoshida T, Shibata S, Koike M, Takano H, Ueda S, Uchiyama Y, Noda T, Okano H 2002, 'RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation', Proceedings of the National Academy of Sciences, vol. 99, no. 23, pp. 15194-15199.

Sakurai A, Doci C, Gutkind JS 2012, 'Semaphorin signaling in angiogenesis, lymphangiogenesis and cancer', Cell Research, vol. 22, no. 1, pp. 23.

Samarra FIP, Miller PJO 2015, 'Prey-induced behavioural plasticity of herring-eating killer whales', Marine Biology, vol. 162, no. 4, pp. 809-821.

Sandoval-Castillo J, Robinson NA, Hart AM, Strain LW, Beheregaray LB 2018, 'Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient', Molecular Ecology, vol. 27, no. 7, pp. 1603-1620.

Santora JA, Hazen EL, Schroeder ID, Bograd SJ, Sakuma KM, Field JC 2017, 'Impacts of ocean climate variability on biodiversity of pelagic forage species in an upwelling ecosystem', Marine Ecology Progress Series, vol. 580, pp. 205-220.

Sarachana T, Hu VW 2013, 'Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder', Molecular Autism, vol. 4, no. 1, pp. 14.

Sargeant BL, Mann J, Berggren P, Krützen M 2005, 'Specialization and development of beach hunting, a rare foraging behavior, by wild bottlenose dolphins (*Tursiops* sp.)', Canadian Journal of Zoology, vol. 83, no. 11, pp. 1400-1410.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T 2019, 'Database resources of the national center for biotechnology information', Nucleic Acids Research, vol. 47, no. Database issue, pp. D23.

Sazzini M, Schiavo G, De Fanti S, Martelli P, Casadio R, Luiselli D 2014, 'Searching for signatures of cold adaptations in modern and archaic humans: hints from the brown adipose tissue genes', Heredity, vol. 113, no. 3, pp. 259.

Scheffers BR, De Meester L, Bridge TC, Hoffmann AA, Pandolfi JM, Corlett RT, Butchart SH, Pearce-Kelly P, Kovacs KM, Dudgeon D 2016, 'The broad footprint of climate change from genes to biomes to people', Science, vol. 354, no. 6313.

Schluter D 2001, 'Ecology and the origin of species', Trends in Ecology & Evolution, vol. 16, no. 7, pp. 372-380.

Schrauwen P, Westerterp KR 2000, 'The role of high-fat diets and physical activity in the regulation of body weight', British Journal of Nutrition, vol. 84, no. 4, pp. 417-427.

Schröder-Adams CJ, Boyd RL, Tran T 2014, 'Estuarine foraminiferal biofacies pattern compared to the brackish ichnofacies model: Port Stephens, southeast Australia', Estuarine, Coastal and Shelf Science, vol. 139, pp. 78-87.

Schuliga M, Westall G, Xia Y, Stewart AG 2013, 'The plasminogen activation system: new targets in lung inflammation and remodeling', Current Opinion in Pharmacology, vol. 13, no. 3, pp. 386-393.

Schunter C, Carreras-Carbonell J, Macpherson E, Tintoré J, Vidal-Vijande E, Pascual A, Guidetti P, Pascual M 2011, 'Matching genetics with oceanography: directional gene flow in a Mediterranean fish species', Molecular Ecology, vol. 20, no. 24, pp. 5167-5181.

Seale P, Bjork B, Yang W, Kajimura S, Chin S, Kuang S, Scime A, Devarakonda S, Conroe HM, Erdjument-Bromage H 2008, 'PRDM16 controls a brown fat/skeletal muscle switch', Nature, vol. 454, no. 7207, pp. 961.

Seale P, Kajimura S, Yang W, Chin S, Rohas LM, Uldry M, Tavernier G, Langin D, Spiegelman BM 2007, 'Transcriptional control of brown fat determination by PRDM16', Cell Metabolism, vol. 6, no. 1, pp. 38-54.
Secchi E 2007. Population Viability Analysis of a small resident population of bottlenose dolphins, *Tursiops truncatus*, in southern Brazil. Rio Grande: Yaqupacha/Laboratório Mamíferos Marinhos, Museu Oceanográfico,.
Sellas AB, Wells RS, Rosel PE 2005, 'Mitochondrial and nuclear DNA analyses reveal fine scale geographic structure in bottlenose dolphins (*Tursiops truncatus*) in the Gulf of Mexico', Conservation Genetics, vol. 6, no. 5, pp. 715-728.
Shaddick K, Gilligan DM, Burridge CP, Jerry DR, Truong K, Beheregaray LB 2011, 'Historic divergence with contemporary connectivity in a catadromous fish, the estuary perch (*Macquaria colonorum*)', Canadian Journal of Fisheries and Aquatic Sciences, vol. 68, no. 2, pp. 304-318.
Shen Y-Y, Liang L, Li G-S, Murphy RW, Zhang Y-P 2012a, 'Parallel evolution of auditory genes for echolocation in bats and toothed whales', PLoS Genetics, vol. 8, no. 6, pp. e1002788.
Shen Y-Y, Liang L, Zhu Z-H, Zhou W-P, Irwin DM, Zhang Y-P 2010, 'Adaptive evolution of energy metabolism genes and the origin of flight in bats', Proceedings of the National Academy of Sciences, vol. 107, no. 19, pp. 8666-8671.
Shen Y-Y, Zhou W-P, Zhou T-C, Zeng Y-N, Li G-M, Irwin DM, Zhang Y-P 2012b, 'Genome-wide scan for bats and dolphin to detect their genetic basis for new locomotive styles', PLoS One, vol. 7, no. 11, pp. e46455.
Shinohara M, Nakamura M, Masuda H, Hirose J, Kadono Y, Iwasawa M, Nagase Y, Ueki K, Kadowaki T, Sasaki T 2012, 'Class IA phosphatidylinositol 3-kinase regulates osteoclastic bone resorption through protein kinase B–mediated vesicle transport', Journal of Bone and Mineral Research, vol. 27, no. 12, pp. 2464-2475.
Shirakihara M, Yoshida H, Shirakihara K 2003, 'Indo-Pacific bottlenose dolphins *Tursiops aduncus* in Amakusa, western Kyushu, Japan', Fisheries Science, vol. 69, no. 3, pp. 654-656.
Silva G, Horne JB, Castilho R 2014, 'Anchovies go north and west without losing diversity: Post-glacial range expansions in a small pelagic fish', Journal of Biogeography, vol. 41, no. 6, pp. 1171-1182.
Silveira AC, Morrison MA, Ji F, Xu H, Reinecke JB, Adams SM, Arneberg TM, Janssian M, Lee J-E, Yuan Y 2010, 'Convergence of linkage, gene expression and association data demonstrates the influence of the RAR-related orphan receptor alpha (RORA) gene on neovascular AMD: A systems biology based approach', Vision Research, vol. 50, no. 7, pp. 698-715.
Simmons NB 2005, 'An Eocene big bang for bats', Science, vol. 307, no. 5709, pp. 527-528.
Simões-Lopes PC, Fabian ME 1999, 'Residence patterns and site fidelity in bottlenose dolphins, *Tursiops truncatus* (Montagu) (Cetacea, Delphinidae) off Southern Brazil', Brazilian Journal of Zoology, vol. 16, no. 4, pp. 1017-1024.
Simões-Lopes PC, Fabián ME, Menegheti JO 1998, 'Dolphin interactions with the mullet artisanal fishing on southern Brazil: a qualitative and quantitative approach', Brazilian Journal of Zoology, vol. 15, no. 3, pp. 709-726.
Simonson T, Huff C, Witherspoon D, Prchal J, Jorde L 2015, 'Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans', Experimental Physiology, vol. 100, no. 11, pp. 1263-1268.
Sirinupong N, Brunzelle J, Ye J, Pirzada A, Nico L, Yang Z 2010, 'Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture', Journal of Biological Chemistry, vol. 285, no. 52, pp. 40635-40644.
Siriwardhana N, Kalupahana NS, Moustaid-Moussa N 2012, 'Health benefits of n-3 polyunsaturated fatty acids: eicosapentaenoic acid and docosahexaenoic acid'. In: Kim, S-K (ed.) *Advances in Food and Nutrition Research*. Elsevier, pp. 211-222.
Slatkin M 1987, 'Gene flow and the geographic structure of natural-populations', Science, vol. 236, no. 4803, pp. 787-792.

Sol D, Duncan RP, Blackburn TM, Cassey P, Lefebvre L 2005, 'Big brains, enhanced cognition, and response of birds to novel environments', Proceedings of the National Academy of Sciences, vol. 102, no. 15, pp. 5460-5465.

Speakman JR 2001, 'The evolution of flight and echolocation in bats: Another leap in the dark', Mammal Review, vol. 31, no. 2, pp. 111-130.

Speakman JR 2018, 'Obesity and thermoregulation'. In: Romanovsky, A (ed.) *Handbook of Clinical Neurology*. Elsevier, pp. 431-443.

Squillace RM, Chenault DM, Wang EH 2002, 'Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR', Developmental Biology, vol. 250, no. 1, pp. 218-230.

Stamatakis A, Ludwig T, Meier H 2004, 'RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees', Bioinformatics, vol. 21, no. 4, pp. 456-463.

Steckenreuter A, Moller L, Harcourt R 2012, 'How does Australia's largest dolphin-watching industry affect the behaviour of a small and resident population of Indo-Pacific bottlenose dolphins?', Journal of Environmental Management, vol. 97, pp. 14-21.

Steeman ME, Hebsgaard MB, Fordyce RE, Ho SY, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sørensen MV, Willerslev E 2009, 'Radiation of extant cetaceans driven by restructuring of the oceans', Systematic Biology, vol. 58, no. 6, pp. 573-585.

Stern DL 2013, 'The genetic causes of convergent evolution', Nature Reviews Genetics, vol. 14, no. 11, pp. 751.

Stinchcombe J, Hoekstra H 2008, 'Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits', Journal of Heredity, vol. 100, no. 2, pp. 158-170.

Stobdan T, Zhou D, Ao-Ieong E, Ortiz D, Ronen R, Hartley I, Gan Z, McCulloch AD, Bafna V, Cabrales P 2015, 'Endothelin receptor B, a candidate gene from human studies at high altitude, improves cardiac tolerance to hypoxia in genetically engineered heterozygote mice', Proceedings of the National Academy of Sciences, vol. 112, no. 33, pp. 10425-10430.

Storz JF 1999, 'Genetic consequences of mammalian social structure', Journal of Mammalogy, vol. 80, no. 2, pp. 553-569.

Stroud JT, Losos JB 2016, 'Ecological opportunity and adaptive radiation', Annual Review of Ecology, Evolution, and Systematics, vol. 47, pp. 507-532.

Su M, Zhou J, Duan Z, Zhang J 2019, 'Transcriptional analysis of renal dopamine-mediated Na+ homeostasis response to environmental salinity stress in *Scatophagus argus*', BMC Genomics, vol. 20, no. 1, pp. 418.

Sun Y-B, Zhou W-P, Liu H-Q, Irwin DM, Shen Y-Y, Zhang Y-P 2012, 'Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins', Genome Biology and Evolution, vol. 5, no. 1, pp. 130-139.

Sun Y, Liu C-H, SanGiovanni JP, Evans LP, Tian KT, Zhang B, Stahl A, Pu WT, Kamenecka TM, Solt LA 2015, 'Nuclear receptor RORα regulates pathologic retinal angiogenesis by modulating SOCS3-dependent inflammation', Proceedings of the National Academy of Sciences, vol. 112, no. 33, pp. 10401-10406.

Sunnucks P, Hales DF 1996, 'Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae)', Molecular Biology and Evolution, vol. 13, no. 3, pp. 510-524.

Suthers IM, Young JW, Baird ME, Roughan M, Everett JD, Brassington GB, Byrne M, Condie SA, Hartog JR, Hassler CS 2011, 'The strengthening East Australian Current, its eddies and biological effects—an introduction and overview', Deep-Sea Research II, vol. 58, no. 5, pp. 538-546.

Sydeman WJ, Poloczanska E, Reed TE, Thompson SA 2015, 'Climate change and marine vertebrates', Science, vol. 350, no. 6262, pp. 772-777.

Tassabehji M, Hammond P, Karmiloff-Smith A, Thompson P, Thorgeirsson SS, Durkin ME, Popescu NC, Hutton T, Metcalfe K, Rucka A 2005, 'GTF2IRD1 in craniofacial development of humans and mice', Science, vol. 310, no. 5751, pp. 1184-1187.

Taylor BL, Perrin WF, Reeves RR, Rosel PE, Wang JY, Cipriano F, Scott Baker C, Brownell Jr RL 2017, 'Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans', Marine Mammal Science, vol. 33, no. S1, pp. 12-26.

Taylor EB, McPhail J 1999, 'Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA', Biological Journal of the Linnean Society, vol. 66, no. 3, pp. 271-291.

Teske PR, Sandoval-Castillo J, Golla TR, Emami-Khoyi A, Tine M, von der Heyden S, Beheregaray LB 2019, 'Thermal selection as a driver of marine ecological speciation', Proceedings of the Royal Society B, vol. 286, no. 1896, pp. 20182023.

Teske PR, Sandoval-Castillo J, van Sebille E, Waters J, Beheregaray LB 2015, 'On-shelf larval retention limits population connectivity in a coastal broadcast spawner', Marine Ecology Progress Series, vol. 532, pp. 1-12.

Tezanos-Pinto G, Baker CS, Russell K, Martien K, Baird RW, Hutt A, Stone G, Mignucci-Giannoni AA, Caballero S, Endo T 2009, 'A worldwide perspective on the population structure and genetic diversity of bottlenose dolphins (*Tursiops truncatus*) in New Zealand', Journal of Heredity, vol. 100, no. 1, pp. 11-24.

Tezanos-Pinto G, Constantine R, Brooks L, Jackson JA, Mourão F, Wells S, Scott Baker C 2013, 'Decline in local abundance of bottlenose dolphins (*Tursiops truncatus*) in the Bay of Islands, New Zealand', Marine Mammal Science, vol. 29, no. 4.

The UniProt Consortium 2018, 'UniProt: a worldwide hub of protein knowledge', Nucleic Acids Research, vol. 47, no. D1, pp. D506-D515.

Thewissen J, Bajpai S 2001, 'Whale Origins as a Poster Child for Macroevolution: Fossils collected in the last decade document the ways in which Cetacea (whales, dolphins, and porpoises) became aquatic, a transition that is one of the best documented examples of macroevolution in mammals', BioScience, vol. 51, no. 12, pp. 1037-1049.

Thewissen J, Cooper LN, George JC, Bajpai S 2009, 'From land to water: The origin of whales, dolphins, and porpoises', Evolution: Education and Outreach, vol. 2, no. 2, pp. 272.

Thewissen J, George J, Rosa C, Kishida T 2011, 'Olfaction and brain size in the bowhead whale (*Balaena mysticetus*)', Marine Mammal Science, vol. 27, no. 2, pp. 282-294.

Thewissen JGM, Cooper LN, Clementz MT, Bajpai S, Tiwari BN 2007, 'Whales originated from aquatic artiodactyls in the Eocene epoch of India', Nature, vol. 450, no. 7173, pp. 1190-1194.

Thomson SJ, Hansen A, Sanguinetti MC 2015, 'Identification of the intracellular Na+ sensor in Slo2. 1 potassium channels', Journal of Biological Chemistry, vol. 290, no. 23, pp. 14528-14535.

Tian R, Wang ZF, Niu X, Zhou KY, Xu SX, Yang G 2016, 'Evolutionary genetics of hypoxia tolerance in cetaceans during diving', Genome Biology and Evolution, vol. 8, no. 3, pp. 827-839.

Tiedemann M, Brehmer P 2017, 'Larval fish assemblages across an upwelling front: Indication for active and passive retention', Estuarine, Coastal and Shelf Science, vol. 187, pp. 118-133.

Toischer K, Hartmann N, Wagner S, Fischer TH, Herting J, Danner BC, Sag CM, Hund TJ, Mohler PJ, Belardinelli L 2013, 'Role of late sodium current as a potential arrhythmogenic mechanism in the progression of pressure-induced heart disease', Journal of Molecular and Cellular Cardiology, vol. 61, pp. 111-122.

Tomasello DL (2017) Slick (Kcnt2) Potassium Channels Regulate Pain. State University of New York, Buffalo, NY.

Tomasello DL, Gancarz-Kausch AM, Dietz DM, Bhattacharjee A 2015, 'Transcriptional regulation of the sodium-activated potassium channel SLICK (KCNT2) promoter by nuclear factor-κB', Journal of Biological Chemistry, vol. 290, no. 30, pp. 18575-18583.

Tomasello DL, Hurley E, Wrabetz L, Bhattacharjee A 2017, 'Slick (Kcnt2) sodium-activated potassium channels limit peptidergic nociceptor excitability and hyperalgesia', Journal of Experimental Neuroscience, vol. 11, pp. 1-16.

Treberg JR, Martin RA, Driedzic WR 2003, 'Muscle enzyme activities in a deep-sea squaloid shark, *Centroscyllium fabricii*, compared with its shallow-living relative, *Squalus acanthias*', Journal of Experimental Zoology Part A-Comparative Experimental Biology, vol. 300A, no. 2, pp. 133-139.

Tse W, Chow SC, Wong CK 2012, 'Eel osmotic stress transcriptional factor 1 (Ostf1) is highly expressed in gill mitochondria-rich cells, where ERK phosphorylated', Frontiers In Zoology, vol. 9, no. 1, pp. 3.

Van Citters RL, Franklin DL, Smith Jr OA, Watson NW, Elsner RW 1965, 'Cardiovascular adaptations to diving in the northern elephant seal *Mirounga angustirostris*', Comparative Biochemistry and Physiology, vol. 16, no. 3, pp. 267-276.

Van Gennip SJ, Popova EE, Yool A, Pecl GT, Hobday AJ, Sorte CJB 2017, 'Going with the flow: the role of ocean circulation in global marine ecosystems under a changing climate', Global Change Biology, vol. 23, no. 7, pp. 2602-2617.

Vaz RAN, Lennon GW, Bowers DG 1990, 'Physical behaviour of a large, negative or inverse estuary', Continental Shelf Research, vol. 10, no. 3, pp. 277-304.

Vendrami DL, Telesca L, Weigand H, Weiss M, Fawcett K, Lehman K, Clark MS, Leese F, McMinn C, Moore H 2017, 'RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity', Royal Society Open Science, vol. 4, no. 2, pp. 1-16.

Vermeren M, Lyraki R, Wani S, Airik R, Albagha O, Mort R, Hildebrandt F, Hurd T 2017, 'Osteoclast stimulation factor 1 (Ostf1) KNOCKOUT increases trabecular bone mass in mice', Mammalian Genome, vol. 28, no. 11-12, pp. 498-514.

Vermeulen E, Fruet P, Costa A, Coscarella M, Laporta P (2019) *Tursiops truncatus* ssp. *gephyreus* [Online]. The IUCN Red List of Threatened Species. Available: https://www.iucnredlist.org/species/134822416/135190824 [Accessed 18/01/2020].

Vial G, Dubouchaud H, Couturier K, Cottet-Rousselle C, Taleux N, Athias A, Galinier A, Casteilla L, Leverve XM 2011, 'Effects of a high-fat diet on energy metabolism and ROS production in rat liver', Journal of Hepatology, vol. 54, no. 2, pp. 348-356.

Viaud-Martinez KA, Brownell Jr RL, Komnenou A, Bohonak AJ 2008, 'Genetic isolation and morphological divergence of Black Sea bottlenose dolphins', Biological Conservation, vol. 141, no. 6, pp. 1600-1611.

Vilstrup JT, Ho SY, Foote AD, Morin PA, Kreb D, Krützen M, Parra GJ, Robertson KM, de Stephanis R, Verborgh P 2011, 'Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae', BMC Evolutionary Biology, vol. 11, no. 1, pp. 65.

Vitelli F, Viola A, Morishima M, Pramparo T, Baldini A, Lindsay E 2003, 'TBX1 is required for inner ear morphogenesis', Human Molecular Genetics, vol. 12, no. 16, pp. 2041-2048.

Wahlberg M, Jensen FH, Aguilar Soto N, Beedholm K, Bejder L, Oliveira C, Rasmussen M, Simon M, Villadsgaard A, Madsen PT 2011, 'Source parameters of echolocation clicks from wild bottlenose dolphins (*Tursiops aduncus* and *Tursiops truncatus*)', The Journal of the Acoustical Society of America, vol. 130, no. 4, pp. 2263-2274.

Walker W 1981, 'Geographic variation in morphology and biology of bottlenose dolphins (*Tursiops*) in the eastern North Pacific', NOAA/NMFS Southwest Fisheries Science Center Administrative Report, vol. LJ-81-3c, pp. 21.

Wang G, Yang E, Smith KJ, Zeng Y, Ji G, Connon R, Fangue NA, Cai JJ 2014, 'Gene expression responses of threespine stickleback to salinity: implications for salt-sensitive hypertension', Frontiers in Genetics, vol. 5, pp. 312.

Wang JY, Chou LS, White BN 1999, 'Mitochondrial DNA analysis of sympatric morphotypes of bottlenose dolphins (genus: *Tursiops*) in Chinese waters', Molecular Ecology, vol. 8, no. 10, pp. 1603-1612.

Wang JY, Chou LS, White BN 2000, 'Osteological differences between two sympatric forms of bottlenose dolphins (genus *Tursiops*) in Chinese waters', Journal of Zoology, vol. 252, no. 2, pp. 147-162.

Wang Z, Chen Z, Xu S, Ren W, Zhou K, Yang G 2015, ''Obesity' is healthy for cetaceans? Evidence from pervasive positive selection in genes related to triacylglycerol metabolism', Scientific Reports, vol. 5, pp. 14187.
Wang Z, Xu S, Du K, Huang F, Chen Z, Zhou K, Ren W, Yang G 2016, 'Evolution of digestive enzymes and RNASE1 provides insights into dietary switch of cetaceans', Molecular Biology and Evolution, vol. 33, no. 12, pp. 3144-3157.
Ward TM, McLeay LJ, Dimmlich WF, Rogers PJ, McClatchie S, Matthews R, KÄMpf J, Van Ruth PD 2006, 'Pelagic ecology of a northern boundary current system: effects of upwelling on the production and distribution of sardine (*Sardinops sagax*), anchovy (*Engraulis australis*) and southern bluefin tuna (*Thunnus maccoyii*) in the Great Australian Bight', Fisheries Oceanography, vol. 15, no. 3, pp. 191-207.
Wegiel J, Gong CX, Hwang YW 2011, 'The role of DYRK1A in neurodegenerative diseases', The FEBS Journal, vol. 278, no. 2, pp. 236-245.
Wei X, Cheng Y, Luo Y, Shi X, Nelson S, Hyde DR 2004, 'The zebrafish Pard3 ortholog is required for separation of the eye fields and retinal lamination', Developmental Biology, vol. 269, no. 1, pp. 286-301.
Weiss J 2006, 'Foraging habitats and associated preferential foraging specializations of bottlenose dolphin (*Tursiops truncatus*) mother-calf pairs', Aquatic Mammals, vol. 32, no. 1, pp. 10.
Wellborn GA, Langerhans RB 2015, 'Ecological opportunity and the adaptive diversification of lineages', Ecology and Evolution, vol. 5, no. 1, pp. 176-195.
Wells RS, Rhinehart HL, Cunningham P, Whaley J, Baran M, Koberna C, Costa DP 1999, 'Long distance offshore movements of bottlenose dolphins', Marine Mammal Science, vol. 15, no. 4, pp. 1098-1114.
Werth AJ 2006, 'Mandibular and dental variation and the evolution of suction feeding in Odontoceti', Journal of Mammalogy, vol. 87, no. 3, pp. 579-588.
Werth AJ 2007, 'Adaptations of the cetacean hyolingual apparatus for aquatic feeding and thermoregulation', The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, vol. 290, no. 6, pp. 546-568.
Westgate AJ 2007, 'Geographic variation in cranial morphology of short-beaked common dolphins (*Delphinus delphis*) from the North Atlantic', Journal of Mammalogy, vol. 88, no. 3, pp. 678-688.
Whitlock MC 1999, 'Neutral additive genetic variance in a metapopulation', Genetics Research, vol. 74, no. 3, pp. 215-221.
Whitlock MC 2000, 'Fixation of new alleles and the extinction of small populations: Drift load, beneficial alleles, and sexual selection', Evolution, vol. 54, no. 6, pp. 1855-1861.
Whitmarsh S (2018) Investigating the Anthropocene influence on temperate fish assemblages through baited remote underwater video stations (BRUVS). PhD Thesis, Flinders University, Adelaide, Australia.
Whitmarsh S, Fairweather P, Huveneers C in review, 'Multi-scale analysis of temporal changes in fish assemblages'.
Wickert JC, von Eye SM, Oliveira LR, Moreno IB 2016, 'Revalidation of *Tursiops gephyreus* Lahille, 1908 (Cetartiodactyla: Delphinidae) from the southwestern Atlantic Ocean', Journal of Mammalogy, vol. 97, no. 6, pp. 1728-1737.
Wickham H 2011, 'The split-apply-combine strategy for data analysis', Journal of Statistical Software, vol. 40, no. 1, pp. 1-29.
Willi Y, Van Buskirk J, Schmid B, Fischer M 2007, 'Genetic isolation of fragmented populations is exacerbated by drift and selection', Journal of Evolutionary Biology, vol. 20, no. 2, pp. 534-542.
Williams GC (1996) Adaptation and natural selection: A critique of some current evolutionary thought. Princeton University Press, New Jersey.
Wilson GA, Rannala B 2003, 'Bayesian inference of recent migration rates using multilocus genotypes', Genetics, vol. 163, no. 3, pp. 1177-1191.

Wiszniewski J, Allen S, Möller L 2009, 'Social cohesion in a hierarchically structured embayment population of Indo-Pacific bottlenose dolphins', Animal Behaviour, vol. 77, no. 6, pp. 1449–1457.
Wiszniewski J, Beheregaray L, Allen S, Möller L 2010, 'Environmental and social influences on the genetic structure of bottlenose dolphins (*Tursiops aduncus*) in southeastern Australia', Conservation Genetics, vol. 11, no. 1, pp. 1405–1419.
Wiszniewski J, Brown C, Möller L 2012, 'Complex patterns of male alliance formation in a dolphin social network', Journal of Mammalogy, vol. 93, no. 1, pp. 239-250.
Wood TE, Burke JM, Rieseberg LH 2005, 'Parallel genotypic adaptation: when evolution repeats itself'. In: Mauricio, R (ed.) *Genetics of Adaptation*. Springer, The Netherlands, pp. 157-170.
Wright S 1922, 'Coefficients of inbreeding and relationship', The American Naturalist, vol. 56, no. 645, pp. 330-338.
Wright S 1931, 'Evolution in Mendelian populations', Genetics, vol. 16, no. 2, pp. 97-159.
Wu L, Cai W, Zhang L, Nakamura H, Timmermann A, Joyce T, McPhaden MJ, Alexander M, Qiu B, Visbeck M 2012, 'Enhanced warming over the global subtropical western boundary currents', Nature Climate Change, vol. 2, no. 3, pp. 161.
Xia J-H, Jiang J, Shi Y-H, Gui J-F 2009, 'Predominant expression and cellular distribution of fish Agr2 in renal collecting system', Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, vol. 152, no. 4, pp. 397-404.
Xu H, Morishima M, Wylie JN, Schwartz RJ, Bruneau BG, Lindsay EA, Baldini A 2004, 'Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract', Development, vol. 131, no. 13, pp. 3217-3227.
Yamaguchi M, Nakayama T, Fu Z, Sato N, Soma M, Morita A, Hinohara S, Doba N, Mizutani T 2010, 'The haplotype of the CACNA1B gene associated with cerebral infarction in a Japanese population', Hereditas, vol. 147, no. 6, pp. 313-319.
Yamamoto T, Izumi-Yamamoto K, Iizuka Y, Shirota M, Nagase M, Fujita T, Gotoda T 2013, 'A novel link between *Slc22a18* and fat accumulation revealed by a mutation in the spontaneously hypertensive rat', Biochemical and Biophysical Research Communications, vol. 440, no. 4, pp. 521-526.
Yang TH, Lai NC, Graham JB, Somero GN 1992, 'Respiratory, blood, and heart enzymatic adaptations of *Sebastolobus alascanus* (Scorpaenidae; Teleostei) to the oxygen minimum zone: A comparative study', Biological Bulletin, vol. 183, no. 3, pp. 490-499.
Yang W-K, Chung C-H, Cheng HC, Tang C-H, Lee T-H 2016, 'Different expression patterns of renal Na+/K+-ATPase α-isoform-like proteins between tilapia and milkfish following salinity challenges', Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, vol. 202, pp. 23-30.
Yim HS, Cho YS, Guang XM, Kang SG, Jeong JY, Cha SS, Oh HM, Lee JH, Yang EC, Kwon KK, Kim YJ, Kim TW, Kim W, Jeon JH, Kim SJ, Choi DH, Jho S, Kim HM, Ko J, Kim H, Shin YA, Jung HJ, Zheng Y, Wang Z, Chen Y, Chen M, Jiang AW, Li EL, Zhang S, Hou HL, Kim TH, Yu LL, Liu S, Ahn K, Cooper J, Park SG, Hong CP, Jin W, Kim HS, Park C, Lee K, Chun S, Morin PA, O'Brien SJ, Lee H, Kimura J, Moon DY, Manica A, Edwards J, Kim BC, Kim S, Wang J, Bhak J, Lee HS, Lee JH 2014, 'Minke whale genome and aquatic adaptation in cetaceans', Nature Genetics, vol. 46, no. 1, pp. 88-92.
Yuan Y, Zhang PJ, Wang K, Liu MZ, Li J, Zheng JS, Wang D, Xu WJ, Lin ML, Dong LJ, Zhu CL, Qiu Q, Li SH 2018, 'Genome sequence of the freshwater Yangtze finless porpoise', Genes, vol. 9, no. 4, pp. 9.
Zanardo N, Bilgmann K, Parra G, Möller L 2016a, 'Socio-genetic structure of short-beaked common dolphins in southern Australia', Journal of Zoology, vol. 299, no. 2, pp. 89-97.
Zanardo N, Parra GJ, Diaz-Aguirre F, Pratt EAL, Möller LM 2018, 'Social cohesion and intra-population community structure in southern Australian bottlenose dolphins (*Tursiops* sp.)', Behavioral Ecology and Sociobiology, vol. 72, no. 9, pp. 156.

Zanardo N, Parra GJ, Möller LM 2016b, 'Site fidelity, residency, and abundance of bottlenose dolphins (*Tursiops* sp.) in Adelaide's coastal waters, South Australia', Marine Mammal Science, vol. 32, no. 4, pp. 1381-1401.

Zeng J, Zhou S-W, Zhao J, Jin M-H, Kang D-J, Yang Y-X, Wang X-L, Chen Y-L 2019, 'Role of OXCT1 in ovine adipose and preadipocyte differentiation', Biochemical and Biophysical Research Communications, vol. 512, no. 4, pp. 779-785.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW 2014, 'Comparative genomics reveals insights into avian genome evolution and adaptation', Science, vol. 346, no. 6215, pp. 1311-1320.

Zhang M, Hu C, Tong D, Xiang S, Williams K, Bai W, Li G-M, Bepler G, Zhang X 2016, 'Ubiquitin-specific peptidase 10 (USP10) deubiquitinates and stabilizes MutS homolog 2 (MSH2) to regulate cellular sensitivity to DNA damage', Journal of Biological Chemistry, vol. 291, no. 20, pp. 10783-10791.

Zhang X, Wang D, Dongye M, Zhu Y, Chen C, Wang R, Long E, Liu Z, Wu X, Lin D 2019, 'Loss-of-function mutations in FREM2 disrupt eye morphogenesis', Experimental Eye Research, vol. 181, pp. 302-312.

Zhivotovsky LA 2015, 'Relationships between Wright’s FST and FIS statistics in a context of Wahlund effect', Journal of Heredity, vol. 106, no. 3, pp. 306-309.

Zhou X, Seim I, Gladyshev VN 2015, 'Convergent evolution of marine mammals is associated with distinct substitutions in common genes', Scientific Reports, vol. 5, pp. 16550.

Zhou X, Sun D, Guang X, Ma S, Fang X, Mariotti M, Nielsen R, Gladyshev VN, Yang G 2018a, 'Molecular footprints of aquatic adaptation including bone mass changes in cetaceans', Genome Biology and Evolution, vol. 10, no. 3, pp. 967-975.

Zhou XM, Guang XM, Sun D, Xu SX, Li MZ, Seim I, Jie WC, Yang LF, Zhu QH, Xu JB, Gao Q, Kaya A, Dou QH, Chen BY, Ren WH, Li SC, Zhou KY, Gladyshev VN, Nielsen R, Fang XD, Yang G 2018b, 'Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater', Nature Communications, vol. 9.

Zhou XM, Sun FM, Xu SX, Fan GY, Zhu KL, Liu X, Chen Y, Shi CC, Yang YX, Huang ZY, Chen J, Hou HL, Guo XJ, Chen WB, Chen YF, Wang XH, Lv T, Yang D, Zhou JJ, Huang BQ, Wang ZF, Zhao W, Tian R, Xiong ZQ, Xu JX, Liang XM, Chen BY, Liu WQ, Wang JY, Pan SK, Fang XD, Li M, Wei FW, Xu X, Zhou KY, Wang J, Yang G 2013, 'Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations', Nature Communications, vol. 4, no. 1, pp. 1-6.

Zhuang P, Lu YH, Shou QY, Mao L, He LL, Wang J, Chen JN, Zhang Y, Jiao JJ 2019, 'Differential anti-adipogenic effects of eicosapentaenoic and docosahexaenoic acids in obesity', Molecular Nutrition & Food Research, vol. 63, no. 14, pp. 11.

Appendices

A. Abbreviations

Abbreviation	Meaning
ACC	Antarctic Circumpolar Current
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
ASPSP	St. Peter and St. Paul Archipelago
BB	Byron Bay
BI	Broughton Island
BIC	Bayesian information criterion
bp	Base pairs
BLAST	Basic local alignment search tool
B-Y correction	Benjamini and Yekutieli's correction
CB	Coffin Bay
CBI	Coffin Bay inner
CBO	Coffin Bay outer
ChloA	Chlorophyll A concentration
CJ	Cape Jervis
CV error	Cross validation error
DAPC	Discriminant analysis of principal components
ddRAD	Double-digest restriction site-associated DNA
ddRADseq	Double-digest restriction site-associated DNA sequencing
DMSO	Dimethyl sulphoxide
DHA	Docosahexaenoic acid
e-value	Expectation value
EAC	East Australian Current
EPA	Eicosapentaenoic acid
FDR	False discovery rate
GEA	Genotype-environment association
GO	Gene ontology
GSV	Gulf St. Vincent
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy-Weinberg Equilibrium
IBD	Isolation by distance
INDEL	Insertions and deletions
IPBD	Indo-Pacific bottlenose dolphin (*T. aduncus*)
IUCN	International Union for Conservation of Nature
JB	Jervis Bay
K	Number of populations
KB	Kilobases (i.e. one thousand base pairs)
LGM	Last Glacial Maximum
MCMC	Markov chain Monte Carlo
MEM	Moran's eigenvector map
mtDNA	Mitochondrial DNA
MYA	Million years ago
N	Sample size
N_e	Effective population size

NCBI	National Center for Biotechnology Information
NSG	Northern Spencer Gulf
NSW	New South Wales
PA	Private alleles
PCA	Principal component analysis
PCR	Polymerase chain reaction
%PL	Percentage of polymorphic loci
PP	Primary productivity
PS	Port Stephens
PW	Port Wakefield
qPCR	Quantitative polymerase chain reaction
RDA	Redundancy analysis
RELL	Resampling estimated log-likelihood
SABD	Southern Australian bottlenose dolphin (*T. australis*)
SAHMRI	South Australian Health and Medical Research Institute
SD	Standard deviations
SFI	St. Francis Island
SG	Spencer Gulf
SNP	Single nucleotide polymorphism
SST	Sea surface temperature
SWAO	Southwestern Atlantic Ocean
VCF	Variant call file
VIF	Variance inflation factor

B. Chapter 2 Appendix

B.i. Figures

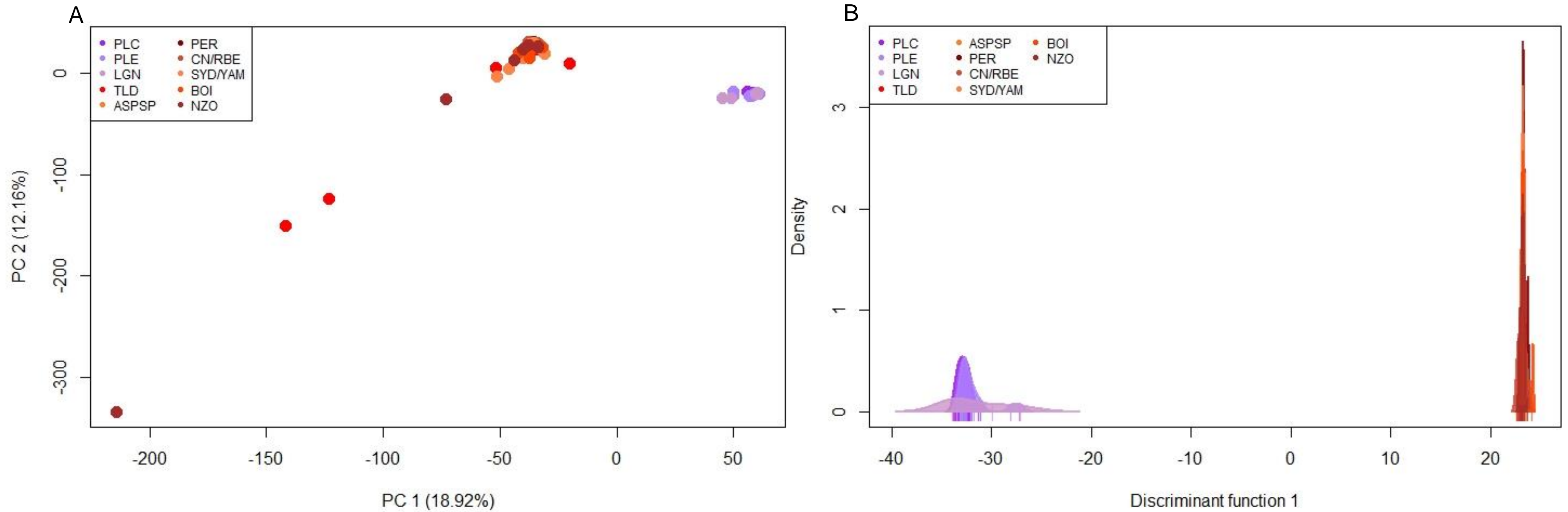


Figure B.i.1 Population genomic structure of bottlenose dolphins (*T. t. truncatus* (red shades) and *T. t. gephyreus* (purple shades)) as estimated by **A**) principal component analysis (PCA) and **B**) discriminant analysis of principal components (DAPC), with two clusters being the most supported number of populations. Sampling locations abbreviations are explained in Appendix B: Table B.ii.1.

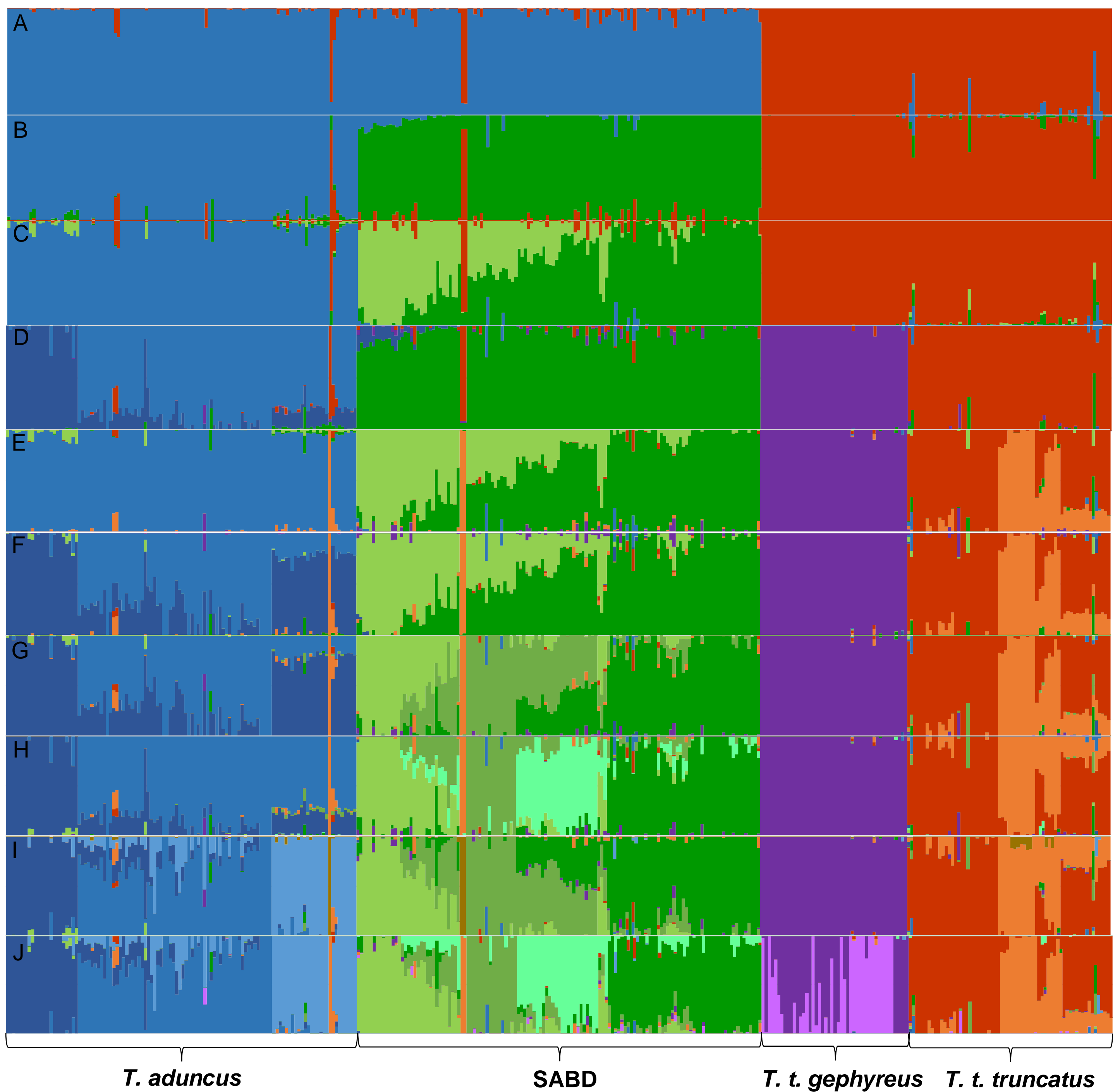


Figure B.i.2 Population genomic structure of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere as estimated by *Admixture*. Plots **A-J** represent scenarios of 2-11 populations (*K*). High statistical support was provided for *K* between six (**E**) and 11 (**J**), while eight (**G**) was the most highly supported *K* value. Colour groups are as per the putative lineage of membership.

B.ii. Tables

Table B.ii.1 Sampling regime breakdown for genomic analysis of bottlenose dolphins (*Tursiops* spp.) across the Southern Hemisphere. Note: sample sizes (N) reported here are before filtering for low number of reads and high missing data in the bioinformatics process.

Putative Taxa	Ocean	Site	Abbreviation	Year Sampled	N
T. aduncus	South Pacific	Byron Bay/Ballina	BB	2007	11
		Yamba	YAM	2005-2007	12
		Forster	FOR	2001-2004	7
		Broughton Island	BI	1999-2004	16
		Port Stephens	PS	1999-2004	28
		Newcastle	NC	2001-2005	11
		Jervis Bay	JB	1998-1999	14
		Eden	ED	2006-2007	13
SABD	South Indian	Albany/Esperance (Western Australia)	WA	2010	14
		St. Francis Island	SFI	2005	11
		Coffin Bay outer	CBO	2014	13
		Coffin Bay inner	CBI	2014	16
		Pt. Lincoln	PL	2005	14
		Northern Spencer Gulf	NSG	2005	11
		Southeastern Spencer Gulf	SESG	2004	11
		Stansbury	SB	2015	12
		Pt. Wakefield	PW	2015	13
		Adelaide	ADE	2013-2015	16
		Cape Jervis	CJ	2015	11
T. t. gephyreus	South Atlantic	Patos Lagoon coastal community	PLC	2003-2012	27
		Patos Lagoon estuarine community	PLE	2003-2012	16
		Laguna	LGN	2011	7
T. t. truncatus	South Atlantic	Talude	TLD	2009-2012	26
		St. Peter and St. Paul Archipelago	ASPSP	2006-2009	7
	South Indian	Perth	PER	2010	4
		Cape Nelson/Robe	CN	2012	8
	South Pacific	Sydney/Yamba	SYD	2005-2007	8
		Bay of Islands	BOI	2003-2005	9
		NZ Offshore (North Island)	NZO	2015-2016	9

Table B.ii.2 Parameters and results of the quality filtering process for the double-digest restriction site-associated sequencing (ddRADseq) of the *Tursiops* dataset and the dataset including nine common dolphins (*Delphinus delphis*) for use in phylogenomic analyses (*D. delphis* as outgroup).

		Tursiops **Dataset**	***Tursiops*** **+** ***Delphinus*** **Dataset**
	Average no. of retained reads per individual	~2,465,060	~2,444,334
	No. of SNPs before filtering	196,751	223,408
	Sample size before filtering	375 individuals	386 individuals
Filtering Parameter	<50% missing data and >500,000 reads per individual (final sample size)	353 individuals	362 individuals
	Variants called in >20% of individuals, minor allele frequency (MAF) >0.02, minimum quality score (MinQ) >30 & biallelic SNPs only	49,471 SNPs	56,640 SNPs
	Removing complex variance and INDELs	48,723 SNPs	53,770 SNPs
	Av. coverage < av. depth + 2 standard deviations	47,459 SNPs	52,522 SNPs
	Allele balance 0.20 - 0.80 (proportion of alternate to reference alleles)	46,264 SNPs	50,130 SNPs
	Mapping quality score 0.9 - 1.05	45,264 SNPs	48,619 SNPs
	Read quality score >20% of read depth	44,667 SNPs	47,951 SNPs
	<10% missing data per variant	34,735 SNPs	37,326 SNPs
	<19% of populations out of HWE - default of <25% used for *Tursiops* + *Delphinus* dataset	34,140 SNPs	36,813 SNPs
	Best quality SNP per locus	23,224 SNPs	24,747 SNPs
	Linkage Disequilibrium - r^2 <0.8	18,112 SNPs	18,338 SNPs
	Aligned to *T. aduncus* genome (final number of SNPs)	18,060 SNPs	18,282 SNPs

Table B.ii.3 Genomic diversity indices for sampling localities of *Tursiops* spp. across the Southern Hemisphere, whereby N = sample size before filtering, %PL = percentage of polymorphic loci, H_O = observed heterozygosity, H_E = expected heterozygosity, F_{IS} = Wright's inbreeding coefficient and PA = number of private alleles. Standard deviations are given in parentheses.

		N	%PL	H_O	H_E	F_{IS}	PA
T. aduncus	Byron Bay (BB)	11	24.787	0.321 (0.182)	0.333 (0.151)	0.036	
	Yamba (YAM)	12	26.912	0.305 (0.179)	0.325 (0.154)	0.062	
	Forster (FOR)	7	23.499	0.338 (0.193)	0.356 (0.144)	0.051	
	Broughton Island (BI)	16	29.882	0.294 (0.183)	0.296 (0.162)	0.007	
	Port Stephens (PS)	28	29.616	0.301 (0.195)	0.286 (0.171)	-0.052	
	Newcastle (NC)	11	24.503	0.333 (0.185)	0.337 (0.151)	0.012	
	Jervis Bay (JB)	14	28.243	0.313 (0.187)	0.306 (0.159)	-0.023	
	Eden (ED)	13	32.429	0.241 (0.176)	0.298 (0.153)	0.191	
	Average/Total	**112**	**27.484**	**0.304 (0.185)**	**0.315 (0.156)**	**0.035**	**512**
SABD	Albany/Esperance (WA)	14	29.505	0.293 (0.179)	0.304 (0.158)	0.036	
	St. Francis Island (SFI)	8	27.266	0.302 (0.189)	0.339 (0.148)	0.109	
	Coffin Bay outer (CBO)	13	32.624	0.249 (0.172)	0.327 (0.147)	0.239	
	Coffin Bay inner (CBI)	16	27.947	0.318 (0.193)	0.305 (0.163)	-0.043	
	Pt. Lincoln (PL)	14	28.725	0.308 (0.189)	0.310 (0.159)	0.006	
	Northern SG (NSG)	6	23.940	0.342 (0.198)	0.353 (0.146)	0.031	
	Southeastern SG (SESG)	9	29.326	0.293 (0.187)	0.326 (0.153)	0.101	
	Stansbury (SB)	10	26.005	0.322 (0.190)	0.322 (0.155)	0.000	
	Pt. Wakefield (PW)	12	26.121	0.305 (0.185)	0.312 (0.160)	0.022	
	Adelaide (ADE)	16	31.443	0.302 (0.195)	0.285 (0.165)	-0.060	
	Cape Jervis (CJ)	11	29.693	0.327 (0.203)	0.311 (0.159)	-0.051	
	Average/Total	**129**	**28.418**	**0.306 (0.189)**	**0.318 (0.156)**	**0.036**	**370**
T. t. gephyreus	Patos Lagoon coastal (PLC)	27	3.284	0.290 (0.233)	0.257 (0.182)	-0.128	
	Patos Lagoon estuary (PLE)	15	6.917	0.160 (0.172)	0.154 (0.145)	-0.039	
	Laguna (LGN)	5	4.244	0.258 (0.168)	0.285 (0.127)	0.095	
	Average/Total	**47**	**4.815**	**0.236 (0.191)**	**0.232 (0.151)**	**-0.024**	**53**
T. t. truncatus	Talude (TLD)	22	33.579	0.264 (0.194)	0.258 (0.174)	-0.023	
	St. Peter and St. Paul Archipelago (ASPSP)	7	18.169	0.400 (0.206)	0.376 (0.138)	-0.064	
	Perth (PER)	4	15.866	0.411 (0.232)	0.418 (0.127)	0.017	
	Cape Nelson/Robe (CN)	8	21.169	0.344 (0.204)	0.357 (0.147)	0.036	
	Sydney/Yamba (SYD)	8	25.824	0.267 (0.173)	0.329 (0.153)	0.188	
	Bay of Islands (BOI)	8	23.328	0.357 (0.200)	0.351 (0.147)	-0.017	
	NZ Offshore North Island (NZO)	8	33.776	0.312 (0.203)	0.304 (0.152)	-0.026	
	Average/Total	**65**	**24.530**	**0.336 (0.202)**	**0.342 (0.148)**	**0.016**	**642**

Table B.ii.4 Analysis of molecular variance (AMOVA) testing the amount of genomic variation explained by the division into four putative bottlenose dolphin lineages in the Southern Hemisphere compared to among individual populations across the study region. An asterisk (*) denotes a significant p-value of <0.001.

Source of Variation	Degrees of Freedom	Sum of Squares	Variance Components	Percentage of Variation
Among lineages	3	279,536.95	540.49	42.81*
Among populations within each lineage	25	47,548.96	52.99	4.20*
Within populations	677	452,904.56	668.99	52.99*
Total	705	779,990.47	1262.48	

Table B.ii.5 Significantly over-enriched (p <0.05) gene ontology (GO) terms in Southern Hemisphere inshore bottlenose dolphin (*Tursiops* spp.) ecotype parallel evolution candidate loci dataset, compared to the full (reference) dataset as identified by a functional enrichment analysis using a Fisher's exact test in Blast2GO.

GO Category	GO ID	GO Term	*p*	Associated genes in: Candidate Dataset (%)	Associated genes in: Reference Dataset (%)
Biological Process	GO:0042307	positive regulation of protein import into nucleus	0.0463	0.7092	0.0279
	GO:0090090	negative regulation of canonical Wnt signaling pathway	0.0488	1.4286	0.2406
	GO:0031333	negative regulation of protein complex assembly	0.0292	1.4286	0.1789
	GO:0042074	cell migration involved in gastrulation	0.0387	0.7092	0.0223
	GO:0030206	chondroitin sulfate biosynthetic process	0.0463	0.7092	0.0279
	GO:0030205	dermatan sulfate metabolic process	0.0234	0.7092	0.0112
	GO:0030208	dermatan sulfate biosynthetic process	0.0234	0.7092	0.0112
	GO:0030207	chondroitin sulfate catabolic process	0.0387	0.7092	0.0223
	GO:0030203	glycosaminoglycan metabolic process	0.0035	2.1583	0.2013
	GO:0042147	retrograde transport, endosome to Golgi	0.0164	1.4286	0.1285
	GO:0043484	regulation of RNA splicing	0.0292	1.4286	0.1789
	GO:0030166	proteoglycan biosynthetic process	0.0139	1.4286	0.1173
	GO:0016482	cytosolic transport	0.0342	1.4286	0.1957
	GO:0090389	phagosome-lysosome fusion involved in apoptotic cell clearance	0.0079	0.7092	0.0000
	GO:0090386	phagosome maturation involved in apoptotic cell clearance	0.0079	0.7092	0.0000
	GO:0090387	phagolysosome assembly involved in apoptotic cell clearance	0.0079	0.7092	0.0000
	GO:0090385	phagosome-lysosome fusion	0.0157	0.7092	0.0056
	GO:0015015	heparan sulfate proteoglycan biosynthetic process, enzymatic modification	0.0234	0.7092	0.0112
	GO:0042989	sequestering of actin monomers	0.0079	0.7092	0.0000
	GO:0038128	ERBB2 signaling pathway	0.0311	0.7092	0.0167
	GO:0090659	walking behavior	0.0463	0.7092	0.0279
	GO:1990569	UDP-N-acetylglucosamine transmembrane transport	0.0234	0.7092	0.0112
	GO:0072180	mesonephric duct morphogenesis	0.0311	0.7092	0.0167
	GO:0072138	mesenchymal cell proliferation involved in ureteric bud development	0.0387	0.7092	0.0223
	GO:0001732	formation of cytoplasmic translation initiation complex	0.0311	0.7092	0.0167

	GO:0061209	cell proliferation involved in mesonephros development	0.0463	0.7092	0.0279
	GO:0000290	deadenylation-dependent decapping of nuclear-transcribed mRNA	0.0234	0.7092	0.0112
	GO:0050651	dermatan sulfate proteoglycan biosynthetic process	0.0234	0.7092	0.0112
	GO:0050655	dermatan sulfate proteoglycan metabolic process	0.0234	0.7092	0.0112
	GO:0034356	NAD biosynthesis via nicotinamide riboside salvage pathway	0.0387	0.7092	0.0223
	GO:0072334	UDP-galactose transmembrane transport	0.0079	0.7092	0.0000
	GO:0033120	positive regulation of RNA splicing	0.0387	0.7092	0.0223
	GO:0010607	negative regulation of cytoplasmic mRNA processing body assembly	0.0079	0.7092	0.0000
	GO:0010603	regulation of cytoplasmic mRNA processing body assembly	0.0311	0.7092	0.0167
	GO:0045807	positive regulation of endocytosis	0.0342	1.4286	0.1957
	GO:0033962	cytoplasmic mRNA processing body assembly	0.0311	0.7092	0.0167
	GO:0018298	protein-chromophore linkage	0.0234	0.7092	0.0112
	GO:0008643	carbohydrate transport	0.0449	1.4286	0.2293
	GO:0006047	UDP-N-acetylglucosamine metabolic process	0.0311	0.7092	0.0167
	GO:0006024	glycosaminoglycan biosynthetic process	0.0015	2.1583	0.1453
	GO:0006022	aminoglycan metabolic process	0.0041	2.1583	0.2125
	GO:0006023	aminoglycan biosynthetic process	0.0018	2.1583	0.1565
	GO:0006029	proteoglycan metabolic process	0.0246	1.4286	0.1621
	GO:0006026	aminoglycan catabolic process	0.0151	1.4286	0.1229
	GO:0006027	glycosaminoglycan catabolic process	0.0151	1.4286	0.1229
Cellular Component	GO:0070993	translation preinitiation complex	0.0387	0.7092	0.0223
	GO:0030120	vesicle coat	0.0096	1.4286	0.0950
	GO:0030125	clathrin vesicle coat	0.0027	1.4286	0.0447
	GO:0030118	clathrin coat	0.0151	1.4286	0.1229
	GO:0030117	membrane coat	0.0468	1.4286	0.2350
	GO:0005794	Golgi apparatus	0.0194	10.0775	5.0108
	GO:0005796	Golgi lumen	0.0128	1.4286	0.1117
	GO:0005775	vacuolar lumen	0.0308	1.4286	0.1845
	GO:0005852	eukaryotic translation initiation factor 3 complex	0.0387	0.7092	0.0223
	GO:0016282	eukaryotic 43S preinitiation complex	0.0311	0.7092	0.0167
	GO:0030662	coated vesicle membrane	0.0449	1.4286	0.2293
	GO:0030665	clathrin-coated vesicle membrane	0.0151	1.4286	0.1229

	GO:0030904	retromer complex	0.0038	1.4286	0.0558
	GO:1990578	perinuclear endoplasmic reticulum membrane	0.0234	0.7092	0.0112
	GO:0048475	coated membrane	0.0468	1.4286	0.2350
	GO:0033290	eukaryotic 48S preinitiation complex	0.0311	0.7092	0.0167
	GO:0044431	Golgi apparatus part	0.0448	7.5758	3.9267
	GO:0031225	anchored component of membrane	0.0488	1.4286	0.2406
	GO:0043202	lysosomal lumen	0.0151	1.4286	0.1229
	GO:0033588	Elongator holoenzyme complex	0.0157	0.7092	0.0056
	GO:0043033	isoamylase complex	0.0079	0.7092	0.0000
Molecular Function	GO:0005338	nucleotide-sugar transmembrane transporter activity	0.0311	0.7092	0.0167
	GO:0102500	beta-maltose 4-alpha-glucanotransferase activity	0.0079	0.7092	0.0000
	GO:0043559	insulin binding	0.0157	0.7092	0.0056
	GO:0004117	calmodulin-dependent cyclic-nucleotide phosphodiesterase activity	0.0157	0.7092	0.0056
	GO:0005462	UDP-N-acetylglucosamine transmembrane transporter activity	0.0234	0.7092	0.0112
	GO:0004133	glycogen debranching enzyme activity	0.0079	0.7092	0.0000
	GO:0004135	amylo-alpha-1,6-glucosidase activity	0.0079	0.7092	0.0000
	GO:0004134	4-alpha-glucanotransferase activity	0.0079	0.7092	0.0000
	GO:0005459	UDP-galactose transmembrane transporter activity	0.0079	0.7092	0.0000
	GO:0017095	heparan sulfate 6-O-sulfotransferase activity	0.0157	0.7092	0.0056
	GO:0015165	pyrimidine nucleotide-sugar transmembrane transporter activity	0.0311	0.7092	0.0167
	GO:1905394	retromer complex binding	0.0387	0.7092	0.0223
	GO:0090599	alpha-glucosidase activity	0.0079	0.7092	0.0000
	GO:0015926	glucosidase activity	0.0234	0.7092	0.0112
	GO:0048101	calcium- and calmodulin-regulated 3',5'-cyclic-GMP phosphodiesterase activity	0.0157	0.7092	0.0056
	GO:0052856	NADHX epimerase activity	0.0157	0.7092	0.0056
	GO:0052857	NADPHX epimerase activity	0.0157	0.7092	0.0056
	GO:0060422	peptidyl-dipeptidase inhibitor activity	0.0311	0.7092	0.0167
	GO:0032027	myosin light chain binding	0.0387	0.7092	0.0223
	GO:0009881	photoreceptor activity	0.0311	0.7092	0.0167
	GO:0043125	ErbB-3 class receptor binding	0.0079	0.7092	0.0000
	GO:0070853	myosin VI binding	0.0157	0.7092	0.0056

GO:0070856	myosin VI light chain binding	0.0157	0.7092	0.0056
GO:0008607	phosphorylase kinase regulator activity	0.0079	0.7092	0.0000

Table B.ii.6 Functional annotation and gene ontology (GO) terms associated with candidate loci identified as outliers in all three inshore - offshore comparisons in *Arlequin* and RandomForest and thus, potentially involved in the parallel evolution of the inshore bottlenose dolphin (*Tursiops* spp.) ecotype in the Southern Hemisphere.

Candidate Locus	Average F_{ST}	e-value	% Identity	Candidate Gene	Protein	GO Terms	
						Molecular Function	Biological Process
42	0.341	4.00E-164	86.77	*LOC109551830* (exonic)	No available information		
211	0.612	3.00E-37	100	*PATL2* (exonic)	Protein PAT1 homolog 2	RNA binding	negative regulation of translation; cytoplasmic mRNA processing body assembly; deadenylation-dependent decapping of nuclear-transcribed mRNA; negative regulation of cytoplasmic mRNA processing body assembly
317	0.292	3.00E-37	100	*LOC101331711 (NSG1*-like*)*	neuron-specific protein family member 1	clathrin light chain binding	clathrin coat assembly; dopamine receptor signaling pathway
380	0.494	1.00E-37	100	*LOC109549816*	No available information		
623	0.450	6.60E-01	68.8	*KIF13A*	Kinesin-like protein KIF13A	ATPase activity; ATP binding; microtubule binding; microtubule motor activity	cell cycle; cell division; endosome to lysosome transport; Golgi to plasma membrane protein transport; intracellular protein transport; melanosome organization; microtubule-based movement; plus-end-directed vesicle transport along microtubule; regulation of cytokinesis; vesicle cargo loading
840	0.534	9.00E-38	100	*PGPEP1L*	Pyroglutamyl-peptidase 1-like protein	cysteine-type peptidase activity	proteolysis
1430	0.665	7.00E-39	100	*NKX2-3*	Homeobox protein nkx-2.3	DNA-binding transcription factor activity; DNA-binding transcription factor activity, RNA polymerase II-specific; RNA polymerase II proximal promoter sequence-specific DNA binding; sequence-specific DNA binding	CD4-positive, alpha-beta T cell differentiation; cell differentiation; gland morphogenesis; leukocyte homeostasis; leukocyte migration; lymph node development; macrophage differentiation; odontogenesis of dentin- containing tooth; Peyer's patch development; plasma cell differentiation; positive regulation of transcription by RNA polymerase II; post-embryonic digestive tract morphogenesis; regulation of cell population proliferation; spleen development; triglyceride metabolic process

2123	0.289	2.00E-39	100	*IL2RB*	Interleukin-2 receptor subunit beta	interleukin-2 receptor activity; interleukin-2 binding; interleukin-15 receptor activity	cytokine-mediated signaling pathway; interleukin-15-mediated signaling pathway; interleukin-2-mediated signaling pathway; MAPK cascade; negative regulation of apoptotic process; positive regulation of phagocytosis; protein-containing complex assembly; signal transduction; viral process
2139	0.299	7.00E-39	100	*ZCWPW2*	Zinc finger CW-type PWWP domain protein 2	zinc ion binding	None given
2148	0.703	7.00E-39	100	*SHROOM4*	protein shroom4	actin filament binding; myosin II binding	brain development; cognition; actin filament organization; actin cytoskeleton organization
2892	0.365	3.00E-38	100	*LOC109547652* (exonic)	No available information		
2963	0.334	2.00E-39	100	*PDE9A*	High affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A	metal ion binding; 3',5'-cyclic-GMP phosphodiesterase activity; 3',5'-cyclic-nucleotide phosphodiesterase activity; identical protein binding	cGMP catabolic process; cGMP-mediated signaling; cGMP metabolic process; positive regulation of cardiac muscle hypertrophy; signal transduction
3165	0.350	2.00E-39	100	*RORA*	Nuclear receptor ROR-alpha	beta-catenin binding; DNA binding; ligand-activated transcription factor activity; nuclear receptor activity; oxysterol binding; RNA polymerase II proximal promoter sequence-specific DNA binding; RNA polymerase II regulatory region sequence-specific DNA binding; steroid hormone receptor activity; transcription coactivator binding; transcription corepressor binding; zinc ion binding	angiogenesis; cellular response to hypoxia; cellular response to tumor necrosis factor; cerebellar granule cell precursor proliferation; cerebellar Purkinje cell differentiation; cGMP metabolic process; cholesterol homeostasis; circadian regulation of gene expression; intracellular receptor signaling pathway; muscle cell differentiation; negative regulation of fat cell differentiation; negative regulation of inflammatory response; nitric oxide biosynthetic process; positive regulation of circadian rhythm; positive regulation of vascular endothelial growth factor production; regulation of glucose metabolic process; regulation of transcription, DNA-templated; regulation of transcription involved in cell fate commitment; T-helper 17 cell differentiation; triglyceride homeostasis
3278	0.353	7.00E-39	100	*IFT43*	Intraflaggellar transport protein 43 homolog	None given	cilium assembly; intraciliary retrograde transport; intraciliary transport involved in cilium assembly
4248	0.332	2.00E-39	100	*PKIB*	cAMP-dependent protein kinase inhibitor beta	cAMP-dependent protein kinase inhibitor activity	negative regulation of cAMP-dependent protein kinase activity; positive regulation of telomerase activity; positive regulation of telomere capping; positive regulation of telomere maintenance via telomerase

4893	0.462	3.00E-38	100	*RBM20*	RNA-binding protein 20	RNA binding; zinc ion binding	heart development; positive regulation of RNA splicing; mRNA processing
4906	0.483	2.00E-39	100	*LOC101334605*	No available information		
5223	0.541	4.70E-01	68.8	*TREES_T100007710*	Extracellular matrix protein FRAS1	structural constituent of ribosome	cell communication; translation
5666	0.261	2.00E-39	100	*LOC101322629* (*COX8A*-like)	cytochrome c oxidase subunit 8A, mitochondrial	cytochrome-c oxidase activity	generation of precursor metabolites and energy; mitochondrial electron transport, cytochrome c to oxygen
5718	0.431	2.00E-39	100	*IMPDH1*	Inosine-5'-monophosphate dehydrogenase 1	DNA binding; IMP dehydrogenase activity; metal ion binding; nucleic acid binding; nucleotide binding; RNA binding	GMP biosynthetic process; GTP biosynthetic process; lymphocyte proliferation; neutrophil degranulation; purine ribonucleotide monophosphate biosynthetic process
5718	0.431	2.00E-39	100	*PRRT4*	Proline-rich transmembrane protein 4	None given	None given
5746	0.448	3.00E-37	98.88	*TG*	Thyroglobulin	hormone activity	hormone biosynthetic process; iodide transport; regulation of myelination; signal transduction; thyroid gland development; thyroid hormone metabolic process
6505	0.658	7.00E-39	100	*EDA*	Ectodysplasin-A	death receptor agonist activity; death receptor binding; signaling receptor binding; tumor necrosis factor receptor binding	cell differentiation; cell-matrix adhesion; cytokine-mediated signaling pathway; gene expression; hair follicle placode formation; immune response; odontogenesis of dentin-containing tooth; pigmentation; positive regulation of canonical Wnt signaling pathway; positive regulation of gene expression; positive regulation of I-kappaB kinase/NF-kappaB signaling; positive regulation of NF-kappaB transcription factor activity; positive regulation of NIK/NF-kappaB signaling; regulation of NIK/NF-kappaB signaling; salivary gland cavitation; trachea gland development; tumor necrosis factor-mediated signaling pathway
6922	0.341	7.00E-39	100	*LOC101339008*	No available information		

7219	0.325	2.00E-39	100	*IFNGR*	Interferon gamma receptor 1	cytokine binding; cytokine receptor activity; interferon-gamma receptor activity	astrocyte activation; cytokine-mediated signaling pathway; interferon-gamma-mediated signaling pathway; microglial cell activation; negative regulation of amyloid-beta clearance; positive regulation of amyloid-beta formation; positive regulation of gene expression; positive regulation of tumor necrosis factor secretion; regulation of interferon-gamma-mediated signaling pathway; response to virus; signal transduction
7309	0.392	2.00E-39	100	*NMRAL1*	NmrA-like family domain-containing protein 1	identical protein binding	urea cycle; nitrogen catabolite repression of transcription from RNA polymerase II promoter
7957	0.375	2.00E-39	100	*FREM2*	FRAS1-related extracellular matrix protein 2	metal ion binding	cell adhesion; cell communication; embryonic digit morphogenesis; eye development; heart development; inner ear development; morphogenesis of an epithelium
8734	0.267	9.00E-38	98.88	*THOC3* (exonic)	THO complex subunit 3	RNA binding	RNA splicing; viral mRNA export from host cell nucleus; RNA export from nucleus; mRNA 3'-end processing; mRNA export from nucleus
8734	0.267	2.00E-39	100	*IL22RA2*	Interleukin-22 receptor subunit alpha-2	interleukin-22 binding; interleukin-22 receptor activity; cytokine receptor activity	negative regulation of inflammatory response; cytokine-mediated signaling pathway; regulation of tyrosine phosphorylation of STAT protein
9040	0.340	2.00E-39	100	*TESK2*	Dual specificity testis-specific protein kinase 2	ATP binding; metal ion binding; protein kinase activity; protein serine/threonine/tyrosine kinase activity; protein serine/threonine kinase activity; protein tyrosine kinase activity	actin cytoskeleton organisation; focal adhesion assembly; protein phosphorylation; spermatogenesis
9115	0.250	7.00E-39	100	*BTC*	Probetacellulin	epidermal growth factor receptor binding; growth factor activity	epidermal growth factor receptor signaling pathway; ERBB2 signaling pathway; MAPK cascade; membrane organization; negative regulation of apoptotic process; negative regulation of epidermal growth factor receptor signaling pathway; positive regulation of cell differentiation; positive regulation of cell division; positive regulation of cell population proliferation; positive regulation of fibroblast proliferation; positive regulation of mitotic nuclear division; positive regulation

							of protein kinase B signaling; positive regulation of urine volume; regulation of cell motility; signal transduction
9474	0.472	2.00E-39	100	*OTUD5*	OTU domain-containing protein 5	Lys48-specific deubiquitinase activity; Lys63-specific deubiquitinase activity; thiol-dependent ubiquitin-specific protease activity; ubiquitinyl hydrolase activity	negative regulation of type I interferon production; protein deubiquitination; protein K48-linked deubiquitination; protein K63-linked deubiquitination; response to lipopolysaccharide
9962	0.430	2.00E-39	100	*ARHGAP12*	Rho GTPase-activating protein 12	GTPase activator activity	actin filament organization; morphogenesis of an epithelial sheet; negative regulation of small GTPase mediated signal transduction; phagocytosis, engulfment; regulation of GTPase activity; regulation of small GTPase mediated signal transduction; signal transduction
10179	0.308	2.00E-39	100	*MARCH3*	E3 ubiquitin-protein ligase MARCH3	zinc ion binding; transferase activity	endocytosis; protein ubiquitination
10240	0.398	2.00E-39	100	*HS6ST2*	Heparan-sulfate 6-O-sulfotransferase 2	heparan sulfate 6-O-sulfotransferase activity	glycosaminoglycan biosynthetic process; heparan sulfate proteoglycan biosynthetic process, enzymatic modification
10251	0.330	2.00E-39	99.67	*FAM186A*	Protein FAM186A	None given	None given
10280	0.328	7.00E-39	100	*LARP4B*	La-related protein 4B	RNA binding	positive regulation of translation
10340	0.481	9.00E-38	100	*NPC1*	NPC intracellular cholesterol transporter 1	cholesterol binding; virus receptor activity; transmembrane signaling receptor activity; signaling receptor activity; sterol transporter activity	adult walking behaviour; autophagy; bile acid metabolic process; cellular response to low-density lipoprotein particle stimulus; cellular response to steroid hormone stimulus; cholesterol efflux; cholesterol homeostasis; cholesterol metabolic process; cholesterol transport; endocytosis; establishment of protein localization to membrane; intracellular cholesterol transport; low-density lipoprotein particle clearance; lysosomal transport; membrane raft organization; negative regulation of cell death; negative regulation of macroautophagy; protein glycosylation; response to cadmium ion; response to drug; viral entry into host cell

10424	0.344	2.00E-39	100	*DUSP23* (exonic)	Dual specificity protein phosphatase 23	phosphatase activity; protein serine/threonine phosphatase activity; protein tyrosine/serine/threonine phosphatase activity; protein tyrosine phosphatase activity	dephosphorylation
10466	0.357	2.00E-39	100	*KCTD19*	BTB/POZ domain-containing protein KCTD19	None given	protein homooligomerisation
10501	0.393	2.00E-39	100	*NKX2-2*	Homeobox protein nkx-2.2	chromatin binding; DNA-binding transcription factor activity; DNA-binding transcription factor activity, RNA polymerase II-specific; proximal promoter sequence-specific DNA binding; RNA polymerase II proximal promoter sequence-specific DNA binding; sequence-specific DNA binding; transcription coactivator activity; transcription factor binding	astrocyte differentiation; brain development; cell differentiation; digestive tract development; endocrine pancreas development; negative regulation of neuron differentiation; neuron fate specification; oligodendrocyte development; optic nerve development; positive regulation of neuron differentiation; response to glucose; response to progesterone; spinal cord motor neuron differentiation; type B pancreatic cell development; ventral spinal cord interneuron fate determination
10752	0.682	9.00E-38	98.88	*TMSB15B*	Thymosin beta-15B	actin monomer binding; DNA-binding transcription factor activity, RNA polymerase II-specific	actin filament organization; positive regulation of cell migration; regulation of cell migration; sequestering of actin monomers
11532	0.422	2.00E-39	100	*PDE1C*	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C	metal ion binding; calcium- and calmodulin-regulated 3',5'-cyclic-GMP phosphodiesterase activity; calmodulin binding; calmodulin-dependent cyclic-nucleotide phosphodiesterase activity	signal transduction
11545	0.343	2.00E-39	100	*UBE2E2*	Ubiquitin-conjugating enzyme E2	ATP binding; ISG15 transferase activity; ubiquitin conjugating enzyme activity; ubiquitin-protein transferase activity	cellular response to DNA damage stimulus; ISG15-protein conjugation; positive regulation of G1/S transition of mitotic cell cycle; protein K11-linked ubiquitination; protein K48-linked ubiquitination; protein K63-linked ubiquitination
11814	0.345	2.00E-39	100	*SCUBE2*	Signal peptide, CUB and EGF-like domain-containing protein 2	calcium ion binding; lipid binding	multicellular organism development

11912	0.390	1.00E-35	98.82	*LOC101317361* (*CDH23*-like)	cadherin-23	calcium ion binding	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules; calcium ion transport; equilibrioception; homophilic cell adhesion via plasma membrane adhesion molecules; inner ear receptor cell stereocilium organization; locomotory behavior; photoreceptor cell maintenance; regulation of cytosolic calcium ion concentration; response to stimulus; sensory perception of light stimulus; sensory perception of sound; visual perception
11915	0.389	4.00E-36	97.75	*LOC101324575* (*SPIDR*-like)	DNA repair-scaffolding protein	None given	cellular response to camptothecin; cellular response to DNA damage stimulus; cellular response to hydroxyurea; cellular response to ionizing radiation; double-strand break repair via homologous recombination; positive regulation of double-strand break repair; positive regulation of protein complex assembly; regulation of double-strand break repair via homologous recombination; regulation of establishment of protein localization to chromosome
11940	0.394	7.00E-39	100	*NRXN3*	neurexin-3-beta	cell adhesion molecule binding; neuroligin family protein binding; transmembrane signaling receptor activity	adult behaviour; angiogenesis; learning; social behaviour; vocalisation behaviour; neuron cell-cell adhesion; signal transduction
12207	0.368	2.00E-39	100	*LOC101318341*	No available information		
12463	0.370	2.00E-39	100	*PRKG1*	cGMP-dependent protein kinase 1	ATP binding; calcium channel regulator activity; cGMP binding; cGMP-dependent protein kinase activity; identical protein binding; protein kinase activity	actin cytoskeleton organization; cGMP-mediated signaling; dendrite development; forebrain development; negative regulation of platelet aggregation; negative regulation of vascular associated smooth muscle cell migration; negative regulation of vascular smooth muscle cell proliferation; neuron migration; protein phosphorylation; regulation of GTPase activity; relaxation of vascular smooth muscle; signal transduction
12507	0.325	2.00E-39	100	*NECAP2*	Adaptin ear-binding coat-associated protein 2	None given	endocytosis; protein transport
12710	0.556	3.00E-37	98.86	*LOC101322849* (*TYW1B*-like)	S-adenosyl-L-methionine-dependent tRNA 4-demethylwyosine synthase	4 iron, 4 sulfur cluster binding; FMN binding; metal ion binding; tRNA-4-demethylwyosine synthase activity	oxidation-reduction process; tRNA processing
12775	0.586	3.00E-37	97.34	*SPATA21*	Spermatogenesis-associated protein 21	calcium ion binding	None given
12844	0.323	3.00E-38	100	*EMID1* (exonic)	EMI domain-containing protein 1	None given	positive regulation of cell-substrate adhesion

13158	0.329	2.00E-39	100	*CARD8* (exonic)	Caspase recruitment domain-containing protein 8	CARD domain binding; cysteine-type endopeptidase activator activity involved in apoptotic process; NACHT domain binding; protein homodimerization activity	activation of cysteine-type endopeptidase activity involved in apoptotic process; inhibition of cysteine-type endopeptidase activity; negative regulation of I-kappaB kinase/NF-kappaB signaling; negative regulation of interleukin-1 beta secretion; negative regulation of lipopolysaccharide-mediated signaling pathway; negative regulation of NF-kappaB transcription factor activity; negative regulation of tumor necrosis factor-mediated signaling pathway; positive regulation of cysteine-type endopeptidase activity involved in apoptotic process; positive regulation of interleukin-1 beta secretion
13282	0.323	2.00E-39	100	*MBNL3*	Muscleblind-like protein 3	metal ion binding; RNA binding	mRNA processing; multicellular organism development; negative regulation of myoblast differentiation; regulation of alternative mRNA splicing, via spliceosome; regulation of RNA splicing; RNA splicing
13484	0.291	3.50E-12	89.3	*DYRK1A*	Dual specificity tyrosine-phosphorylation-regulated kinase 1A	actin binding; ATP binding; cytoskeletal protein binding; identical protein binding; non-membrane spanning protein tyrosine kinase activity; protein kinase activity; protein self-association; protein serine/threonine/tyrosine kinase activity; protein serine/threonine kinase activity; protein tyrosine kinase activity; tau protein binding; tau-protein kinase activity; tubulin binding	amyloid-beta formation; circadian rhythm; negative regulation of DNA damage response, signal transduction by p53 class mediator; negative regulation of microtubule polymerization; negative regulation of mRNA splicing, via spliceosome; nervous system development; positive regulation of protein deacetylation; positive regulation of RNA splicing; protein autophosphorylation; protein phosphorylation; regulation of alternative mRNA splicing, via spliceosome; viral process
13484	0.291	2.00E-39	100	*LOC101326814* (exonic)	No available information		
13625	0.321	2.00E-39	100	*ADARB2*	Double-stranded RNA-specific editase B2	adenosine deaminase activity; double-stranded RNA adenosine deaminase activity; double-stranded RNA binding; metal ion binding; RNA binding; single-stranded RNA binding; tRNA-specific adenosine deaminase activity	mRNA processing; RNA processing; adenosine to inosine editing
13707	0.578	2.00E-39	100	*OXCT1*	Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial	3-oxoacid CoA-transferase activity; protein homodimerization activity	adipose tissue; brain development; cellular ketone body metabolic process; heart development; ketone catabolic process; positive regulation of insulin secretion involved in cellular response to glucose stimulus; response to activity; response to ethanol; response to hormone; response to nutrient; response to starvation

13791	0.394	2.00E-39	100	*IQGAP3*	Ras GTPase-activating-like protein IQGAP3	calmodulin binding; myosin VI light chain binding; Rho GTPase binding	activation of MAPK activity; cellular response to organic substance; ERK1 and ERK2 cascade; G1/S transition of mitotic cell cycle; negative regulation of gene expression; positive regulation of gene expression; positive regulation of mammary gland epithelial cell proliferation; Ras protein signal transduction; regulation of cell size; regulation of GTPase activity
13968	0.419	2.00E-39	100	*SEMA3E*	semaphorin-3E	neuropilin binding; chemorepellent activity; semaphorin receptor binding	branching involved in blood vessel morphogenesis; negative chemotaxis; negative regulation of angiogenesis; negative regulation of axon extension involved in axon guidance; negative regulation of cell-matrix adhesion; neural crest cell migration; positive regulation of cell migration; regulation of actin cytoskeleton reorganization; regulation of cell shape; semaphorin-plexin signaling pathway; sprouting angiogenesis; synapse organization
14153	0.366	7.00E-39	100	*ADCK1*	Uncharacterized aarF domain-containing protein kinase 1	ATP binding; protein serine/threonine kinase activity	None given
14156	0.431	2.00E-39	100	*NSDHL*	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	3-beta-hydroxy-delta5-steroid dehydrogenase activity; 4alpha-carboxy-4beta-methyl-5alpha-cholesta-8-en-3beta-ol:NAD(P)+ 3-oxidoreductase (decarboxylating) activity; 4alpha-carboxy-5alpha-cholesta-8-en-3beta-ol:NAD(P)+ 3-dehydrogenase (decarboxylating) activity; C-3 sterol dehydrogenase (C-4 sterol decarboxylase) activity; oxidoreductase activity; oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; sterol-4-alpha-carboxylate 3-dehydrogenase (decarboxylating) activity	cholesterol biosynthetic process; cholesterol metabolic process; hair follicle development; labyrinthine layer blood vessel development; smoothened signaling pathway
14398	0.483	2.00E-39	100	*DMRTC2* (exonic)	Doublesex- and mab-3-related transcription factor C2	DNA-binding transcription factor activity, RNA polymerase II-specific; metal ion binding; protein homodimerization activity; sequence-specific DNA binding	male meiosis I; positive regulation of histone H3-K9 dimethylation; positive regulation of histone H3-K9 trimethylation; sex differentiation; spermatid nucleus elongation

14587	0.474	2.00E-39	100	*VIPR2*	Vasoactive intestinal polypeptide receptor 2	G protein-coupled peptide receptor activity; G protein-coupled receptor activity; peptide hormone binding; vasoactive intestinal polypeptide receptor activity	activation of adenylate cyclase activity; cell-cell signaling; cell surface receptor signaling pathway; G protein-coupled receptor signaling pathway; negative regulation of smooth muscle cell proliferation; signal transduction
14601	0.374	2.00E-39	100	*JDP2*	Jun dimerization protein 2	chromatin binding; DNA-binding transcription factor activity, RNA polymerase II-specific; DNA-binding transcription repressor activity, RNA polymerase II-specific; protein heterodimerization activity; RNA polymerase II proximal promoter sequence-specific DNA binding; transcription corepressor activity	negative regulation of fat cell differentiation; negative regulation of transcription by RNA polymerase II; positive regulation of histone deacetylation
14643	0.465	2.00E-39	100	*APH1B*	Gamma-secretase subunit APH-1B	endopeptidase activity; peptidase activity	cellular protein metabolic process; ephrin receptor signaling pathway; locomotory behavior; membrane protein intracellular domain proteolysis; Notch receptor processing; Notch receptor processing, ligand-dependent; Notch signaling pathway; positive regulation of apoptotic process; positive regulation of catalytic activity; protein processing
14795	0.648	3.00E-38	100	*AGR2*	Anterior gradient protein 2 homolog	dystroglycan binding; epidermal growth factor receptor binding; protein homodimerization activity	digestive tract morphogenesis; lung goblet cell differentiation; mucus secretion; negative regulation of cell death; positive regulation of cell-substrate adhesion; positive regulation of developmental growth; positive regulation of epidermal growth factor receptor signaling pathway; positive regulation of gene expression; positive regulation of IRE1-mediated unfolded protein response; positive regulation of PERK-mediated unfolded protein response; positive regulation of protein localization to plasma membrane
14911	0.541	2.00E-39	100	*LOC109552183*	No available information		
14973	0.545	1.00E-35	98.82	*PIK3R1*	Phosphatidylinositol 3-kinase regulatory subunit alpha	1-phosphatidylinositol-3-kinase regulator activity; ErbB-3 class receptor binding; insulin binding; insulin-like growth factor receptor binding; insulin receptor binding; insulin receptor substrate binding; neurotrophin TRKA receptor	axon guidance; B cell differentiation; cellular glucose homeostasis; cellular response to insulin stimulus; cellular response to UV; epidermal growth factor receptor signaling pathway; growth hormone receptor signaling pathway; insulin-like growth factor receptor signaling pathway; insulin receptor signaling pathway; intrinsic apoptotic signaling pathway in response to

						binding; phosphatidylinositol 3-kinase binding; phosphatidylinositol 3-kinase regulator activity; phosphatidylinositol 3-kinase regulatory subunit binding; phosphotyrosine residue binding; protein heterodimerization activity; protein phosphatase binding; transcription factor binding; transmembrane receptor protein tyrosine kinase adaptor activity	DNA damage; leukocyte migration; negative regulation of apoptotic process; negative regulation of cell-matrix adhesion; negative regulation of osteoclast differentiation; platelet activation; positive regulation of glucose import; positive regulation of protein import into nucleus; positive regulation of protein localization to plasma membrane; positive regulation of RNA splicing; positive regulation of transcription by RNA polymerase II; positive regulation of tumor necrosis factor production; protein import into nucleus; protein phosphorylation; regulation of insulin receptor signaling pathway; T cell receptor signaling pathway; vascular endothelial growth factor receptor signaling pathway; viral process
15258	0.295	2.00E-39	100	*PARD3*	Partitioning defective 3 homolog	identical protein binding; phosphatidylinositol-3,4,5-trisphosphate binding; phosphatidylinositol-3-phosphate binding; phosphatidylinositol-4,5-bisphosphate binding; phosphatidylinositol binding; protein phosphatase binding	asymmetric cell division; axonogenesis; bicellular tight junction assembly; cell adhesion; cell cycle; establishment of cell polarity; establishment of centrosome localization; establishment of epithelial cell polarity; establishment or maintenance of cell polarity; establishment or maintenance of epithelial cell apical/basal polarity; microtubule cytoskeleton organization; myelination in peripheral nervous system; positive regulation of myelination; protein-containing complex assembly; protein localization; protein targeting to membrane; regulation of cellular localization; transforming growth factor beta receptor signaling pathway
15307	0.389	4.00E-36	97.75	*PLXNA2*	Plexin-A2	identical protein binding; semaphorin receptor activity	centrosome localization cerebellar granule cell precursor tangential migration; limb bud formation; negative regulation of cell adhesion; neural tube development; pharyngeal system development; positive regulation of axonogenesis; regulation of cell migration; regulation of cell shape; regulation of GTPase activity; semaphorin-plexin signaling pathway; semaphorin-plexin signaling pathway involved in axon guidance; somitogenesis
15355	0.340	2.00E-39	100	*LOC109547740* (exonic) (*GTF2IRD1*-like)	general transcription factor II-I repeat domain-containing protein 1	DNA binding; DNA-binding transcription factor activity; DNA-binding transcription factor activity, RNA polymerase II-specific	multicellular organism development; regulation of transcription, DNA-templated; transcription by RNA polymerase II; transition between slow and fast fiber

15418	0.446	2.00E-39	100	*TBX1*	T-box transcription factor TBX1	DNA binding; DNA-binding transcription factor activity, RNA polymerase II-specific; protein dimerization activity; protein homodimerization activity; sequence-specific DNA binding	angiogenesis; anterior/posterior pattern specification; aorta morphogenesis; artery morphogenesis; blood vessel development; blood vessel morphogenesis; cell fate specification; cell population proliferation; cellular response to fibroblast growth factor stimulus; cellular response to retinoic acid; cochlea morphogenesis; coronary artery morphogenesis; determination of left/right symmetry; ear morphogenesis; embryonic cranial skeleton morphogenesis; embryonic viscerocranium morphogenesis; enamel mineralization; epithelial cell differentiation; face morphogenesis; heart development; heart morphogenesis; inner ear morphogenesis; lymph vessel development; mesoderm development; middle ear morphogenesis; muscle organ morphogenesis; muscle tissue morphogenesis; negative regulation of cell differentiation; negative regulation of mesenchymal cell apoptotic process; neural crest cell migration; odontogenesis of dentin-containing tooth; outer ear morphogenesis; outflow tract septum morphogenesis; parathyroid gland development; pattern specification process; pharyngeal system development; positive regulation of cell population proliferation; positive regulation of epithelial cell proliferation; positive regulation of MAPK cascade; positive regulation of mesenchymal cell proliferation; positive regulation of protein phosphorylation; positive regulation of tongue muscle cell differentiation; positive regulation of transcription, DNA-templated; positive regulation of transcription by RNA polymerase II; regulation of animal organ morphogenesis; regulation of transcription by RNA polymerase II; retinoic acid receptor signaling pathway; semicircular canal morphogenesis; sensory perception of sound; social behavior; soft palate development; thymus development; thyroid gland development; tongue morphogenesis; vagus nerve morphogenesis
15492	0.369	3.00E-37	100	*AGL*	Glycogen debranching enzyme	4-alpha-glucanotransferase activity; amylo-alpha-1,6-glucosidase activity; beta-maltose 4-alpha-glucanotransferase activity; glycogen debranching enzyme activity; polysaccharide binding; polyubiquitin modification-dependent protein binding	glycogen biosynthetic process; glycogen catabolic process; neutrophil degranulation; response to glucocorticoid; response to nutrient

15722	0.626	2.00E-39	100	*MAPKAPK3*	MAP kinase-activated protein kinase 3	ATP binding; calcium-dependent protein serine/threonine kinase activity; calmodulin binding; calmodulin-dependent protein kinase activity; MAP kinase kinase activity; mitogen-activated protein kinase binding; protein serine/threonine kinase activity	activation of MAPK activity; intracellular signal transduction; macropinocytosis; peptidyl-serine phosphorylation; protein autophosphorylation; response to cytokine; response to lipopolysaccharide; signal transduction; toll-like receptor signaling pathway; vascular endothelial growth factor receptor signaling pathway
16121	0.277	7.00E-39	100	*MYH11*	Myosin-11	actin filament binding; motor activity; structural constituent of muscle; ATP binding; calmodulin binding	cardiac muscle fiber development; elastic fiber assembly; muscle contraction; skeletal muscle myosin thick filament assembly; smooth muscle contraction
16440	0.415	2.00E-39	100	*EEFSEC*	Selenocysteine-specific elongation factor	GTPase activity; GTP binding; ribonucleoprotein complex binding; selenocysteine insertion sequence binding; translation elongation factor activity; tRNA binding	selenocysteine incorporation; translational elongation
16603	0.325	9.00E-38	98.88	*BAIAP2L1*	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	actin binding; cadherin binding involved in cell-cell adhesion; proline-rich region binding	actin crosslink formation; actin filament bundle assembly; plasma membrane organization; positive regulation of actin cytoskeleton reorganization; positive regulation of actin filament polymerization; regulation of insulin receptor signaling pathway; response to bacterium
16682	0.341	2.00E-39	100	*USP10*	Ubiquitin carboxyl-terminal hydrolase 10	cysteine-type endopeptidase activity; ion channel binding; p53 binding; RNA binding; thiol-dependent ubiquitin-specific protease activity; thiol-dependent ubiquitinyl hydrolase activity	autophagy; cellular response to DNA damage stimulus; cellular response to interleukin-1; DNA damage response, signal transduction by p53 class mediator; negative regulation of I-kappaB kinase/NF-kappaB signaling; protein deubiquitination; regulation of autophagy; translesion synthesis; ubiquitin-dependent protein catabolic process
16751	0.320	1.00E-36	98.85	*SEC14L1* (exonic)	SEC14-like protein 1	molecular function regulator; RIG-I binding	innate immune response; choline transport; negative regulation of RIG-I signaling pathway
16896	0.573	1.00E-36	99.33	*GLC3*	Glypican-3	peptidyl-dipeptidase inhibitor activity	anatomical structure morphogenesis; anterior/posterior axis specification; body morphogenesis; bone mineralization; branching involved in ureteric bud morphogenesis; cell migration; cell migration involved in gastrulation; cell proliferation involved in kidney development; cell proliferation involved in metanephros development; cellular protein metabolic process; coronary vasculature development; embryonic hindlimb morphogenesis; glycosaminoglycan biosynthetic process; glycosaminoglycan catabolic process; lung

							development; mesenchymal cell proliferation involved in ureteric bud development; mesonephric duct morphogenesis; negative regulation of canonical Wnt signaling pathway; negative regulation of epithelial cell proliferation; negative regulation of growth; negative regulation of smoothened signaling pathway; osteoclast differentiation; positive regulation of BMP signaling pathway; positive regulation of Wnt signaling pathway, planar cell polarity pathway; positive regulation of canonical Wnt signaling pathway; positive regulation of endocytosis; positive regulation of glucose import; positive regulation of protein catabolic process; positive regulation of smoothened signaling pathway; post-translational protein modification; regulation of canonical Wnt signaling pathway; regulation of non-canonical Wnt signaling pathway; regulation of protein localization to membrane; response to bacterium; retinoid metabolic process
16924	0.281	2.00E-39	100	*LOC109550070*	No available information		
17230	0.427	2.00E-39	100	*EVL*	Ena/VASP-like protein	actin binding; profilin binding; SH3 domain binding	actin filament organisation; actin nucleation; actin polymerization or depolymerization; animal organ morphogenesis; axon guidance; cell surface receptor signaling pathway; cellular response to interferon-gamma; negative regulation of epithelial cell migration; negative regulation of ruffle assembly; nervous system development; positive regulation of actin filament polymerization; positive regulation of stress fiber assembly; protein homotetramerization
17344	0.260	2.00E-39	100	*LOC109548740*	No available information		
17413	0.498	9.00E-38	100	*LRRC1*	Leucine-rich repeat-containing protein 1	None given	None given
17429	0.411	2.00E-39	100	*FOXRED2*	FAD-dependent oxidoreductase domain-containing protein 2	flavin adenine dinucleotide binding; oxidoreductase activity	ubiquitin-dependent ERAD pathway

17453	0.463	3.00E-37	100	*IFIH1*	Interferon-induced helicase C domain-containing protein 1	ATP binding; DNA binding; double-stranded RNA binding; helicase activity; identical protein binding; ribonucleoprotein complex binding; single-stranded RNA binding; zinc ion binding	cellular response to exogenous dsRNA; cytoplasmic pattern recognition receptor signaling pathway in response to virus; defense response to virus; detection of virus; innate immune response; MDA-5 signaling pathway; negative regulation of type I interferon production; positive regulation of interferon-alpha production; positive regulation of interferon-alpha secretion; positive regulation of interferon-beta production; positive regulation of interferon-beta secretion; positive regulation of interleukin-6 secretion; positive regulation of response to cytokine stimulus; positive regulation of tumor necrosis factor secretion; protein deubiquitination; protein sumoylation; regulation of type III interferon production; response to virus; viral process
17652	0.309	2.00E-39	100	*AP3M2*	AP-3 complex subunit mu-2	None given	anterograde axonal transport; anterograde synaptic vesicle transport; intracellular protein transport; vesicle-mediated transport
17652	0.309	2.00E-39	100	*PLAT*	Tissue-type plasminogen activator	phosphoprotein binding; serine-type endopeptidase activity; signaling receptor binding	blood coagulation; cellular protein modification process; fibrinolysis; negative regulation of proteolysis; plasminogen activation; platelet-derived growth factor receptor signaling pathway; proteolysis; response to hypoxia; smooth muscle cell migration; trans-synaptic signaling by BDNF, modulating synaptic transmission
17665	0.350	7.00E-39	100	*DSE*	Dermatan-sulfate epimerase	chondroitin-glucuronate 5-epimerase activity	chondroitin sulfate biosynthetic process; dermatan sulfate biosynthetic process; heparan sulfate proteoglycan biosynthetic process
17665	0.350	7.00E-39	100	*CALHM6*	Calcium homeostasis modulator protein 6	None given	ion transport
17815	0.550	2.00E-39	100	*DIAPH2*	Protein diaphanous homolog 2	actin binding; Rho GTPase binding; signaling receptor binding	actin filament organization; female gamete generation; multicellular organism development; oogenesis
17872	0.338	2.00E-39	100	*DNAJB6*	DnaJ homolog subfamily B member 6	ATPase activator activity; chaperone binding; DNA binding; heat shock protein binding; identical protein binding; unfolded protein binding	actin cytoskeleton organization; chorio-allantoic fusion; chorion development; extracellular matrix organization; intermediate filament organization; negative regulation of cysteine-type endopeptidase activity involved in apoptotic process; negative regulation of inclusion body assembly; negative regulation of transcription, DNA-templated; protein folding; protein localization to nucleus; regulation of cellular response to heat; regulation of protein localization; syncytiotrophoblast cell differentiation involved in labyrinthine layer development
17964	0.394	7.00E-39	100	*MSI2*	RNA-binding protein Musashi homolog 2	RNA binding; poly(U) RNA binding	stem cell development

17996	0.278	2.00E-39	100	*CACNA1B*	Voltage-dependent N-type calcium channel subunit alpha-1B	high voltage-gated calcium channel activity; amyloid-beta binding; ATP binding; calcium ion binding; high voltage-gated calcium channel activity; protein C-terminus binding	calcium ion import; calcium ion transport; chemical synaptic transmission; locomotory behavior; membrane depolarization; modulation of chemical synaptic transmission; neurotransmitter secretion; regulation of blood pressure; regulation of calcium ion transport; regulation of heart contraction; regulation of ion transmembrane transport; response to amyloid-beta; response to pain

Table B.ii.7 Functional annotation and gene ontology (GO) terms associated with candidate loci identified as outliers between the southwestern Atlantic Ocean *T. t. truncatus* and *T. t. gephyreus* (early-stage evolution candidates)

Candidate Locus	e-value	% Identity	Candidate Gene	Protein	GO Terms	
					Molecular Function	**Biological Process**
1081	3E-37	98.86	KCNH5	Potassium voltage-gated channel subfamily H member 5	calmodulin binding; ion channel binding; protein heterodimerization activity; voltage-gated potassium channel activity	potassium ion transmembrane transport; regulation of G2/M transition of mitotic cell cycle; regulation of ion transmembrane transport; regulation of membrane potential
1558	3E-38	100	LOC101323668 (ZNF345-like)	Zinc finger protein 345	DNA-binding transcription factor activity; DNA-binding transcription factor activity, RNA polymerase II-specific; metal ion binding; sequence-specific DNA binding	negative regulation of transcription, DNA-templated; negative regulation of transcription by RNA polymerase II; positive regulation of transcription, DNA-templated; positive regulation of transcription by RNA polymerase II; regulation of transcription by RNA polymerase III; transcription by RNA polymerase II; transcription by RNA polymerase III
1714	2E-39	100	LOC101273739	No information available		
3235	2E-34	97.67	PRDM16	Histone-lysine N-methyltransferase PRDM16	activating transcription factor binding; DNA-binding transcription factor activity, RNA polymerase II-specific; histone methyltransferase activity (H3-K9 specific); metal ion binding; sequence-specific DNA binding; SMAD binding; transcription coactivator activity	brown fat cell differentiation; heterochromatin organization; negative regulation of granulocyte differentiation; negative regulation of transcription, DNA-templated; negative regulation of transcription by RNA polymerase II; negative regulation of transforming growth factor beta receptor signaling pathway; neurogenesis; positive regulation of brown fat cell differentiation; positive regulation of cold-induced thermogenesis; positive regulation of transcription, DNA-templated; regulation of cellular respiration; roof of mouth development; somatic stem cell population maintenance; tongue development; white fat cell differentiation
4632	4E-36	97.75	RYR2	Ryanodine receptor 2	calcium channel activity; calcium-induced calcium release activity; calcium ion binding; calcium-release channel activity; calmodulin binding; enzyme binding; identical protein binding; ion channel binding; protein kinase A catalytic subunit binding; protein kinase A regulatory subunit binding; protein self-association; ryanodine-sensitive calcium-release channel activity; suramin binding	BMP signaling pathway; calcium ion transport; calcium ion transport into cytosol; calcium-mediated signaling; calcium-mediated signaling using intracellular calcium source; canonical Wnt signaling pathway; cardiac muscle contraction; cardiac muscle hypertrophy; cell communication by electrical coupling involved in cardiac conduction; cellular calcium ion homeostasis; cellular response to caffeine; cellular response to epinephrine stimulus; detection of calcium ion; embryonic heart tube morphogenesis; establishment of protein localization to endoplasmic reticulum; ion transmembrane transport; left ventricular cardiac muscle tissue morphogenesis; positive regulation of ATPase-coupled calcium transmembrane transporter activity; positive regulation of heart rate; positive regulation of sequestering of calcium ion; positive regulation of the force of heart contraction; Purkinje myocyte to ventricular cardiac muscle cell signaling; regulation of atrial cardiac

						muscle cell action potential; regulation of AV node cell action potential; regulation of cardiac conduction; regulation of cardiac muscle contraction; regulation of cardiac muscle contraction by calcium ion signaling; regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion; regulation of cytosolic calcium ion concentration; regulation of heart rate; regulation of SA node cell action potential; regulation of ventricular cardiac muscle cell action potential; release of sequestered calcium ion into cytosol; release of sequestered calcium ion into cytosol by sarcoplasmic reticulum; response to caffeine; response to hypoxia; response to muscle activity; response to muscle stretch; response to redox state; sarcoplasmic reticulum calcium ion transport; type B pancreatic cell apoptotic process; ventricular cardiac muscle cell action potential
15595	6E-30	93.48	PRKAG2	5'-AMP-activated protein kinase subunit gamma-2	ADP binding; AMP-activated protein kinase activity; AMP binding; ATP binding; cAMP-dependent protein kinase inhibitor activity; cAMP-dependent protein kinase regulator activity; phosphorylase kinase regulator activity; protein kinase activator activity; protein kinase binding	ATP biosynthetic process; carnitine shuttle; cell cycle arrest; fatty acid biosynthetic process; glycogen metabolic process; intracellular signal transduction; macroautophagy; negative regulation of protein kinase activity; positive regulation of peptidyl-threonine phosphorylation; positive regulation of protein kinase activity; regulation of fatty acid biosynthetic process; regulation of fatty acid metabolic process; regulation of fatty acid oxidation; regulation of glucose import; regulation of glycolytic process; regulation of macroautophagy; regulation of signal transduction by p53 class mediator; sterol biosynthetic process

B.iii. Methods

B.iii.1 *dDocent* v.2.2.19 script for demultiplexing samples

```
#!/usr/bin/perl -w


my $input = $ARGV[0];
my $output = $ARGV[1];
my $barcode_file = $ARGV[2];
my $radtag = $ARGV[3];
if (not defined ($input && $output && $barcode_file)){print "\nusage: stacks_pipeline.pl raw_sequencing_files_directory Output_folder Barcode_file RAD_tag\n\n"; exit;}

my $format ='';
unless (-d $input){ print "No $input folder exists\n";exit;}
unless (-d $output){`mkdir $output`;}
$outdDocent="$output\/dDocent";
unless (-d $outdDocent){`mkdir $outdDocent`;}
unless (-f $barcode_file){print "$barcode_file file exists\n";exit;}
if (not defined $radtag){$radtag = 'TGCAGG'}; ## default RAD-tag
#my @inputfiles = glob("$input/*");

open(FILE, $barcode_file);
my @DATA = <FILE>;
close FILE;

my $trace=0;
open(POPMAP, ">popmapseq");
foreach my $line (@DATA){
chomp ($line);
my @try = split(/\t/,$line);
print "$#try\n";
if ($#try !=4){print "check format of $barcode_file file\nData should be in this order:\nRaw_sequencing_file\tsample_name\tBarcode\ttype(Parent or Single or Progeney)\tPop\n"; exit;}
my ($raw_file,$sample_name, $barcode, $type,$popmap) = split(/\t/,$line);
$hash_raw{$raw_file}=1;
$hash{$sample_name} = $type;
$hash_pop{$sample_name}=$popmap;
$hash{$raw_file}{traceno}{$trace} = $barcode;
$hash{$raw_file}{code}{$barcode}=$sample_name;
$trace++;
print POPMAP "$sample_name\_1.RAD\t$popmap\n";
}
close BARCODE;
open (RADFILE, '>RAD.txt');
print RADFILE $radtag;
close RADFILE;

`mkdir $output/log`;

      my $options="";
#foreach my $inputfile (@inputfiles){
foreach my $raw (sort keys%hash_raw){
chomp($raw);
                  #my $ori_file = $inputfile;
                  #$ori_file=~ s/$input\///;
  my @inputfile = glob("$input/$raw*");
```

```
  if ($inputfile[0] =~ /R1/){print "single-end files found\n";}else{print
"single-end files not found or check file names (file name should contain
R1)\n";}
open (BARCODE , '>barcode_tmp.txt');
      foreach my $file_number (keys %{$hash{$raw}{'traceno'}}){
                        print                                       BARCODE
$hash{$raw}{traceno}{$file_number},"\n";
                        }
                  close BARCODE;

      if ($inputfile[0] =~ /\.fastq$|\.fq$/i){$format = 'fastq';}
      if ($inputfile[0] =~ /\.gz$/i) {$format = 'gzfastq';}
      `mkdir outtemp radtemp`;
      `process_radtags -f $inputfile[0] -o outtemp -b barcode_tmp.txt -e sbfI
-E phred33 -r --disable_rad_check --barcode_dist_1 2 -i $format`;
      `mv outtemp/process_radtags\.log $output/log/$raw\_barcode.log`;
      `mkdir $output/$raw`;
      `mkdir $output/$raw/remain_reads`;
      `mv outtemp/*.rem.*.fq $output/$raw/remain_reads`;

      my @outfiles = glob("outtemp/*.fq.gz");
#     foreach my $outfile (@outfiles) {
open (BARCODE , 'barcode_tmp.txt');
      while(<BARCODE>){
      chomp($_);
      my $string = $_;
          my @outfiles = glob("outtemp/*_$string.*fq.gz");
            my $outfile1 = $outfiles[0];
#           print "$outfile1\n";
                  #my $string = $outfile;
                              #if ($outfile =~ /sample_(\w+)/){
                                          if                        (defined
$hash{$raw}{code}{$string}){

      $s_name = $hash{$raw}{code}{$string};

      $outfiles[0] =~ s/sample_$string/$s_name/;

#my $new_name= $outfile;

      rename ($outfile1, $outfiles[0]) or die "the rename operation failed
$!:";

#`mkdir rad_sub_temp`;

#`cp $new_name rad_sub_temp`;

`process_radtags -f $outfiles[0] -o radtemp -b RAD.txt -e sbfI -E phred33 -
r --barcode_dist_1 3 --disable_rad_check -i gzfastq`;

`mv radtemp/process_radtags\.log $output/log/$raw\_$s_name\_radtag.log`;

`mv                                         radtemp/sample_$radtag.fq.gz
$outdDocent/pop$hash_pop{$s_name}\_$s_name.F.fq.gz`;

#`rm -r rad_sub_temp`;

                                    }
                                                else{print
$outfile1,"\n";exit;}
```

```
}

`rm -r barcode_tmp.txt`;
}
```

B.iii.2 *dDocent* v.2.2.19 script for processing demultiplexed raw reads

```
#!/usr/bin/env bash
export LC_ALL=en_US.UTF-8

##########dDocent##########
VERSION='2.2.19'
#This script serves as an interactive bash wrapper to QC, assemble, map, and
call SNPs from double digest RAD (SE or PE), ezRAD (SE or PE) data, or SE
RAD data.
#It requires that your raw data are split up by tagged individual and follow
the naming convention of:

#Pop_Sample1.F.fq and Pop_Sample1.R.fq

#Prints out title and contact info
echo -e "dDocent" $VERSION "\n"
echo -e "Contact jpuritz@gmail.com with any problems \n\n "

###Code to check for the required software for dDocent

echo "Checking for required software"
DEP=(freebayes mawk bwa samtools vcftools rainbow gnuplot gawk seqtk cd-hit-
est bamToBed bedtools parallel vcfcombine bamtools pearRM)
NUMDEP=0
for i in "${DEP[@]}"
do
	if which $i &> /dev/null; then
		foo=0
	else
		echo "The dependency" $i "is not installed or is not in your"
'$PATH'"."
		NUMDEP=$((NUMDEP + 1))
	fi
done

if find ${PATH//:/ } -maxdepth 1 -name trimmomatic*jar 2> /dev/null| grep -
q 'trim' ; then
	TRIMMOMATIC=$(find ${PATH//:/ } -maxdepth 1 -name trimmomatic*jar 2>
/dev/null | head -1)
	else
    echo "The dependency trimmomatic is not installed or is not in your"
'$PATH'"."
    NUMDEP=$((NUMDEP + 1))
	fi

#if find ${PATH//:/ } -maxdepth 1 -name TruSeq2-PE.fa 2> /dev/null | grep -
q 'Tru' ; then
```

```
#     ADAPTERS=$/usr/local/lib/Trimmomatic-0.36/adapters
      ADAPTERS=$(find /usr/local/lib/Trimmomatic-0.36/adapters/ -maxdepth 1
-name TruSeq2-PE.fa 2> /dev/null | head -1)
#     echo "$ADAPTERS"
#     else
#     echo "The file listing adapters (included with trimmomatic) is not
installed or is not in your" '$PATH'"."
#    NUMDEP=$((NUMDEP + 1))
#    fi

SAMV1=$(samtools 2>&1 >/dev/null | grep Ver | sed -e 's/Version://' | cut -
f2 -d " " | sed -e 's/-.*//' | cut -c1)
SAMV2=$(samtools 2>&1 >/dev/null | grep Ver | sed -e 's/Version://' | cut -
f2 -d " " | sed -e 's/-.*//' | cut -c3)
      if [ "$SAMV1"  -ge "1" ]; then
            if [ "$SAMV2"  -lt "3" ]; then
            echo "The version of Samtools installed in your" '$PATH' "is not
optimized for dDocent."
            echo "Please install at least version 1.3.0"
                  echo -en "\007"
                  echo -en "\007"
                  exit 1
            fi

      else
                echo "The version of Samtools installed in your" '$PATH' "is
not optimized for dDocent."
            echo "Please install at least version 1.3.0"
                  echo -en "\007"
                  echo -en "\007"
                  exit 1
      fi

RAINV=(`rainbow | head -1 | cut -f2 -d' ' `)
      if [[ "$RAINV" != "2.0.2" && "$RAINV" != "2.0.3" && "$RAINV" != "2.0.4"
]]; then
            echo "The version of Rainbow installed in your" '$PATH' "is not
optimized for dDocent."
            echo -en "\007"
                  echo -en "\007"
                  echo -en "\007"
            echo "Is the version of rainbow installed newer than 2.0.2?
Enter yes or no."
                  read TEST
                  if [ "$TEST" != "yes" ]; then
                  echo "Please install a version newer than 2.0.2"
                  exit 1
            fi
        fi
FREEB=(`freebayes | grep -oh 'v[0-9].*' | cut -f1 -d "." | sed 's/v//' `)

      if [ "$FREEB" != "1" ]; then
            echo "The version of FreeBayes installed in your" '$PATH' "is
not optimized for dDocent."
            echo "Please install at least version 1.0.0"
            exit 1
        fi
VCFTV=$(vcftools | grep VCF | grep -oh '[0-9]*[a-z]*)$' | sed 's/[a-z)]//')
      if [ "$VCFTV" -lt "10" ]; then
            echo "The version of VCFtools installed in your" '$PATH' "is not
optimized for dDocent."
```

```
            echo "Please install at least version 0.1.11"
            exit 1
        elif [ "$VCFTV" == "11" ]; then
                VCFGTFLAG="--geno"
        elif [ "$VCFTV" -ge "12" ]; then
                VCFGTFLAG="--max-missing"
      fi
BWAV=$(bwa 2>&1 | mawk '/Versi/' | sed 's/Version: //g' | sed 's/0.7.//g' |
sed 's/-.*//g' | sed 's/a//g')
      if [ "$BWAV" -lt "13" ]; then
            echo "The version of bwa installed in your" '$PATH' "is not
optimized for dDocent."
            echo "Please install at least version 0.7.13"
            exit 1
      fi

BTC=$( bedtools --version | mawk '{print $2}' | sed 's/v//g' | cut -f1,2 -
d"." | sed 's/2\.//g' | sed 's/a//g' )
      if [ "$BTC" -ge "26" ]; then
            BEDTOOLSFLAG="NEW"
            elif [ "$BTC" == "23" ]; then
            BEDTOOLSFLAG="OLD"
            elif [ "$BTC" != "23" ]; then
            echo "The version of bedtools installed in your" '$PATH' "is not
optimized for dDocent."
            echo "Please install version 2.23.0 or version 2.26.0 and above"
            exit 1
      fi

if ! awk --version | fgrep -v GNU &>/dev/null; then
         awk=gawk
    else
         awk=awk
fi

if [ $NUMDEP -gt 0 ]; then
      echo -e "\nPlease install all required software before running dDocent
again."
      exit 1
else
      echo -e "\nAll required software is installed!"
fi

#This code checks for individual fastq files follow the correct naming
convention and are gziped
TEST=$(ls *.fq 2> /dev/null | wc -l )
if [ "$TEST" -gt 0 ]; then
echo -e "\ndDocent is now configured to work on compressed sequence files.
Please run gzip to compress your files."
echo "This is as simple as 'gzip *.fq'"
echo "Please rerun dDocent after compressing files."
exit 1
fi

#Count number of individuals in current directory
NumInd=$(ls *.F.fq.gz | wc -l)
NumInd=$(($NumInd - 0))

#Create list of sample names
ls *.F.fq.gz > namelist
sed -i'' -e 's/.F.fq.gz//g' namelist
```

```
#Create an array of sample names
NUMNAMES=$(mawk '/_/' namelist | wc -l)

if [ "$NUMNAMES" -eq "$NumInd" ]; then
      NAMES=( `cat "namelist" `)
else
      echo "Individuals do not follow the dDocent naming convention."
      echo "Please rename individuals to: Locality_Individual.F.fq.gz"
      echo "For example: LocA_001.F.fq.gz"
      exit 1
fi

#Wrapper for main program functions.  This allows the entire file to be read
first before execution
main(){
##########User Input Section##########
#This code gets input from the user and assigns variables
######################################

#Sets a start time variable
STARTTIME=$(date)

echo -e "\ndDocent run started" $STARTTIME "\n"

#dDocent can now accept a configuration file instead of running interactively
#Checks if a configuration file is being used, if not asks for user input
if [ -n "$1" ]; then
      CONFIG=$1
      NUMProc=$(grep -A1 Processor $CONFIG | tail -1)
      MAXMemory=$(grep -A1 Memory $CONFIG | tail -1)
      TRIM=$(grep -A1 Trim $CONFIG | tail -1)
      ASSEMBLY=$(grep -A1 '^Assembly' $CONFIG | tail -1)
      ATYPE=$(grep -A1 Type $CONFIG | tail -1)
      simC=$(grep -A1 Simi $CONFIG | tail -1)
      MAP=$(grep -A1 Mapping_R $CONFIG | tail -1)
      optA=$(grep -A1 _Match $CONFIG | tail -1)
      optB=$(grep -A1 MisMatch $CONFIG | tail -1)
      optO=$(grep -A1 Gap $CONFIG | tail -1)
      SNP=$(grep -A1 SNP $CONFIG | tail -1)
      MAIL=$(grep -A1 Email $CONFIG | tail -1)
      if [ "$ASSEMBLY" == "no" ]; then
            #Prints instructions on how to move analysis to background and
disown process
            echo  "At  this  point,  all  configuration  information  has  been
entered and dDocent may take several hours to run."
            echo "It is recommended that you move this script to a background
operation and disable terminal input and output."
            echo "All data and logfiles will still be recorded."
            echo "To do this:"
            echo "Press control and Z simultaneously"
            echo "Type 'bg' without the quotes and press enter"
            echo "Type 'disown -h' again without the quotes and press enter"
            echo ""
            echo "Now sit back, relax, and wait for your analysis to finish."
      else
            echo  "dDocent  will  require  input  during  the  assembly  stage.
Please wait until prompt says it is safe to move program to the background."
      fi

else
      GetInfo
```

```
fi

#Creates (or appends to) a dDcoent run file recording variables
echo "Variables used in dDocent Run at" $STARTTIME >> dDocent.runs
echo "Number of Processors" >> dDocent.runs
echo $NUMProc >> dDocent.runs
echo "Maximum Memory" >> dDocent.runs
echo $MAXMemory >> dDocent.runs
echo "Trimming" >> dDocent.runs
echo $TRIM >> dDocent.runs
echo "Assembly?" >> dDocent.runs
echo $ASSEMBLY >> dDocent.runs
echo "Type_of_Assembly" >> dDocent.runs
echo $ATYPE >> dDocent.runs
echo "Clustering_Similarity%" >> dDocent.runs
echo $simC >> dDocent.runs
echo "Mapping_Reads?" >> dDocent.runs
echo $MAP >> dDocent.runs
echo "Mapping_Match_Value" >> dDocent.runs
echo $optA >> dDocent.runs
echo "Mapping_MisMatch_Value" >> dDocent.runs
echo $optB >> dDocent.runs
echo "Mapping_GapOpen_Penalty" >> dDocent.runs
echo $optO >> dDocent.runs
echo "Calling_SNPs?" >> dDocent.runs
echo $SNP >> dDocent.runs
echo "Email" >> dDocent.runs
echo $MAIL >> dDocent.runs


##Section of logic statements that dictates the order and function of
processing the pipeline

if [[ "$TRIM" == "yes" && "$ASSEMBLY" == "yes" ]]; then
        echo "Trimming reads and simultaneously assembling reference
sequences"
        TrimReads & 2> trim.log
        Assemble
        #setupRainbow 2> rainbow.log
        wait
fi
if [[ "$TRIM" == "yes" && "$ASSEMBLY" != "yes" ]]; then
        echo "Trimming reads"
        TrimReads 2> trim.log
fi
if [[ "$TRIM" != "yes" && "$ASSEMBLY" == "yes" ]]; then
        Assemble
        #setupRainbow 2> rainbow.log
fi

#Checks to see if reads will be mapped.
if [ "$MAP" != "no" ]; then
echo "Using BWA to map reads."
      if [ reference.fasta -nt reference.fasta.fai ]; then
        samtools faidx reference.fasta
        bwa index reference.fasta &> index.log
      fi
#dDocent now checks for trimmed read files before attempting mapping
        if [[ "$MAP" != "no" && ! -f "${NAMES[@]:(-1)}".R1.fq.gz ]]; then
            echo "dDocent cannot locate trimmed reads files"
            echo "Please rerun dDocent with quality trimming"
```

```
            exit 1
        fi
#This next section of code checks to see if the reference was assembled by
dDocent
#and if so, modifies the expected insert length distribution for BWA's metric
for proper pairing
        if head -1 reference.fasta | grep -e 'dDocent' reference.fasta
1>/dev/null; then
            rm lengths.txt &> /dev/null
            for i in "${NAMES[@]}";
                  do
                  if [ -f "$i.R.fq.gz" ]; then
                  zcat $i.R.fq.gz | head -2 | tail -1 >> lengths.txt
                  fi
                  done
            if [ -f "lengths.txt" ]; then
            MaxLen=$(mawk '{ print length() | "sort -rn" }' lengths.txt| head
-1)
            INSERT=$(($MaxLen * 2 ))
            INSERTH=$(($INSERT + 100 ))
            INSERTL=$(($INSERT - 100 ))
            SD=$(($INSERT / 5))
            fi
#BWA for mapping for all samples.  As of version 2.0 can handle SE or PE
reads by checking for PE read files
            for i in "${NAMES[@]}"
            do
            if [ -f "$i.R2.fq.gz" ]; then
                  bwa mem reference.fasta $i.R1.fq.gz $i.R2.fq.gz -L 20,5 -
I $INSERT,$SD,$INSERTH,$INSERTL -t $NUMProc -a -M -T 10 -A $optA -B $optB -
O $optO -R "@RG\tID:$i\tSM:$i\tPL:Illumina" 2> bwa.$i.log | mawk '$6 !~/[2-
9].[SH]/ && $6 !~ /[1-9][0-9].[SH]/' | samtools view -@$NUMProc -q 1 -SbT
reference.fasta - > $i.bam 2>$i.bam.log
            else
                  bwa mem reference.fasta $i.R1.fq.gz -L 20,5 -t $NUMProc -
a -M -T 10 -A $optA -B $optB -O $optO -R "@RG\tID:$i\tSM:$i\tPL:Illumina" 2>
bwa.$i.log | mawk '$6 !~/[2-9].[SH]/ && $6 !~ /[1-9][0-9].[SH]/' | samtools
view -@$NUMProc -q 1 -SbT reference.fasta - > $i.bam 2>$i.bam.log
            fi
            samtools sort -@$NUMProc $i.bam -o $i.bam
            mv $i.bam $i-RG.bam
            samtools index $i-RG.bam
            done
        else
            for i in "${NAMES[@]}"
            do
            if [ -f "$i.R2.fq.gz" ]; then
                  bwa mem reference.fasta $i.R1.fq.gz $i.R2.fq.gz -L 20,5 -
t $NUMProc -a -M -T 10 -A $optA -B $optB -O $optO -R
"@RG\tID:$i\tSM:$i\tPL:Illumina" 2> bwa.$i.log | mawk '$6 !~/[2-9].[SH]/ &&
$6 !~ /[1-9][0-9].[SH]/' | samtools view -@$NUMProc -q 1 -SbT reference.fasta
- > $i.bam 2>$i.bam.log
            else
                  bwa mem reference.fasta $i.R1.fq.gz -L 20,5 -t $NUMProc -
a -M -T 10 -A $optA -B $optB -O $optO -R "@RG\tID:$i\tSM:$i\tPL:Illumina" 2>
bwa.$i.log | mawk '$6 !~/[2-9].[SH]/ && $6 !~ /[1-9][0-9].[SH]/' | samtools
view -@$NUMProc -q 1 -SbT reference.fasta - > $i.bam 2>$i.bam.log
            fi
            samtools sort -@$NUMProc $i.bam -o $i.bam
            mv $i.bam $i-RG.bam
            samtools index $i-RG.bam
```

```
            done
        fi
fi

##Creating mapping intervals if needed, CreateIntervals function is defined
later in script
#If mapping is being performed, intervals are created automatically

if [ "$MAP" != "no" ]; then
echo "Creating alignment intervals"
ls *-RG.bam >bamlist.list
CreateIntervals
fi

##SNP Calling Section of code

if [ "$SNP" != "no" ]; then
      #Create list of BAM files
      ls *-RG.bam >bamlist.list
      #If mapping is not being performed, but intervals do not exist they
are created
      if [[ "$MAP" == "no" && ! -f "cat-RRG.bam" ]]; then
            CreateIntervals
      fi
      #Check for runs from older versions to ensure the recreation of cat-
RRG.bam
      if [[ "$MAP" == "no" && -f "map.bed" ]]; then
            CreateIntervals
      fi
      #Check to make sure interval files have been created
      if [[ "$MAP" == "no" && ! -f "mapped.bed" ]]; then
            bamToBed -i cat-RRG.bam > map.bed
            bedtools merge -i map.bed > mapped.bed
            rm map.bed
      fi
      #This code estimates the coverage of reference intervals and removes
intervals in 0.01% of depth
      #This allows genotyping to be more effecient and eliminates extreme
copy number loci from the data
      if [ "cat-RRG.bam" -nt "cov.stats" ]; then
            if [ "$BEDTOOLSFLAG" == "OLD" ]; then
                  coverageBed -abam cat-RRG.bam -b mapped.bed -counts >
cov.stats
            else
                  bedtools coverage -b  cat-RRG.bam -a mapped.bed -counts -
sorted > cov.stats
      fi
      fi
      if head -1 reference.fasta | grep -e 'dDocent' reference.fasta
1>/dev/null; then

            DP=$(mawk '{print $4}' cov.stats | sort -rn | perl -e
'$d=.001;@l=<>;print $l[int($d*@l)]')
            CC=$( mawk -v x=$DP '$4 < x' cov.stats | mawk '{len=$3-
$2;lc=len*$4;tl=tl+lc} END {OFMT = "%.0f";print tl/"'$NUMProc'"}')
      else
            DP=$(mawk '{print $4}' cov.stats | sort -rn | perl -e
'$d=.00005;@l=<>;print $l[int($d*@l)]')
            CC=$( mawk -v x=$DP '$4 < x' cov.stats | mawk '{len=$3-
$2;lc=len*$4;tl=tl+lc} END {OFMT = "%.0f";print tl/"'$NUMProc'"}')
      fi
```

```
	mawk -v x=$DP '$4 < x' cov.stats |sort -V -k1,1 -k2,2 | mawk -v cutoff=$CC 'BEGIN{i=1}
	{
	len=$3-$2;lc=len*$4;cov = cov + lc
	if ( cov < cutoff) {x="mapped."i".bed";print $1"\t"$2"\t"$3 > x}
	else {i=i+1; x="mapped."i".bed"; print $1"\t"$2"\t"$3 > x; cov=0}
	}'

	FB2=$(( $NUMProc / 4 ))

	echo "Using FreeBayes to call SNPs"

	#Creates a population file to use for more accurate genotype calling
	cut -f1 -d "_" namelist > p
	paste namelist p > popmap
	rm p


###New implementation of SNP calling here to save on memory
	call_genos(){
	samtools view -@$FB2 -b -1 -L mapped.$1.bed -o split.$1.bam cat-RRG.bam
	samtools index split.$1.bam
	freebayes -b split.$1.bam -t mapped.$1.bed -v raw.$1.vcf -f reference.fasta -m 5 -q 5 -E 3 --min-repeat-entropy 1 -V --populations popmap -n 10
	rm split.$1.bam*
	}
	export -f call_genos

	#ls mapped.*.bed | sed 's/mapped.//g' | sed 's/.bed//g' | shuf | parallel --env call_genos --memfree $MAXMemory -j $NUMProc --no-notice call_genos {}
	ls mapped.*.bed | sed 's/mapped.//g' | sed 's/.bed//g' | shuf | parallel --env call_genos -j $NUMProc --no-notice call_genos {}
####
	#ls mapped.*.bed | sed 's/mapped.//g' | sed 's/.bed//g' | shuf | parallel --memfree $MAXMemory -j $FB1 --no-notice --delay 1 freebayes -L bamlist.list -t mapped.{}.bed -v raw.{}.vcf -f reference.fasta -m 5 -q 5 -E 3 --min-repeat-entropy 1 -V --populations popmap -n 10
	#ls mapped.*.bed | sed 's/mapped.//g' | sed 's/.bed//g' | shuf | parallel --memfree $MAXMemory -j $FB1 --no-notice "samtools view -b -L mapped.{}.bed | freebayes -c -t mapped.{}.bed -v raw.{}.vcf -f reference.fasta -m 5 -q 5 -E 3 --min-repeat-entropy 1 -V --populations popmap -n 10"

	  mkdir raw.vcf
	  mv mapped.*.bed ./mapped.bed
	mv raw.1.vcf raw.01.vcf
	mv raw.2.vcf raw.02.vcf
	mv raw.3.vcf raw.03.vcf
	mv raw.4.vcf raw.04.vcf
	mv raw.5.vcf raw.05.vcf
	mv raw.6.vcf raw.06.vcf
	mv raw.7.vcf raw.07.vcf
	mv raw.8.vcf raw.08.vcf
	mv raw.9.vcf raw.09.vcf

	vcfcombine raw.*.vcf | sed -e 's/	\.\:/	\.\/\.\:/g' > TotalRawSNPs.vcf

	if [ ! -d "raw.vcf" ]; then
```

```
            mkdir raw.vcf
      fi
      mv raw.*.vcf ./raw.vcf

      echo "Using VCFtools to parse TotalRawSNPS.vcf for SNPs that are called
in at least 90% of individuals"
      vcftools --vcf TotalRawSNPs.vcf $VCFGTFLAG 0.9 --out Final --recode -
-non-ref-af 0.001 --max-non-ref-af 0.9999 --mac 1 --minQ 30 --recode-INFO-
all &>VCFtools.log
fi

##Checking for possible errors

if [ "$MAP" != "no" ]; then
ERROR1=$(mawk '/developer/' bwa* 2>/dev/null | wc -l 2>/dev/null)
fi
ERROR2=$(mawk '/error/' *.bam.log 2>/dev/null | wc -l 2>/dev/null)
ERRORS=$(($ERROR1 + $ERROR2))

#Move various log files to own directory
if [ ! -d "logfiles" ]; then
mkdir logfiles
fi
mv *.txt *.log ./logfiles 2> /dev/null

#Sending a completion email

if [ $ERRORS -gt 0 ]; then
        echo -e "dDocent has finished with errors in" `pwd` "\n\ndDocent
started" $STARTTIME "\n\ndDocent finished" `date` "\n\nPlease check log
files\n\n" `mawk '/After filtering, kept .* out of a possible/'
./logfiles/VCFtools.log` "\n\ndDocent" $VERSION "\nThe 'd' is silent,
hillbilly." | mailx -s "dDocent has finished with ERRORS!" $MAIL
else
        echo -e "dDocent has finished with an analysis in" `pwd` "\n\ndDocent
started" $STARTTIME "\n\ndDocent finished" `date` "\n\n" `mawk '/After
filtering, kept .* out of a possible/' ./logfiles/VCFtools.log` "\n\ndDocent"
$VERSION "\nThe 'd' is silent, hillbilly." | mailx -s "dDocent has finished"
$MAIL
fi
}

##Function definitions

#Function for trimming reads using trimmomatic
trim_reads(){
      TRIMMOMATIC=$(find ${PATH//:/ } -maxdepth 1 -name trimmomatic*jar 2>
/dev/null | head -1)
    ADAPTERS=$(find ${PATH//:/ } -maxdepth 1 -name TruSeq2-PE.fa 2> /dev/null
| head -1)

      if [ -f $1.R.fq.gz ]; then
            java -jar $TRIMMOMATIC PE -threads 2 -phred33 $1.F.fq.gz
$1.R.fq.gz $1.R1.fq.gz $1.unpairedF.fq.gz $1.R2.fq.gz $1.unpairedR.fq.gz
ILLUMINACLIP:$ADAPTERS:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:5:10 $TW
&> $1.trim.log
      else
            java -jar $TRIMMOMATIC SE -threads 2 -phred33 $1.F.fq.gz
$1.R1.fq.gz ILLUMINACLIP:$ADAPTERS:2:30:10 LEADING:20 TRAILING:20
SLIDINGWINDOW:5:10 $TW &> $1.trim.log
      fi
```

```
}

      export -f trim_reads

TrimReads () {
      #STACKS adds a strange _1 or _2 character to the end of processed
reads, this looks for checks for errant characters and replaces them.
      #This functionality is now parallelized and will run if only SE
sequences are used.

      STACKS=$(cat namelist| parallel -j $NUMProc --no-notice "zcat
{}.F.fq.gz | head -1" | mawk '$0 !~ /\/1$/ && $0 !~ /\/1[ , ]/ && $0 !~ /
1:.*[A-Z]*/' | wc -l )
      FB1=$(( $NUMProc / 2 ))
      if [ $STACKS -gt 0 ]; then

            echo "Removing the _1 character and replacing with /1 in the
name of every sequence"
            cat namelist | parallel -j $FB1 --no-notice "zcat {}.F.fq.gz |
sed -e 's:_1$:/1:g' > {}.F.fq"
            rm -f *.F.fq.gz
            cat namelist | parallel -j $FB1 --no-notice "gzip {}.F.fq"
      fi

      if [ -f "${NAMES[@]:(-1)}".R.fq.gz ]; then

            STACKS=$(cat namelist| parallel -j $NUMProc --no-notice "zcat
{}.R.fq.gz | head -1" | mawk '$0 !~ /\/2$/ && $0 !~ /\/2[ , ]/ && $0 !~ /
2:.*[A-Z]*/'| wc -l )

            if [ $STACKS -gt 0 ]; then
                  echo "Removing the _2 character and replacing with /2 in
the name of every sequence"
                  cat namelist | parallel -j $FB1 --no-notice "zcat
{}.R.fq.gz | sed -e 's:_2$:/2:g' > {}.R.fq"
                  rm -f *.R.fq.gz
                  cat namelist | parallel -j $FB1 --no-notice "gzip {}.R.fq"
            fi
      fi

      cat namelist | parallel -j $NUMProc "zcat {}.F.fq.gz | head -2 | tail
-1 >> lengths.txt"
      MLen=$(mawk '{ print length() | "sort -rn" }' lengths.txt| head -1)
    MLen=$(($MLen / 2))
      TW="MINLEN:$MLen"
      cat namelist | parallel --env trim_reads -j $FB1 trim_reads {}
      mkdir unpaired &>/dev/null
      mv *unpaired*.gz ./unpaired &>/dev/null
}


#Main function for assembly
Assemble()
{
AWK1='BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0; P++}'
AWK2='!/>/'
AWK3='!/NNN/'
PERLT='while (<>) {chomp; $z{$_}++;} while(($k,$v) = each(%z)) {print
"$v\t$k\n";}'
SED1='s/^[  ]*//'
SED2='s/ /  /g'
```

```
FRL=$(zcat ${NAMES[0]}.F.fq.gz | mawk '{ print length() | "sort -rn" }' |
head -1)

if [ ${NAMES[@]:(-1)}.F.fq.gz -nt ${NAMES[@]:(-1)}.uniq.seqs ];then
      if [[ "$ATYPE" == "PE" || "$ATYPE" == "RPE" ]]; then
      #If PE assembly, creates a concatenated file of every unique for each
individual in parallel
            cat namelist | parallel --no-notice -j $NUMProc "zcat {}.F.fq.gz
| mawk '$AWK1' | mawk '$AWK2' > {}.forward"
            cat namelist | parallel --no-notice -j $NUMProc "zcat {}.R.fq.gz
| mawk '$AWK1' | mawk '$AWK2' > {}.reverse"
            if [ "$ATYPE" = "RPE" ]; then
                  cat namelist | parallel --no-notice -j $NUMProc "paste -d
'-' {}.forward {}.reverse | mawk '$AWK3'| sed 's/-/NNNNNNNNNN/' | sort |
uniq -c -w $FRL| sed -e '$SED1' | sed -e '$SED2' > {}.uniq.seqs"
            else
                  cat namelist | parallel --no-notice -j $NUMProc "paste -d
'-' {}.forward {}.reverse | mawk '$AWK3'| sed 's/-/NNNNNNNNNN/' | perl -e
'$PERLT' > {}.uniq.seqs"
            fi
            rm *.forward
            rm *.reverse
      fi
      if [ "$ATYPE" == "SE" ]; then
      #if SE assembly, creates files of every unique read for each individual
in parallel
            cat namelist | parallel --no-notice -j $NUMProc "zcat {}.F.fq.gz
| mawk '$AWK1' | mawk '$AWK2' | perl -e '$PERLT' > {}.uniq.seqs"
      fi
      if [ "$ATYPE" == "OL" ]; then
      #If OL assembly, dDocent assumes that the marjority of PE reads will
overlap, so the software PEAR is used to merge paired reads into single reads
            for i in "${NAMES[@]}";
                  do
                  zcat $i.R.fq.gz | head -2 | tail -1 >> lengths.txt
                  done
            MaxLen=$(mawk '{ print length() | "sort -rn" }' lengths.txt| head
-1)
            LENGTH=$(( $MaxLen / 3))
            for i in "${NAMES[@]}"
                  do
                  pearRM -f $i.F.fq.gz -r $i.R.fq.gz -o $i -j $NUMProc -n
$LENGTH
                  done
            cat namelist | parallel --no-notice -j $NUMProc "mawk '$AWK1'
{}.assembled.fastq | mawk '$AWK2' | perl -e '$PERLT' > {}.uniq.seqs"
      fi


fi

#Create a data file with the number of unique sequences and the number of
occurrences

if [ -f "uniq.seqs.gz" ]; then
      if [ uniq.seqs.gz -nt uniq.seqs ]; then
      gunzip uniq.seqs.gz 2>/dev/null
      fi
fi

if [ ! -f "uniq.seqs" ]; then
```

```
      cat *.uniq.seqs > uniq.seqs
fi


for i in {2..20};
do
echo $i >> pfile
done
cat pfile | parallel -j $NUMProc --no-notice "echo -n {}xxx && mawk -v x={}
'\$1 >= x' uniq.seqs | wc -l" | mawk  '{gsub("xxx","\t",$0); print;}'| sort
-g > uniqseq.data
rm pfile


#Plot graph of above data
gnuplot << \EOF
set terminal dumb size 120, 30
set autoscale
set xrange [2:20]
unset label
set title "Number of Unique Sequences with More than X Coverage (Counted
within individuals)"
set xlabel "Coverage"
set ylabel "Number of Unique Sequences"
plot 'uniqseq.data' with lines notitle
pause -1
EOF


echo -en "\007"
echo -en "\007"
echo -en "\007"
echo -e "Please choose data cutoff.  In essence, you are picking a minimum
(within individual) coverage level for a read (allele) to be used in the
reference assembly"

read CUTOFF
if [ "$ATYPE" == "RPE" ]; then
      parallel --no-notice -j $NUMProc mawk -v x=$CUTOFF \''$1 >= x'\' :::
*.uniq.seqs | cut -f2 | sort | uniq -c -w $FRL | sed -e 's/^[      ]*//'   |
sed -e 's/ /      /g' > uniqCperindv
else
      parallel --no-notice -j $NUMProc mawk -v x=$CUTOFF \''$1 >= x'\' :::
*.uniq.seqs | cut -f2 | perl -e 'while (<>) {chomp; $z{$_}++;} while(($k,$v)
= each(%z)) {print "$v\t$k\n";}' > uniqCperindv
fi
if [ "$NumInd" -gt 10 ]; then
      NUM=$(($NumInd / 2))
else
      NUM=$NumInd
fi

for ((i = 2; i <= $NUM; i++));
do
echo $i >> ufile
done

cat ufile | parallel -j $NUMProc --no-notice "echo -n {}xxx && mawk -v x={}
'\$1 >= x' uniqCperindv | wc -l" | mawk  '{gsub("xxx","\t",$0); print;}'|
sort -g > uniqseq.peri.data
rm ufile
```

```
#Plot graph of above data

gnuplot << \EOF
set terminal dumb size 120, 30
set autoscale
unset label
set title "Number of Unique Sequences present in more than X Individuals"
set xlabel "Number of Individuals"
set ylabel "Number of Unique Sequences"
plot 'uniqseq.peri.data' with lines notitle
pause -1
EOF

echo -en "\007"
echo -en "\007"
echo -en "\007"
echo -e "Please choose data cutoff.  Pick point right before the assymptote.
A good starting cutoff might be 10% of the total number of individuals"

read CUTOFF2

#Prints instructions on how to move analysis to background and disown process
echo "At this point, all configuration information has been entered and
dDocent may take several hours to run."
echo "It is recommended that you move this script to a background operation
and disable terminal input and output."
echo "All data and logfiles will still be recorded."
echo "To do this:"
echo "Press control and Z simultaneously"
echo "Type 'bg' without the quotes and press enter"
echo "Type 'disown -h' again without the quotes and press enter"
echo ""
echo "Now sit back, relax, and wait for your analysis to finish."

#Now that data cutoffs have been chosen, reduce data set to specified set of
unique reads, convert to FASTA format,
#and remove reads with substantial amounts of adapters

mawk -v x=$CUTOFF2 '$1 >= x' uniqCperindv > uniq.k.$CUTOFF.c.$CUTOFF2.seqs
cut -f2 uniq.k.$CUTOFF.c.$CUTOFF2.seqs > totaluniqseq
mawk '{c= c + 1; print ">dDocent_Contig_" c "\n" $1}' totaluniqseq >
uniq.full.fasta
LENGTH=$(mawk '!/>/' uniq.full.fasta | mawk
'(NR==1||length<shortest){shortest=length} END {print shortest}')
LENGTH=$(($LENGTH * 3 / 4))
$awk 'BEGIN {RS = ">" ; FS = "\n"} NR > 1 {print
"@"$1"\n"$2"\n+""\n"gensub(/./, "I", "g", $2)}' uniq.full.fasta > uniq.fq
java -jar $TRIMMOMATIC SE -threads $NUMProc -phred33 uniq.fq uniq.fq1
ILLUMINACLIP:$ADAPTERS:2:30:10 MINLEN:$LENGTH
mawk 'BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0; P++}'
uniq.fq1 > uniq.fasta
mawk '!/>/' uniq.fasta > totaluniqseq
rm uniq.fq*

#If this is a PE assebmle
if [[ "$ATYPE" == "PE" || "$ATYPE" == "RPE" ]]; then
	#Reads are first clustered using only the Forward reads using CD-hit
instead of rainbow
	sed -e 's/NNNNNNNNNN/	/g' uniq.fasta | cut -f1 > uniq.F.fasta
```

```
      CDHIT=$(python -c "print(max("$simC" - 0.1,0.8))")
      cd-hit-est -i uniq.F.fasta -o xxx -c $CDHIT -T 0 -M 0 -g 1 -d 100
&>cdhit.log
      mawk '{if ($1 ~ /Cl/) clus = clus + 1; else  print $3 "\t" clus}'
xxx.clstr | sed 's/[>dDocent_Contig_,...]//g' | sort -g -k1 >
sort.contig.cluster.ids
      paste sort.contig.cluster.ids totaluniqseq >
contig.cluster.totaluniqseq
      sort -k2,2 -g contig.cluster.totaluniqseq | sed -e 's/NNNNNNNNNN/ /g'
> rcluster
      #CD-hit output is converted to rainbow format
      rainbow div -i rcluster -o rbdiv.out -f 0.5 -K 10
      rainbow merge -o rbasm.out -a -i rbdiv.out -r 2 -N10000 -R10000 -l 20
-f 0.75
      #This AWK code replaces rainbow's contig selection perl script
      cat rbasm.out <(echo "E") |sed 's/[0-9]*:[0-9]*://g' | mawk ' {
            if (NR == 1) e=$2;
            else if ($1 ~/E/ && lenp > len1) {c=c+1; print ">dDocent_Contig_"
e "\n" seq2 "NNNNNNNNNN" seq1; seq1=0;
seq2=0;lenp=0;e=$2;fclus=0;len1=0;freqp=0;lenf=0}
            else if ($1 ~/E/ && lenp <= len1) {c=c+1; print
">dDocent_Contig_" e "\n" seq1; seq1=0;
seq2=0;lenp=0;e=$2;fclus=0;len1=0;freqp=0;lenf=0}
            else if ($1 ~/C/) clus=$2;
            else if ($1 ~/L/) len=$2;
            else if ($1 ~/S/) seq=$2;
            else if ($1 ~/N/) freq=$2;
            else if ($1 ~/R/ && $0 ~/0/ && $0 !~/1/ && len > lenf) {seq1 =
seq; fclus=clus;lenf=len}
            else if ($1 ~/R/ && $0 ~/0/ && $0 ~/1/) {seq1 = seq; fclus=clus;
len1=len}
            else if ($1 ~/R/ && $0 ~!/0/ && freq > freqp && len >= lenp ||
$1 ~/R/ && $0 ~!/0/ && freq == freqp && len > lenp) {seq2 = seq; lenp = len;
freqp=freq}
        }' > rainbow.fasta

      seqtk seq -r rainbow.fasta > rainbow.RC.fasta
      mv rainbow.RC.fasta rainbow.fasta

      #The rainbow assembly is checked for overlap between newly assembled
Forward and Reverse reads using the software PEAR
      sed -e 's/NNNNNNNNNN/   /g' rainbow.fasta | cut -f1 | gawk 'BEGIN {RS
= ">" ; FS = "\n"} NR > 1 {print "@"$1"\n"$2"\n+""\n"gensub(/./, "I", "g",
$2)}' > ref.F.fq
      sed -e 's/NNNNNNNNNN/   /g' rainbow.fasta | cut -f2 | gawk 'BEGIN {RS
= ">" ; FS = "\n"} NR > 1 {print "@"$1"\n"$2"\n+""\n"gensub(/./, "I", "g",
$2)}' > ref.R.fq

      seqtk seq -r ref.R.fq > ref.RC.fq
      mv ref.RC.fq ref.R.fq
      LENGTH=$(mawk '!/>/' rainbow.fasta | mawk
'(NR==1||length<shortest){shortest=length} END {print shortest}')
      LENGTH=$(( $LENGTH * 5 / 4))


      pearRM -f ref.F.fq -r ref.R.fq -o overlap -p 0.001 -j $NUMProc -n
$LENGTH

      rm ref.F.fq ref.R.fq
```

```
      mawk 'BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0;
P++}' overlap.assembled.fastq > overlap.fasta
      mawk '/>/' overlap.fasta > overlap.loci.names
      mawk 'BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0;
P++}' overlap.unassembled.forward.fastq > other.F
      mawk 'BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0;
P++}' overlap.unassembled.reverse.fastq > other.R
      paste other.F other.R | mawk '{if ($1 ~ />/) print $1; else print $0}'
| sed 's/   /NNNNNNNNNN/g' > other.FR

      cat other.FR overlap.fasta > totalover.fasta

      rm *.F *.R
fi
if [[ "$ATYPE" != "PE" && "$ATYPE" != "RPE" ]]; then
      cp uniq.fasta totalover.fasta
fi
cd-hit-est -i totalover.fasta -o reference.fasta.original -M 0 -T 0 -c $simC

sed -e 's/^C/NC/g' -e 's/^A/NA/g' -e 's/^G/NG/g' -e 's/^T/NT/g' -e
's/T$/TN/g' -e 's/A$/AN/g' -e 's/C$/CN/g' -e 's/G$/GN/g'
reference.fasta.original > reference.fasta


samtools faidx reference.fasta
bwa index reference.fasta

}

##Create alignment intervals
##This takes advantage of the fact that RAD loci are very discrete.  Instead
of calculating intervals for every BAM file,
##this function merges all BAM files together.  This overall BAM file
##is used to create a single list of intervals, saving a large amount of
computational time.

CreateIntervals()
{
samtools merge -@$NUMProc -b bamlist.list -f cat-RRG.bam &>/dev/null
samtools index cat-RRG.bam
wait
bamToBed -i cat-RRG.bam | bedtools merge -i - > mapped.bed
}



#This checks that dDocent has detected the proper number of individuals and
exits if incorrect
GetInfo(){
echo "$NumInd individuals are detected. Is this correct? Enter yes or no and
press [ENTER]"

read Indcorrect

if [ "$Indcorrect" == "no" ]; then
        echo "Please double check that all fastq files are named
Ind01.F.fq.gz and Ind01.R.fq.gz"
        exit 1
elif [ "$Indcorrect" == "yes" ]; then
            echo "Proceeding with $NumInd individuals"
else
```

```
        echo "Incorrect Input"
        exit 1
fi

#Tries to get number of processors, if not asks user
NUMProc=( `grep -c ^processor /proc/cpuinfo 2> /dev/null` )
NUMProc=$(($NUMProc + 0))

echo "dDocent detects $NUMProc processors available on this system."
echo "Please enter the maximum number of processors to use for this analysis."
        read NUMProc

if [ $NUMProc -lt 1 ]; then
        echo "Incorrect. Please enter the number of processing cores on this
computer"
        read NUMProc
fi
if [ $NUMProc -lt 1 ]; then
        echo "Incorrect input, exiting"
        exit 1
fi

#Tries to get maximum system memory, if not asks user
MAXMemory=$(($(grep -Po '(?<=^MemTotal:)\s*[0-9]+' /proc/meminfo | tr -d "
") / 1048576))G

echo "dDocent detects $MAXMemory maximum memory available on this system."
echo "Please enter the maximum memory to use for this analysis. The size can
be postfixed with
K, M, G, T, P, k, m, g, t, or p which would multiply the size with 1024,
1048576, 1073741824,
1099511627776, 1125899906842624, 1000, 1000000, 1000000000, 1000000000000,
or 1000000000000000 respectively."
echo "For example, to limit dDocent to ten gigabytes, enter 10G or 10g"
        read MAXMemory

while [[ -z $MAXMemory ]];
      do
      echo "Incorrect input"
      echo -e "Please enter the maximum memory to use for this analysis. The
size can be postfixed with K, M, G, T, P, k, m, g, t, or p which would
multiply the size with 1024, 1048576, 1073741824, 1099511627776,
1125899906842624, 1000, 1000000, 1000000000, 1000000000000, or
1000000000000000 respectively."
      echo -e "This option does not work with all distributions of Linux.
If runs are hanging at variant calling, enter 0"
      echo -e "Then press [ENTER]"
      read MAXMemory
      done

#Asks if user wants to trim reads.  This allows this part of the pipeline to
be skipped during subsequent analyses
echo -e "\nDo you want to quality trim your reads?"
echo "Type yes or no and press [ENTER]?"

read TRIM

#Asks if user wants to perform an assembly.  This allows this part of the
pipeline to be skipped during subsequent analyses

echo -e "\nDo you want to perform an assembly?"
```

```
echo "Type yes or no and press [ENTER]."

read ASSEMBLY

if [ "$ASSEMBLY" == "no" ]; then
        echo -e "\nReference contigs need to be in a file named
reference.fasta\n"
        sleep 1
else
      echo -e "What type of assembly would you like to perform?  Enter SE
for single end, PE for paired-end, RPE for paired-end sequencing for RAD
protocols with random shearing, or OL for paired-end sequencing that has
substantial overlap."
      echo -e "Then press [ENTER]"
      read ATYPE

      while [[ $ATYPE != "SE" && $ATYPE != "PE" && $ATYPE != "OL" && $ATYPE
!= "RPE" ]];
      do
      echo "Incorrect input"
      echo -e "What type of assembly would you like to perform?  Enter SE
for single end, PE for paired-end, RPE for paired-end sequencing for RAD
protocols with random shearing, or OL for paired-end sequencing that has
substantial overlap."
      echo -e "Then press [ENTER]"
      read ATYPE
      done
fi
#If performing de novo assembly, asks if the user wants to enter a different
-c value
if [ "$ASSEMBLY" == "yes" ]; then
    echo "Reads will be assembled with Rainbow"
    echo "CD-HIT will cluster reference sequences by similarity. The -c
parameter (% similarity to cluster) may need to be changed for your taxa."
    echo "Would you like to enter a new c parameter now? Type yes or no and
press [ENTER]"
    read optC
    if [ "$optC" == "no" ]; then
            echo "Proceeding with default 0.9 value."
            simC=0.9
        elif [ "$optC" == "yes" ]; then
            echo "Please enter new value for c. Enter in decimal form (For
90%, enter 0.9)"
            read newC
            simC=$newC
        else
            echo "Incorrect input. Proceeding with the default value."
            simC=0.9
        fi
fi

#Asks if user wants to map reads and change default mapping variables for
BWA
echo "Do you want to map reads?  Type yes or no and press [ENTER]"
read MAP
if [ "$MAP" == "no" ]; then
        echo "Mapping will not be performed"
        optA=1
      optB=4
      optO=6
        else
```

```
		echo "BWA will be used to map reads.  You may need to adjust -A -B and -O parameters for your taxa."
		echo "Would you like to enter a new parameters now? Type yes or no and press [ENTER]"
		read optq

	if [ "$optq" == "yes" ]; then
	echo "Please enter new value for A (match score).  It should be an integer.  Default is 1."
	read newA
	optA=$newA
		echo "Please enter new value for B (mismatch score).  It should be an integer.  Default is 4."
	read newB
	optB=$newB
		echo "Please enter new value for O (gap penalty).  It should be an integer.  Default is 6."
	read newO
	optO=$newO
	else
		echo "Proceeding with default values for BWA read mapping."
		optA=1
		optB=4
		optO=6
	fi
fi

#Does user wish to call SNPs?
echo "Do you want to use FreeBayes to call SNPs?  Please type yes or no and press [ENTER]"
read SNP

while [[ $SNP != "yes" && $SNP != "no" ]];
      do
      echo "Incorrect input"
      echo -e "Do you want to use FreeBayes to call SNPs?  Please type yes or no and press [ENTER]"
      read SNP
      done

#Asks user for email address to notify when analysis is complete
echo ""
echo "Please enter your email address.  dDocent will email you when it is finished running."
echo "Don't worry; dDocent has no financial need to sell your email address to spammers."
read MAIL
echo ""
echo ""

if [ "$ASSEMBLY" == "no" ]; then
#Prints instructions on how to move analysis to background and disown process
echo "At this point, all configuration information has been entered and dDocent may take several hours to run."
echo "It is recommended that you move this script to a background operation and disable terminal input and output."
echo "All data and logfiles will still be recorded."
echo "To do this:"
echo "Press control and Z simultaneously"
echo "Type 'bg' without the quotes and press enter"
echo "Type 'disown -h' again without the quotes and press enter"
```

```
echo ""
echo "Now sit back, relax, and wait for your analysis to finish."
fi

if [ "$ASSEMBLY" == "yes" ]; then
echo "dDocent will require input during the assembly stage.  Please wait
until prompt says it is safe to move program to the background."
fi
}

#Actually starts program
if [ -n "$1" ]; then
      main $1 2>&1 | tee -a dDocent_main.LOG #Log all output
else
      main 2>&1 | tee -a dDocent_main.LOG  #Log all output
fi

#Compress Large Leftover files
gzip -f concat.fasta concat.seq rcluster rbdiv.out rbasm.out rainbow.fasta
reference.fasta.original  uniq.seqs  uniq.fasta  totaluniqseq  uniq.F.fasta
uniq.RC.fasta 2> /dev/null &
```

Contact GitHub API Training Shop Blog About
© 2017 GitHub, Inc. Terms Privacy Security Status Help

B.iii.3 Custom BASH scripts for filtering raw ddRAD sequences

```
Remove high missing data:
vcftools --vcf TotalRawSNPs.vcf --missing-indv
#create a document with the names of individuals with >20% missing
data
nano missing.data.indiv
vcftools --vcf TotalRawSNPs.vcf --remove missing.data.indiv --recode
--recode-INFO-all --out FinalInd

Remove all individuals with <500,000 reads that were not removed
above:
#create a document with the names of individuals with <500,000 reads
nano low.reads.indiv
vcftools --vcf FinalInd.recode.vcf --remove ../low.reads.indiv --
recode --recode-INFO-all --out FinalInd2


Remove loci with general missing data >20% and a minor allele frequency
<0.02:
vcftools --vcf FinalInd2.recode.vcf --max-missing 0.8 --min-alleles 2
--max-alleles 2 --out FinalInd80_maf0.02 --non-ref-af 0.001 --max-
non-ref-af 0.9999 --mac 1 --maf 0.02 --minQ 30 --recode --recode-INFO-
all &>VCFtools80_maf0.02.log
#check how many loci left
mawk '!/#/' FinalInd80_maf0.02.recode.vcf | wc -l
```

Remove complex haplotypes and INDELs:

```
vcfallelicprimitives FinalInd80_maf0.02.recode.vcf --keep-info --keep-geno > FinalInd80_1.vcf
mawk '!/#/' FinalInd80_1.vcf | wc –l
vcftools --vcf FinalInd80_1.vcf --remove-indels --recode --recode-INFO-all --out FinalInd80_2
```

Calculating mean coverage per locus:

```
#first calculate mean coverage per locus
vcftools --vcf FinalInd80_2.recode.vcf --site-depth --out FinalInd80_2
cut -f3 FinalInd80_2.ldepth > AverageDepthSite2
#x= sample size
mawk '!/D/' AverageDepthSite2| mawk -v x=385 '{print $1/x}' > meandepthpersite2
#calculate mean coverage plus 2 SD to use as a threshold
mawk '{ sum += $1; sumsq += ($1)^2; n++ } END { if (n > 0) print sum / n, sqrt((sumsq-sum^2 / n) / n); }' meandepthpersite2
#x = average + 2 SD
vcftools --vcf FinalInd80_2.recode.vcf --recode-INFO-all --out FinalInd80_3 --max-meanDP 62 --recode
```

Remove loci with allelic balance <0.20 or <0.80:

```
vcffilter -s -f "AB > 0.20 & AB < 0.80 | AB < 0.01" FinalInd80_3.recode.vcf > FinalInd80_AB.recode.vcf
mawk '!/#/' FinalInd80_AB.recode.vcf | wc –l
```

Remove loci with mapping quality score <0.9 or <1.05:

```
vcffilter -f "MQM / MQMR > 0.9 & MQM / MQMR <1.05" FinalInd80_AB.recode.vcf > FinalInd80_AB2.recode.vcf
mawk '!/#/' FinalInd80_AB2.recode.vcf | wc –l
```

Removing any locus that has a quality score below 0.2 (1/5) of the depth:

```
vcffilter -f "QUAL / DP > 0.20" FinalInd80_AB2.recode.vcf > FinalInd80_AB3.recode.vcf
mawk '!/#/' FinalInd80_AB3.recode.vcf | wc –l
```

Removing SNPs with >10% missing data:

```
#(sample size x 0.1=max missing count)
vcftools --vcf FinalInd80_AB3_RR.recode.vcf --max-missing-count 36 --recode --recode-INFO-all --out FinalInd80_AB4
```

Removing loci outside of Hardy-Weinberg Equilibrium:

```
#using a HWE (-h) value of 0.05 and a –c value (being the percentage of populations out of HWE) of the default of 25%
perl ~/Scripts/HWfilteringVCF.pl -v FinalInd80_AB4.recode.vcf -p ./popmap1.txt -h 0.05 -c 0.25 -o FinalInd80_HWE
#using HWfilteringVCF.pl script (custom script)
mawk '!/#/' FinalInd80_HWE.recode.vcf | wc –l
```

Extracting one locus per site:

```
perl ~/Scripts/ExtractBQFRLM_Elly.pl FinalInd80_HWE.recode.vcf BQ.vcf Frst.vcf Rdm.vcf LM.vcf
# using ExtractBQFRLM.pl script (custom script)
```

Removing loci in linkage disequilibrium:

```
vcftools --vcf BQ.vcf --interchrom-geno-r2 --min-r2 0.8
#using the out.interchrom.geno.ld output file extract the list of loci out of LD in R (LD_exclude_list.txt)
vcftools --vcf BQ.vcf --exclude-positions LD_exclude_list.txt --recode --recode-INFO-all --out BQ_LD
```

Turn final SNPs vcf file into fasta file:

```
perl ~/Scripts/ExtSeq.pl ./reference.fasta ./BQ_LD.recode.vcf ./BQ_LD.fasta
```

Aligning to the *T. aduncus* genome:

```
bowtie2 -x ~/Aduncus_seascape/GenomeAssembly/Tadunc_genome -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 -f ./BQ_LD.fasta -S ./SpComp_to_adunc_bowtie2_assembly --no-sq
#list of SNPs that did not align
nano Nonaligning_SNPs.txt
#removing SNPs that did not align
awk 'FNR == NR { h[$1]; next }; !($1 in h)' Nonaligning_SNPs ./BQ_LD.recode.vcf > Mapped_SNPs_speciescomp.vcf
mawk '!/#/' Mapped_SNPs_speciescomp.vcf | wc –l
```

C. Chapter 3 Appendix

C.i. Figures

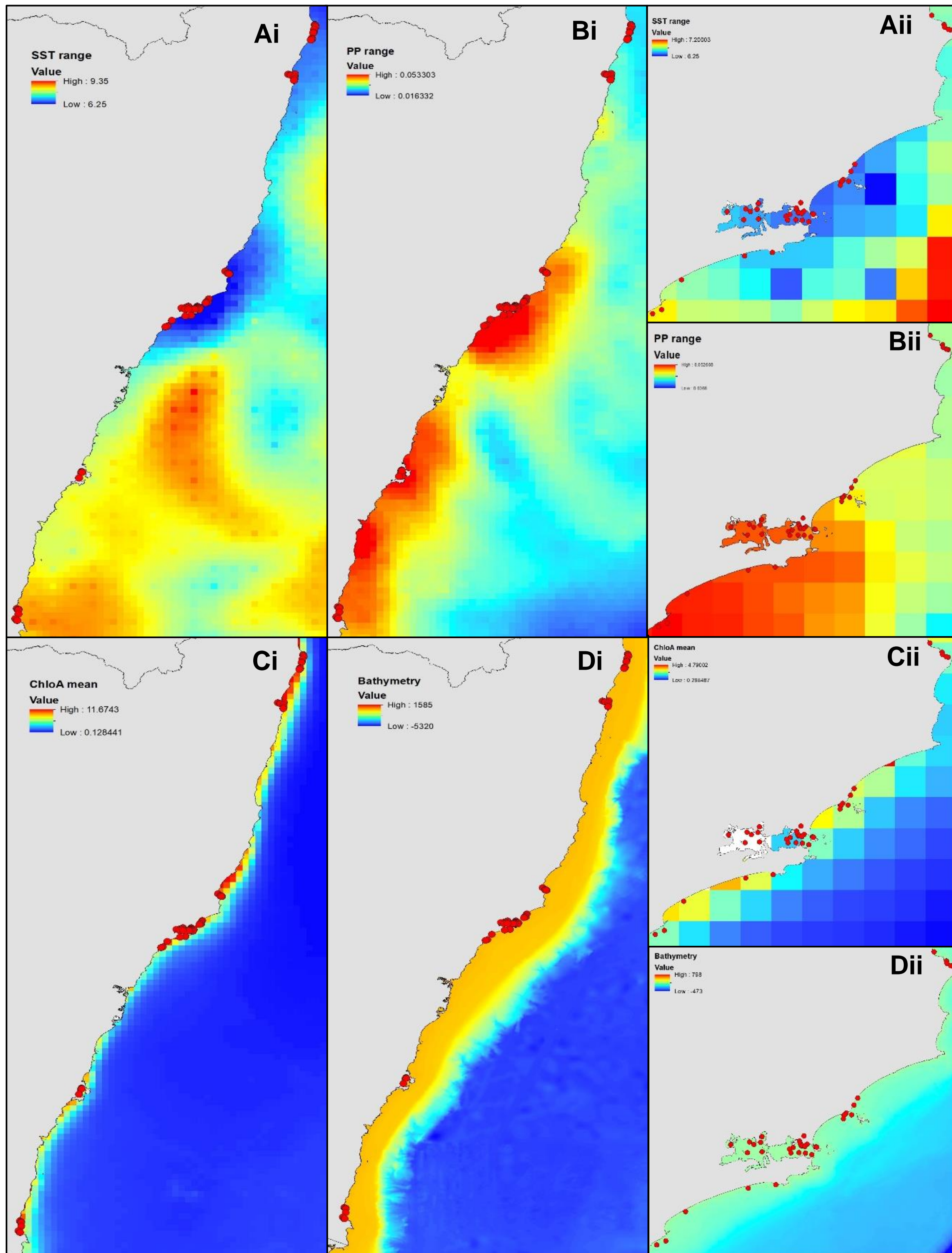


Figure C.i.1 Maps of the four environmental and topological variables included in the final redundancy analysis (RDA) run for the genotype-environment association (GEA) analysis. **A)** sea surface temperature (SST) range, **B)** primary productivity (PP) range, **C)** mean chlorophyll A concentration (chloA) (note: white cells represent no data and were therefore, extrapolated from the geographically closest available data) and **D)** bathymetry gradients **i)** across the entire study region and **ii)** the Hunter Region specifically. Additional information about these ecological variables can be found in Appendix C: Table C.ii.2.

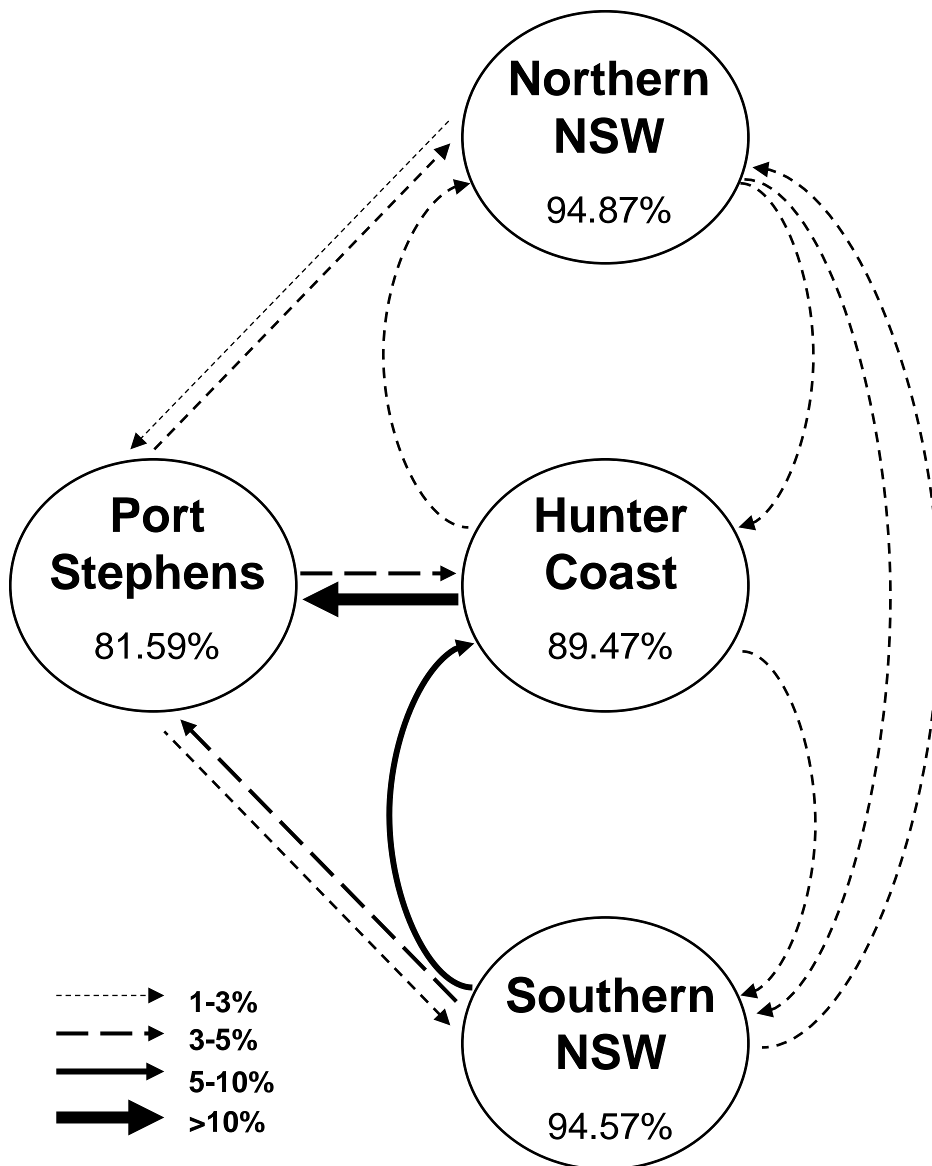


Figure C.i.2 Contemporary migration rates among four populations of Indo-Pacific bottlenose dolphins along the east coast of Australia as estimated by *BayesAss3*. Percentages within circles are the proportion of non-migrants in the population, while arrows are weighted to represent the percentage of migrants among populations.

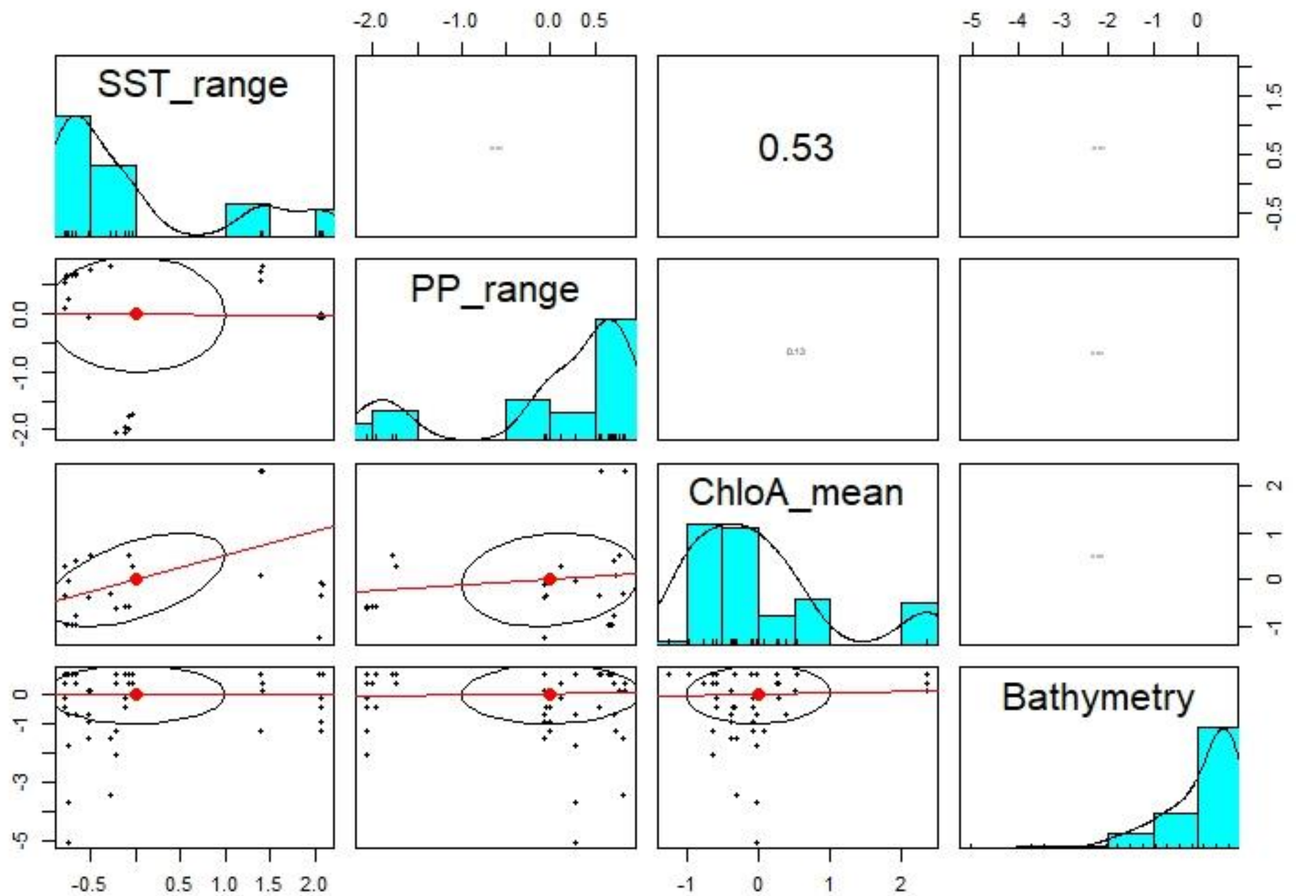


Figure C.i.3 Collinearity analysis of the four ecological variables retained after forward selection for inclusion in the redundancy analysis (RDA). All variables were under the variance inflation factor (VIF) threshold of three, implying low levels of collinearity. Details of the environmental variables used are provided in Appendix C: Table C.ii.2 (SST = sea surface temperature, PP = primary productivity, ChloA = Chlorophyll A concentration).

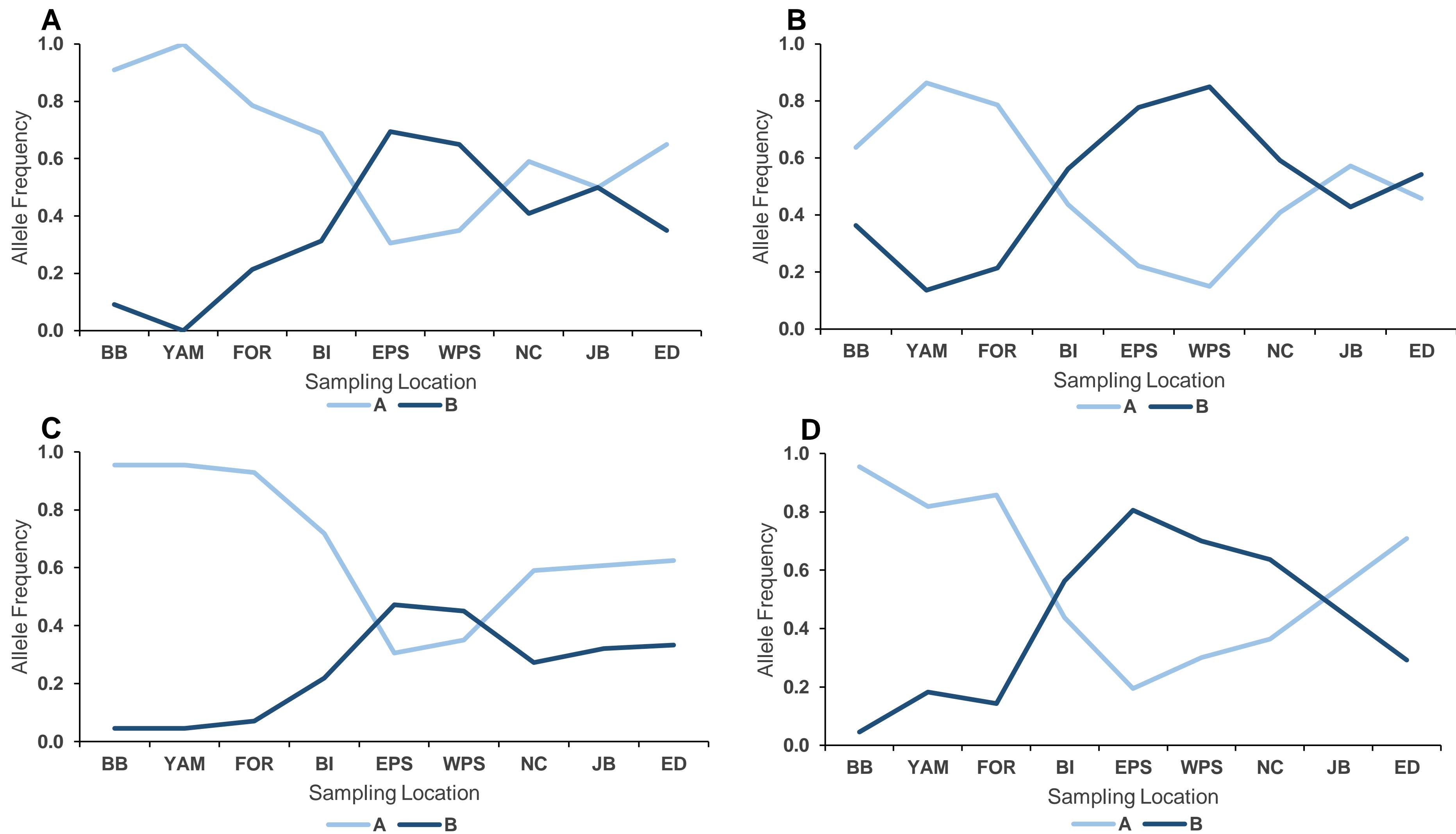


Figure C.i.4 Allele frequency change in four candidate genes: **A)** *THNSL2,* **B)** *OSTF1,* **C)** *SMYD1* and **D)** *VEGFA,* putatively associated with adaptation of several vital bodily systems in Indo-Pacific bottlenose dolphins in eastern Australia. Sampling locations are ordered from north (left) to south (right) and abbreviations are explained in Figure 3.1 (and Appendix C: Table C.ii.3).

C.ii. Tables

Table C.ii.1 Parameters and results for quality filtering double-digest restriction site-associated DNA sequenced (ddRADseq) loci of Indo-Pacific bottlenose dolphins.

	Average no. of reads per individual	2,018,571
	No. of SNPs before filtering	66,884
	Sample size before filtering	111
Filtering Parameter	<20% missing data per individual (final sample size)	110 individuals
	Variants called in >20% of individuals, minor allele frequency (MAF) >0.03, minimum quality score (MinQ) >30 & biallelic SNPs only	28,402 SNPs
	Complex variance and indels	28,368 SNPs
	Av. coverage < av. depth + 2 standard deviations	27,432 SNPs
	Allele balance between 0.20 - 0.80 (proportion of alternate to reference alleles)	26,900 SNPs
	Mapping quality score between 0.9 - 1.05	24,777 SNPs
	Read quality score >20% of read depth	24,654 SNPs
	<4% missing data per variant	21,445 SNPs
	<25% of populations out of HWE	20,738 SNPs
	Best quality SNP per locus	15,395 SNPs
	Linkage Disequilibrium – r^2 <0.8 (final number of SNPs)	14,466 SNPs
	Neutral Dataset (RDA candidates + additional outliers removed)	14,143 SNPs
	Adaptive Dataset (RDA candidates)	281 SNPs

Table C.ii.2 Environmental, topographic and oceanographic layers selected for use in the genotype-environment association (GEA) analysis.

Variable	Description	Spatial Resolution	Source	Temporal Range
Chlorophyll A (ChloA)	Chlorophyll A concentration at mean depth (mean annual maximum, mean, minimum and standard deviation). Units: mg/mŸ	5 arcmin (9.2 km)	Satellite (Aqua-MODIS) Ocean Color (Feldman and McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/ Accessed through Marine Geospatial Ecology Tools (MGET) in ArcGIS	2002-2009
pH	Measure of acidity in the ocean (mean annual mean)	In-situ measurement	World Ocean Database 2009 (Boyer et al. 2013) URL: http://www.nodc.noaa.gov/ Accessed through Bio-ORACLE (URL: http://www.bio-oracle.org/downloads-to-email.php) (Tyberghein et al. 2012; Assis et al. 2018)	1910-2007
Salinity	Model - Dissolved salt concentration at mean depth (mean annual maximum, mean, minimum and range). Units: PSS	0.25 arcdegree	Global Ocean Physics Reanalysis ECMWF ORAP5.0 URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE	2000-2014
Sea Surface Temperature (SST)	Temperature in the topmost one metre of water column (mean annual maximum, mean, minimum and range) Units: ^{O}C	5 arcmin (9.2 km)	Satellite (Aqua-MODIS) Ocean Color (Feldman and McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/ Accessed through Bio-ORACLE	2002-2009
Current Velocity (CV)	Model - Sea water velocity at mean depth (mean annual maximum, mean, minimum and range) Units: m/s	0.25 arcdegree	Global Ocean Physics Reanalysis ECMWF ORAP5.0 URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE	2000-2014
Primary Production (PP)	Model - Net primary productivity of carbon at mean depth (mean annual maximum, mean, minimum and range) Units: g/mŸ/day	0.25 arcdegree	Global Ocean Biogeochemistry Non assimilative hindcast (Pisces) URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE	2000-2014

Bathymetry	Average depth of seafloor Units: Metres (m)	9 arc second (0.0025Ao or ~250 m at equator)	Geoscience Australia Australian Bathymetry and Topography Grid, June 2009 URL: https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search?node=srv#/metadata/a05f7892-fae9-7506-e044-00144fdd4fa6 Accessed through data.gov.au portal	2009
Seafloor Rugosity	Surface area of 3x3 cells of the Aust Bathymetry and Topography Grid above	9 arc second (0.0025A° or ~250 m at equator)	Geoscience Australia Bathymetry derived topographic rugosity grid – created from the Australian Bathymetry and Topography Grid, June 2009 URL: https://data.gov.au/dataset/bathymetry-derived-topographic-rugosity-grid Accessed through data.gov.au portal	2017
Seafloor Slope	Steepness of the seafloor (degree of incline)	1 km^2	Geoscience Australia Australian bathymetry and its derivatives URL: https://data.gov.au/dataset/australian-bathymetry-and-its-derivatives Accessed through data.gov.au portal	2017
Topographic Relief	Amount of topographic change on the seafloor	1 km^2	Geoscience Australia Australian bathymetry and its derivatives URL: https://data.gov.au/dataset/australian-bathymetry-and-its-derivatives Accessed through data.gov.au portal	2017
Distance to River Mouth (DRM)	Distance to the mouth of closest major river. Major river defined as those with a relevance of 1 or 2 (according to river size) in the GIS layer. Units: m	1 m	NSW Government Spatial Services URL: https://datasets.seed.nsw.gov.au/dataset/nsw-hydrography Distance calculated in ArcGIS	2016

Table C.ii.3 Molecular diversity indices for Indo-Pacific bottlenose dolphins (*T. aduncus*) at nine sampling locations along the east coast of Australia, based on 110 individuals and 14,143 SNPs for the neutral dataset and 281 SNPs for the adaptive dataset. N = sample size before filtering (M:F:U represents the number of samples that are male, female or of unknown sex), H_O = observed heterozygosity, H_E = expected heterozygosity, %PL = percentage of polymorphic loci and F_{IS} = Wright's inbreeding coefficient. Standard deviations are shown in parentheses.

Site	Year Sampled	N (M:F:U)	Adaptive Dataset			Neutral Dataset			
			H_O	H_E	% PL	H_O	H_E	% PL	F_{IS}
Byron Bay (BB)	2007	11 (3:8)	0.379 (0.170)	0.393 (0.129)	50.00	0.322 (0.180)	0.338 (0.148)	37.95	0.031
Yamba (YAM)	2007	12 (7:4:1)	0.380 (0.184)	0.385 (0.138)	49.68	0.321 (0.178)	0.333 (0.148)	36.52	0.037
Forster (FOR)	2001 & 2004	7 (2:5)	0.377 (0.189)	0.393 (0.138)	46.40	0.338 (0.193)	0.352 (0.144)	42.35	0.041
Broughton Island (BI)	1999-2004	16 (10:6)	0.397 (0.149)	0.408 (0.116)	50.83	0.312 (0.177)	0.312 (0.153)	42.35	-0.001
Eastern Port Stephens (EPS)	1999-2004	18 (10:8)	0.432 (0.159)	0.420 (0.112)	51.90	0.314 (0.182)	0.306 (0.157)	41.65	-0.027
Western Port Stephens (WPS)	1999-2000	10 (4:3:3)	0.477 (0.199)	0.419 (0.119)	51.27	0.376 (0.209)	0.339 (0.151)	37.70	-0.112
Newcastle (NC)	2001-2005	11 (5:6)	0.397 (0.165)	0.429 (0.113)	51.88	0.326 (0.183)	0.329 (0.151)	38.08	0.008
Jervis Bay (JB)	1998-1999	14 (8:4:2)	0.410 (0.163)	0.404 (0.118)	50.16	0.324 (0.181)	0.318 (0.152)	40.50	-0.019
Eden (ED)	2006-2007	12 (6:6)	0.403 (0.178)	0.405 (0.121)	49.82	0.309 (0.176)	0.323 (0.151)	39.89	0.044
		Average	0.410 (0.034)	0.4132 (0.019)	50.73 (1.81)	0.327 (0.020)	0.327 (0.014)	39.67 (2.19)	0.0002 (0.049)

Table C.ii.4 Analysis of molecular variance (AMOVA) based on 14,143 neutral SNPs, testing the amount of genomic variation explained by the division into four neutral populations: 1) northern New South Wales (NSW), 2) Hunter Coast, 3) Port Stephens and 4) southern NSW, compared to among individual locations across the study region. An asterisk (*) denotes significance at 0.001.

Source of Variation	Degrees of Freedom	Sum of Squares	Variance Components	Percentage of Variation
Among populations	3	10 667.36	43.01	4.49*
Among sampling locations within populations	5	5 900.25	11.92	1.24*
Within pops	211	190 406.46	902.40	94.26*
Total	**219**	**206 974.07**	**957.32**	

Table C.ii.5 Analysis of variance (ANOVA) for testing of genotype-environment association (GEA) based on 14,466 loci in Indo-Pacific bottlenose dolphins. Redundancy analysis (RDA) was used to test the amount of genomic variation explained by the overall model and each of the ecological variables.

	Variable	Degrees of Freedom	Variance Explained	F	*p*
Overall	Model	4	215.5417	1.4317	0.001
	Residual	102	3839.0660		
Marginal	Sea surface temperate (SST) range	1	41.8879	1.1129	0.056
	Primary productivity range	1	64.2631	1.7074	0.001
	Chlorophyll A concentration mean	1	50.5355	1.3427	0.001
	Bathymetry	1	41.2630	1.0963	0.104
	Residual	102	3839.0660		

Table C.ii.6 Significantly over-enriched (*p* <0.05) gene ontology (GO) terms in the putatively adaptive dataset as compared to the full (reference) dataset (14,466 SNPs) using a functional enrichment analysis (Fisher's exact test in Blast2GO).

GO Category	GO ID	GO Term	*p*	Associated genes in: Candidate Dataset (%)	Associated genes in: Reference Dataset (%)
Biological Process	GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.0125	0.7168	0.0495
	GO:0006750	glutathione biosynthetic process	0.0456	1.4440	0.4620
	GO:0006749	glutathione metabolic process	0.0456	1.4440	0.4620
	GO:0042398	cellular modified amino acid biosynthetic process	0.0456	1.4440	0.4620
	GO:0001578	microtubule bundle formation	0.0386	0.3571	0.0071
	GO:0019883	antigen processing and presentation of endogenous antigen	0.0456	1.4440	0.4620
	GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	0.0456	1.4440	0.4620
	GO:0006575	cellular modified amino acid metabolic process	0.0497	1.4440	0.4763
	GO:0030509	BMP signaling pathway	0.0125	0.7168	0.0495
	GO:0030513	positive regulation of BMP signaling pathway	0.0022	0.7168	0.0142
	GO:0030510	regulation of BMP signaling pathway	0.0075	0.7168	0.0354
	GO:0090100	positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.0054	0.7168	0.0283
	GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	0.0476	1.4440	0.4691
	GO:0090287	regulation of cellular response to growth factor stimulus	0.0185	0.7168	0.0637
	GO:0002244	hematopoietic progenitor cell differentiation	0.0386	0.3571	0.0071
	GO:0009719	response to endogenous stimulus	0.0280	2.1818	0.7987
	GO:0007019	microtubule depolymerization	0.0386	0.3571	0.0071
	GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	0.0336	0.7168	0.0921
	GO:0071772	response to BMP	0.0154	0.7168	0.0566
	GO:0071773	cellular response to BMP stimulus	0.0154	0.7168	0.0566
	GO:0019184	nonribosomal peptide biosynthetic process	0.0456	1.4440	0.4620
	GO:0071495	cellular response to endogenous stimulus	0.0241	2.1818	0.7699
	GO:0031109	microtubule polymerization or depolymerization	0.0386	0.3571	0.0071

	GO:0002483	antigen processing and presentation of endogenous peptide antigen	0.0456	1.4440	0.4620
	GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	0.0476	1.4440	0.4691
	GO:0044272	sulfur compound biosynthetic process	0.0497	1.4440	0.4763
Cellular Component	GO:0035102	PRC1 complex	0.0497	1.4440	0.4763
	GO:1990752	microtubule end	0.0195	0.3571	0.0000
	GO:0042824	MHC class I peptide loading complex	0.0456	1.4440	0.4620
	GO:0042825	TAP complex	0.0456	1.4440	0.4620
	GO:0035371	microtubule plus-end	0.0195	0.3571	0.0000
	GO:0016272	prefoldin complex	0.0476	1.4440	0.4691
	GO:0033180	proton-transporting V-type ATPase, V1 domain	0.0386	0.3571	0.0071
	GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	0.0386	0.3571	0.0071
	GO:0000152	nuclear ubiquitin ligase complex	0.0497	1.4440	0.4763
Molecular Function	GO:0042287	MHC protein binding	0.0497	1.4440	0.4763
	GO:0030506	ankyrin binding	0.0386	0.3571	0.0071
	GO:0004357	glutamate-cysteine ligase activity	0.0456	1.4440	0.4620
	GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.0456	1.4440	0.4620
	GO:0016881	acid-amino acid ligase activity	0.0456	1.4440	0.4620
	GO:0008324	cation transmembrane transporter activity	0.0353	1.8116	0.6265
	GO:0008378	galactosyltransferase activity	0.0476	1.4440	0.4691
	GO:0015433	ATPase-coupled peptide antigen transmembrane transporter activity	0.0456	1.4440	0.4620
	GO:0015440	ATPase-coupled peptide transmembrane transporter activity	0.0497	1.4440	0.4763
	GO:0033220	ATPase-coupled amide-transporter activity	0.0497	1.4440	0.4763
	GO:0022890	inorganic cation transmembrane transporter activity	0.0353	1.8116	0.6265
	GO:0039706	co-receptor binding	0.0195	0.3571	0.0000
	GO:0046979	TAP2 binding	0.0456	1.4440	0.4620
	GO:0046977	TAP binding	0.0456	1.4440	0.4620
	GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.0386	0.3571	0.0071

Table C.ii.7 Functional annotation and gene ontology (GO) terms associated with candidate loci identified by genotype-environment association (GEA) analysis to be moderately to highly correlated with mean chlorophyll A concentration (chloA) or primary productivity (PP) range.

Candidate Locus	Correlated Variable	Correlation Coefficient	e-value	% Identity	Candidate Gene	Protein	GO Terms	
							Molecular Function	**Biological Process**
4340	ChloA mean	-0.444	9E-38	98.88	*GRK3*	Beta-adrenergic receptor kinase 2	ATP binding; protein kinase activity; beta-adrenergic receptor kinase activity; G protein-coupled receptor kinase activity	signal transduction; receptor internalisation; G protein-coupled receptor signaling pathway
11682	PP range	-0.508	7E-39	100.00	*VEGFA*	Vascular endothelial growth factor A	chemoattractant activity; growth factor activity; vascular endothelial growth factor receptor 1 (and 2) binding; platelet-derived growth factor receptor binding; heparin binding; neuropilin binding; cytokine activity	angiogenesis; artery morphogenesis; branching involved in blood vessel morphogenesis; camera-type eye morphogenesis; cardiac muscle fibre development; cardiac vascular smooth muscle cell development; cellular response to hypoxia; cell migration involved in sprouting angiogenesis; coronary artery (and vein) morphogenesis; heart morphogenesis; kidney development; lactation; lung development; epithelial cell differentiation; in utero embryonic development; lymph vessel morphogenesis; eye photoreceptor cell development; nervous system development; vasculogenesis (among other terms)
4139	PP range	-0.501	2E-35	97.75	*THNSL2*	Threonine synthase-like 2	cytokine activity; serine binding; pyridoxal phosphate binding; threonine synthase activity	dephosphorylation; serine family amino acid catabolic process; threonine biosynthetic process; 2-oxobutyrate biosynthetic process
8701	PP range	-0.485	2E-35	98.82	*SMYD1*	Histone-lysine N-methyltransferase	DNA binding; metal ion binding; transcription corepressor activity; histone-lysine N-methyltransferase activity	chromatin remodelling; heart development; negative regulation of transcription, DNA-templated; positive regulation of myoblast differentiation; positive regulation of myotube differentiation; skeletal muscle cell differentiation
1979	PP range	-0.449	9E-38	98.88	*TRIM68*	E3 ubiquitin-protein ligase TRIM68	androgen receptor binding; zinc ion binding; histone acetyltransferase binding; ubiquitin-protein transferase activity	protein autoubiquitination; regulation of androgen receptor signaling pathway; interferon-gamma-mediated signaling pathway
13546	PP range	-0.440	4E-36	97.75	*PIP5K1B*	Phosphatidylinositol 4-phosphate 5-kinase type-1 beta	ATP binding; 1-phosphatidylinositol-4-phosphate 5-kinase activity	phosphatidylinositol biosynthetic process; regulation of phosphatidylinositol 3-kinase signaling
8760	PP range	-0.432	9E-38	98.88	*KBTBD11*	Kelch repeat and BTB domain-containing protein 11	None given	None given

4604	PP range	-0.412	1E-36	98.85	*OSTF1*	Osteoclast-stimulating factor 1	SH3 domain binding	ossification; neutrophil degranulation; signal transduction
2386	PP range	-0.402	2E-39	100.00	*DEPDC1B*	DEP domain-containing protein 1B	GTPase activator activity	cell migration; intracellular signal transduction; positive regulation of Wnt signaling pathway; regulation of small GTPase mediated signal transduction
12443	PP range	0.524	2E-39	100.00	*GCNT2*	N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase	N-acetyllactosaminide beta-1,6-N-acetylglucosaminyltransferase activity	glycosaminoglycan biosynthetic process; maintenance of lens transparency; multicellular organism development; positive regulation of cell population proliferation; protein glycosylation; transforming growth factor beta receptor signaling pathway; positive regulation of epithelial mesenchymal transition; negative regulation of cell-substrate adhesion
9517	PP range	0.590	7E-39	100.00	*ELOVL2*	Elongation of very long chain fatty acids protein 2	fatty acid elongase activity; very-long-chain 3-ketoacyl-CoA synthase activity; 3-oxo-arachidoyl-CoA synthase activity; 3-oxo-cerotoyl-CoA synthase activity; 3-oxo-lignoceronyl-CoA synthase activity	fatty acid elongation, monounsaturated fatty acid; fatty acid elongation, polyunsaturated fatty acid; alpha-linolenic acid metabolic process; fatty acid elongation, saturated fatty acid; long-chain fatty-acyl-CoA biosynthetic process, unsaturated fatty acid biosynthetic process; sphingolipid biosynthetic process

C.iii. Methods

C.iii.1 Custom BASH scripts for filtering raw *T. aduncus* ddRAD sequences

```
Remove high missing data:
vcftools --vcf TotalRawSNPs.vcf --missing-indv
#create a document with the names of individuals with >20% missing data
nano missing.data.indiv
vcftools --vcf TotalRawSNPs.vcf --remove missing.data.indiv --recode --recode-
INFO-all --out FinalInd

Remove loci with general missing data >20% and a minor allele frequency <0.03:
vcftools --vcf FinalInd.recode.vcf --max-missing 0.8 --min-alleles 2 --max-
alleles 2 --out FinalInd80 --non-ref-af 0.001 --max-non-ref-af 0.9999 --mac 1 -
-maf 0.03 --minQ 30 --recode --recode-INFO-all &>VCFtools80.log
#check how many loci left
mawk '!/#/' FinalInd80.recode.vcf | wc -l

Remove complex haplotypes and INDELs:
vcfallelicprimitives FinalInd80.recode.vcf --keep-info --keep-geno >
FinalInd80_1.vcf
mawk '!/#/' FinalInd80_1.vcf | wc –l
vcftools --vcf FinalInd80_1.vcf --remove-indels --recode --recode-INFO-all --out
FinalInd80_2

Calculating mean coverage per locus:
#first calculate mean coverage per locus
vcftools --vcf FinalInd80_2.recode.vcf --site-depth --out FinalInd80_2
cut -f3 FinalInd80_2.ldepth > AverageDepthSite2
#x= sample size
mawk '!/D/' AverageDepthSite2| mawk -v x=123 '{print $1/x5' > meandepthpersite2
#calculate mean coverage plus 2 SD to use as a threshold
mawk '{ sum += $1; sumsq += ($1)^2; n++ } END
{if(n>0) print sum/n,sqrt((sumsq-sum^2 / n) / n);
}' meandepthpersite2
#x = average + 2 SD
vcftools --vcf FinalInd80_2.recode.vcf --recode-INFO-all --out FinalInd80_3 --
max-meanDP 65 --recode

Remove loci with allelic balance <0.20 or <0.80:
vcffilter -s -f "AB > 0.20 & AB < 0.80 | AB < 0.01" FinalInd80_3.recode.vcf >
FinalInd80_AB.recode.vcf
mawk '!/#/' FinalInd80_AB.recode.vcf | wc –l

Remove loci with mapping quality score <0.9 or <1.05:
vcffilter -f "MQM / MQMR > 0.9 & MQM / MQMR <1.05" FinalInd80_AB.recode.vcf >
FinalInd80_AB2.recode.vcf
mawk '!/#/' FinalInd80_AB2.recode.vcf | wc –l

Removing any locus that has a quality score below 0.2 (1/5) of the depth:
vcffilter -f "QUAL / DP > 0.20" FinalInd80_AB2.recode.vcf >
FinalInd80_AB3.recode.vcf
mawk '!/#/' FinalInd80_AB3.recode.vcf | wc –l

Removing SNPs with >10% missing data:
#(sample size x 0.1=max missing count)
```

vcftools --vcf FinalInd80_AB3_RR.recode.vcf --max-missing-count 4 --recode --recode-INFO-all --out FinalInd80_AB4

Removing loci outside of hardy-Weinberg Equilibrium:

```
#using a HWE (-h) value of 0.05 and a –c value (being the percentage of populations out of HWE) of the default of 25%
perl ~/Scripts/HWfilteringVCF.pl -v FinalInd80_AB4.recode.vcf -p ./popmap1.txt -h 0.05 -c 0.25 -o FinalInd80_HWE
#using HWfilteringVCF.pl script (custom script)
mawk '!/#/' FinalInd80_HWE.recode.vcf | wc –l
```

Extracting one locus per site:

```
perl ~/Scripts/ExtractBQFRLM_Elly.pl FinalInd80_HWE.recode.vcf BQ.vcf Frst.vcf Rdm.vcf LM.vcf
#using ExtractBQFRLM.pl script (custom script)
```

Removing loci in linkage disequilibrium:

```
vcftools --vcf BQ.vcf --interchrom-geno-r2 --min-r2 0.8
#using the out.interchrom.geno.ld output file extract the list of loci out of LD in R (LD_exclude_list.txt)
vcftools --vcf BQ.vcf --exclude-positions LD_exclude_list.txt --recode --recode-INFO-all --out BQ_LD
```

Turn final SNPs vcf file into fasta file:

```
perl ~/Scripts/ExtSeq.pl  ./reference.fasta ./BQ_LD.recode.vcf  ./BQ_LD.fasta
```

Aligning to the *T. aduncus* genome:

```
bowtie2 -x ~/Aduncus_seascape/GenomeAssembly/Tadunc_genome -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 -f ./BQ_LD.fasta -S ./ Tadunc_to_adunc_bowtie2_assembly --no-sq
#list of SNPs that did not align
nano Nonaligning_SNPs.txt
```

D. Chapter 4 Appendix

D.i. Figures

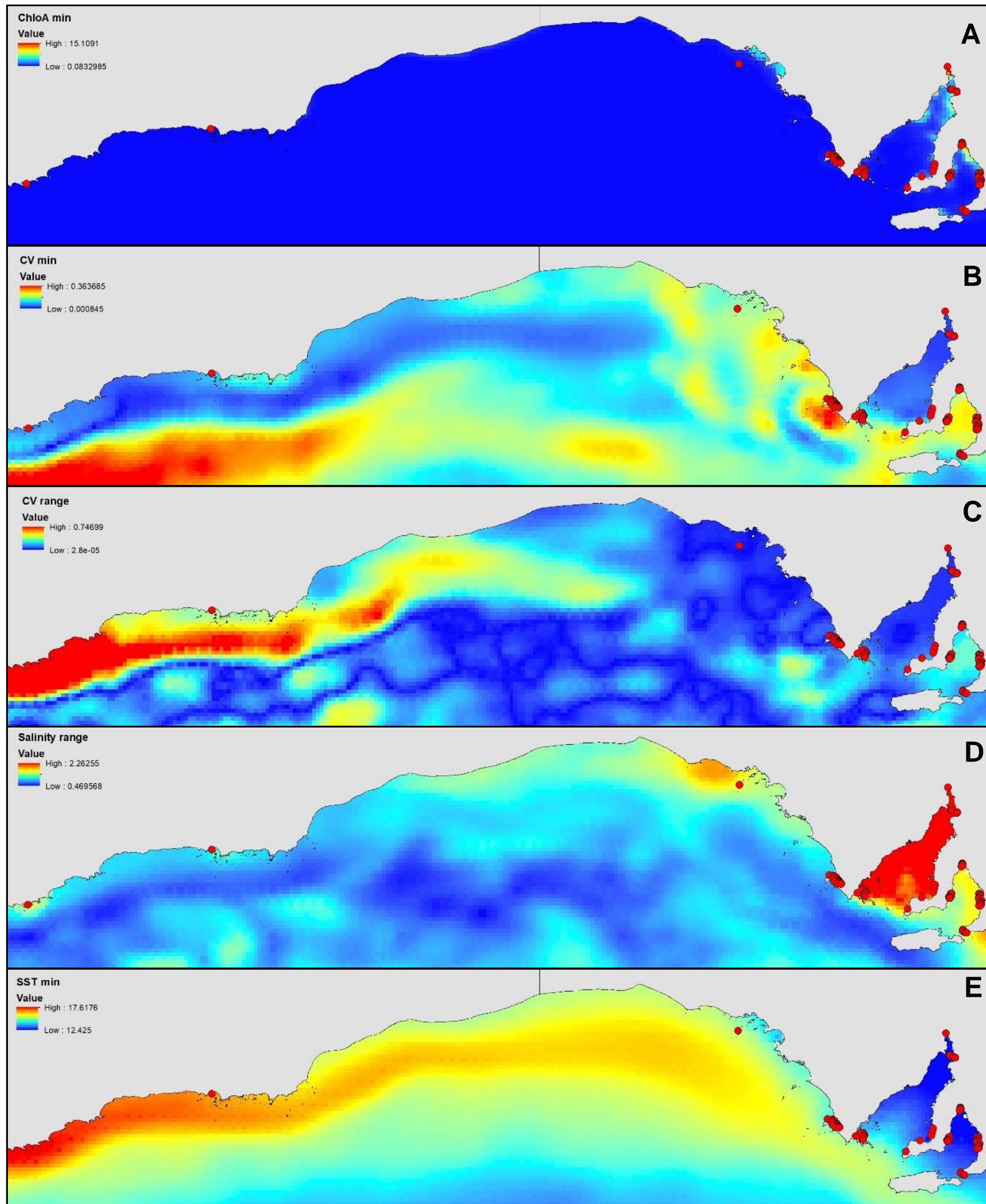


Figure D.i.1 Maps of spatial heterogeneity in the five environmental variables included in the final redundancy analysis (RDA) model run for genotype-environment association (GEA) analysis. **A)** chlorophyll A concentration (chloA) minimum; **B)** current velocity (CV) minimum; **C)** CV range; **D)** salinity range; and **E)** sea surface temperature (SST) minimum**.** Further information on these environmental variables can be found in Appendix D: Table D.ii.2.

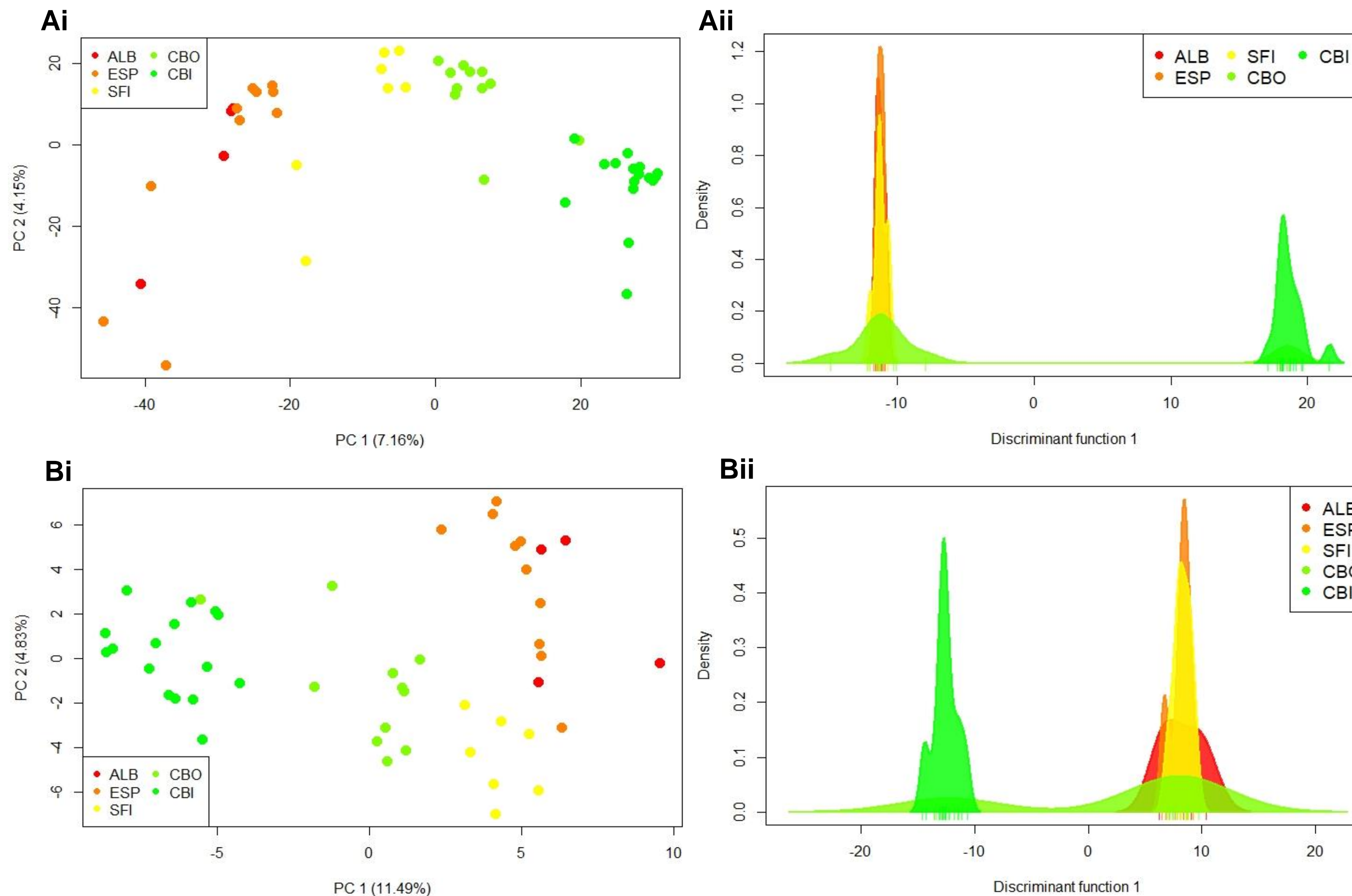


Figure D.i.2 Population genomic structure of southern Australian coastal bottlenose dolphins at sampling sites to the west of Eyre Peninsula, based on 48 individuals and **A)** 7,817 SNPs in the neutral dataset and **B)** 241 outlier SNPs in the adaptive dataset. **i)** Principal component analysis (PCA) and **ii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from west (top) to east (bottom) as per along-shore travel. Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

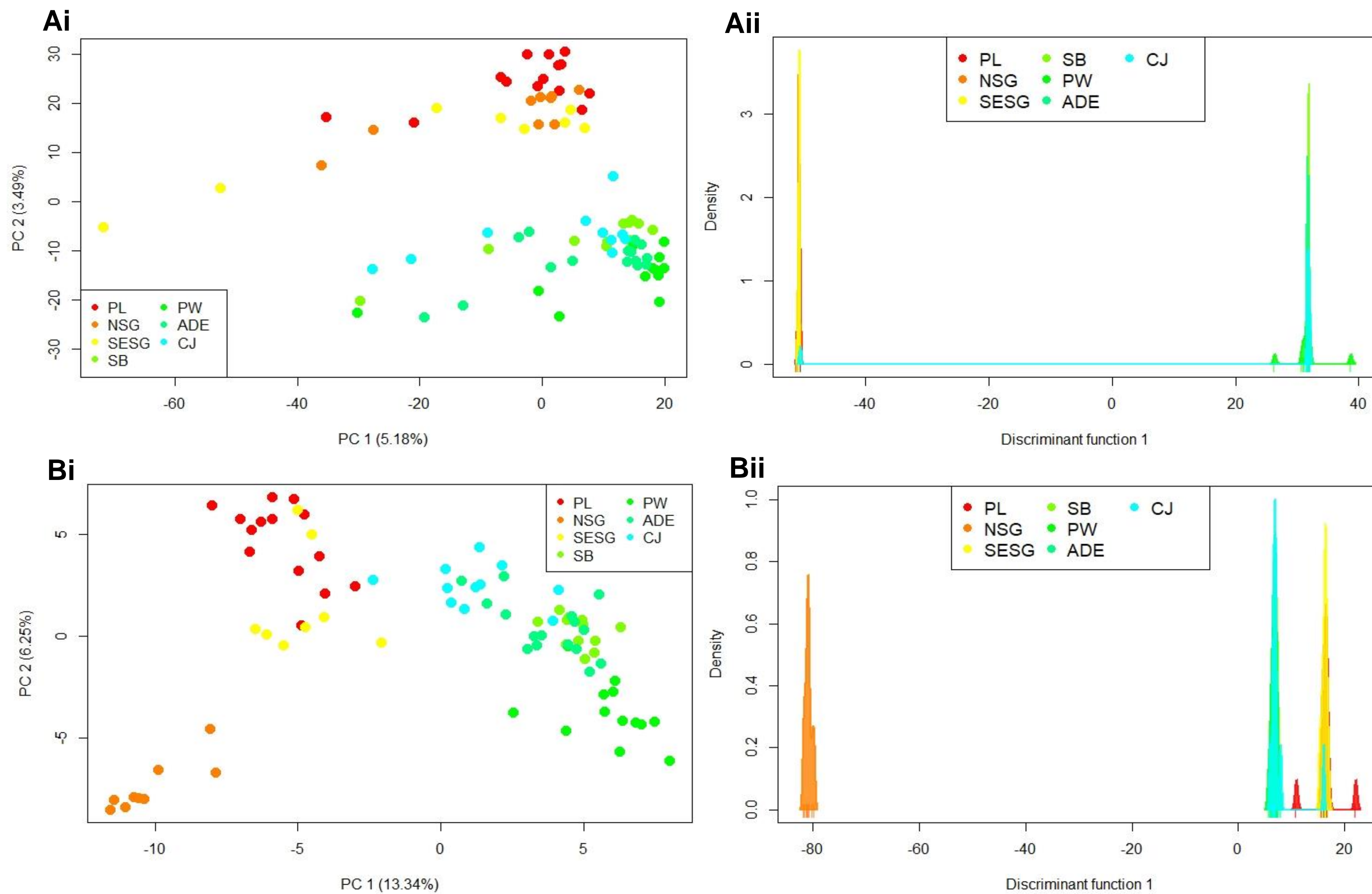


Figure D.i.3 Population genomic structure of southern Australian coastal bottlenose dolphins at sampling sites to the east of Eyre Peninsula, based on 83 individuals and **A)** 7,817 SNPs in the neutral dataset and **B)** 241 outlier SNPs in the adaptive dataset. **i)** Principal component analysis (PCA) and **ii)** discriminant analysis of principal components (DAPC) plots. Sampling locations are ordered from west (top) to east (bottom) as per along-shore travel. Sampling locality abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

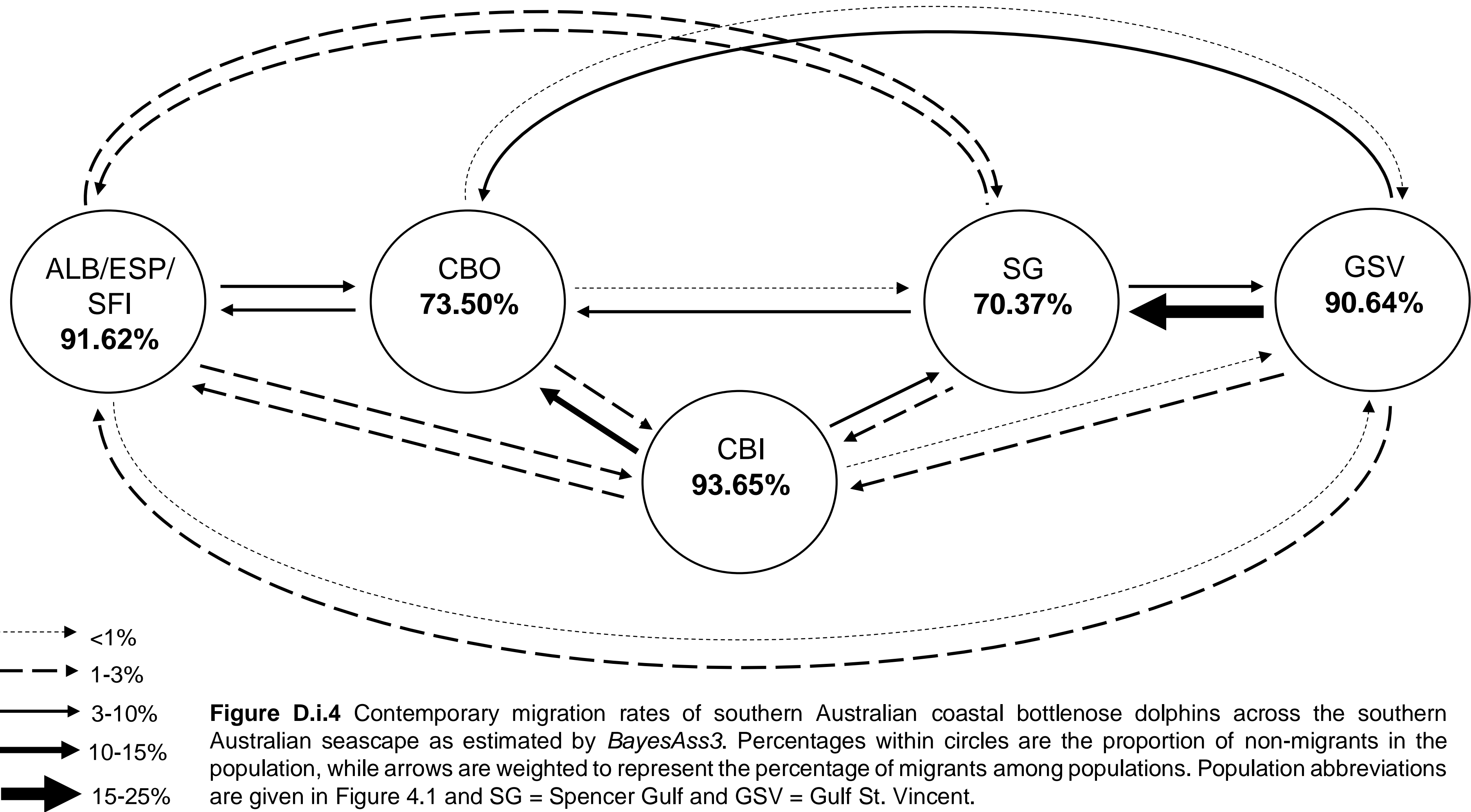


Figure D.i.4 Contemporary migration rates of southern Australian coastal bottlenose dolphins across the southern Australian seascape as estimated by *BayesAss3*. Percentages within circles are the proportion of non-migrants in the population, while arrows are weighted to represent the percentage of migrants among populations. Population abbreviations are given in Figure 4.1 and SG = Spencer Gulf and GSV = Gulf St. Vincent.

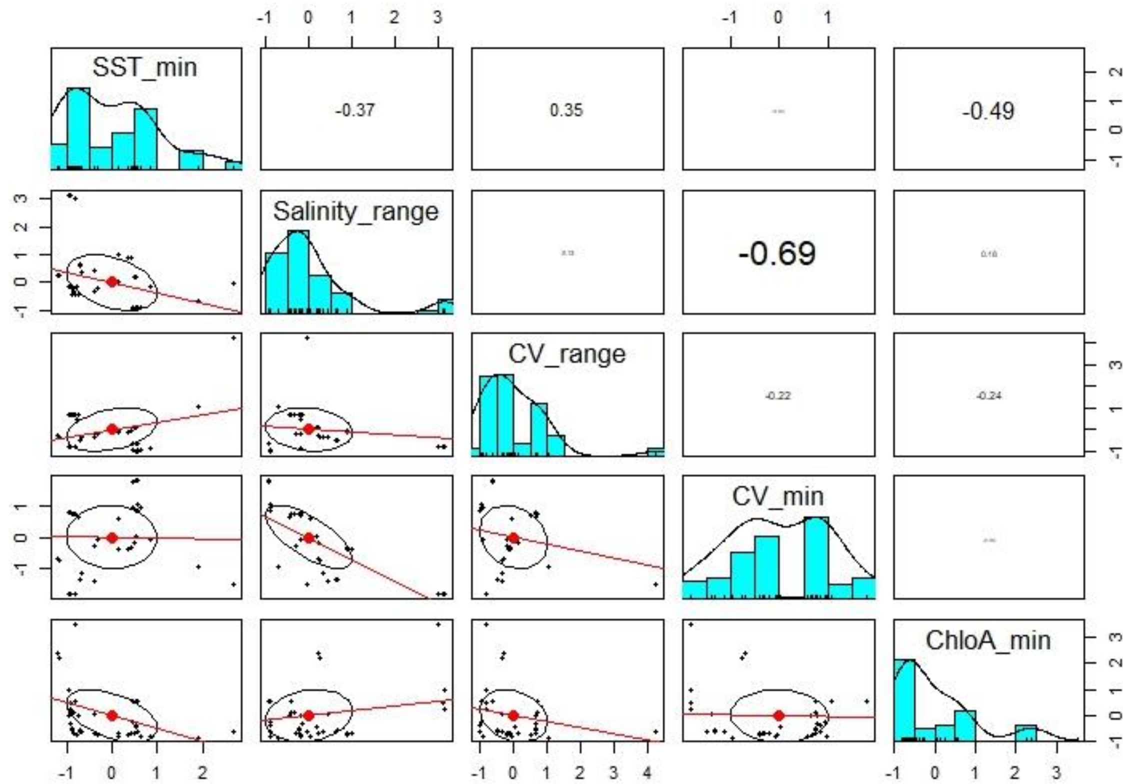


Figure D.i.5 Correlation of the final five environmental variables included in the redundancy analysis (RDA), retained after forward selection and collinearity analysis. Additional information of the ecological variables used are provided in Appendix D: Table D.ii.2 (SST = sea surface temperature, CV = current velocity, ChloA = chlorophyll A concentration; also note min = minimum).

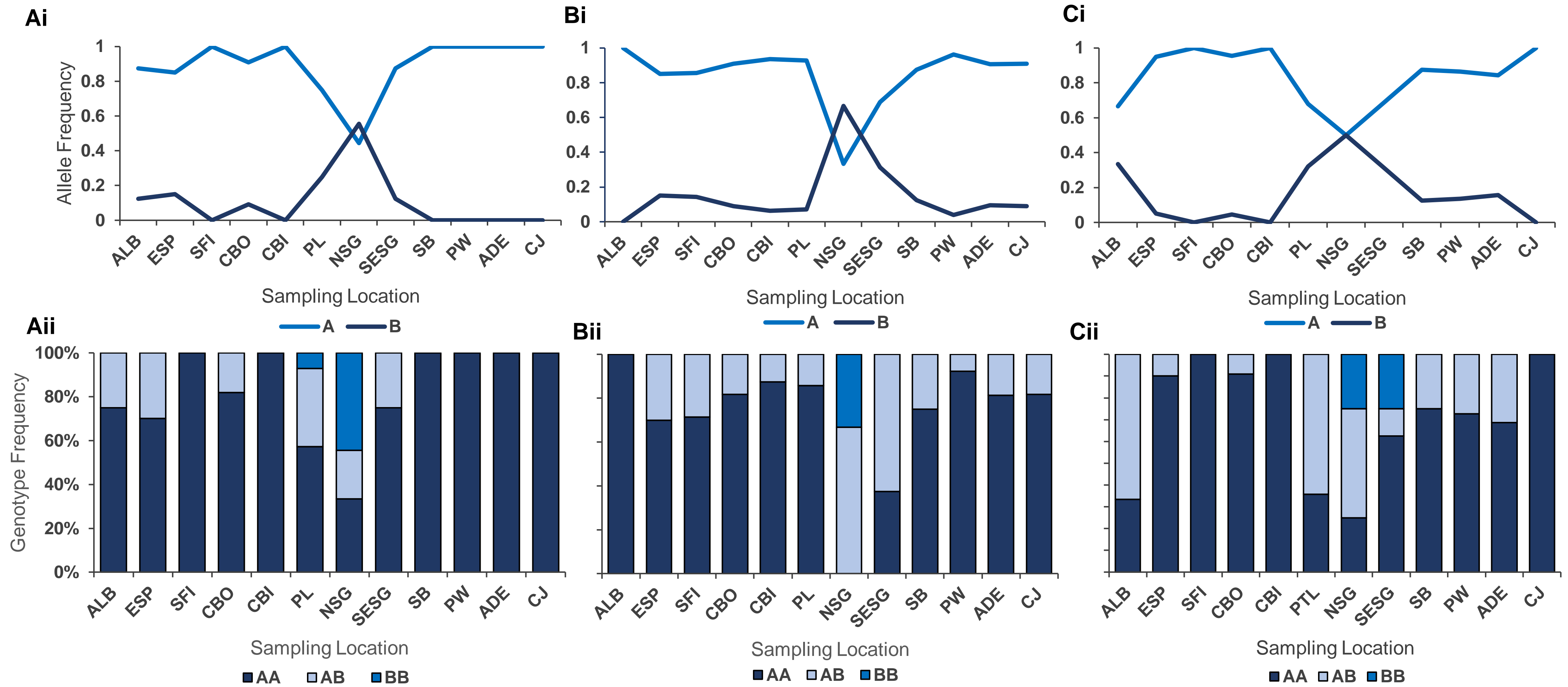


Figure D.i.6 Variation in three candidate genes putatively associated with adaptation to the salinity gradient in southern Australia: **A)** *KCNT2*, **B)** *RYR2* and **C)** *SLC22A18*, showing: **i)** allele frequency change and **ii)** genotype distribution over the southern Australian seascape. Sampling locations are ordered from west (left) to east (right) and abbreviations are explained in Figure 4.1 (and Appendix D: Table D.ii.3).

D.ii. Tables

Table D.ii.1 Southern Australian coastal bottlenose dolphin SNP quality filtering parameters and results.

	Average no. of reads per individual	2,674,445
	No. of SNPs before filtering	66,997
	Sample size before filtering	139
Filtering Parameter	<20% missing data per individual (final sample size)	131
	Variants called in >20% of individuals, minor allele frequency (MAF) >0.03, minimum quality score (MinQ) >30 & biallelic SNPs only	17,776 SNPs
	Removing complex variance and INDELs	18,148 SNPs
	Av. coverage < av. depth + 2 standard deviations	17,547 SNPs
	Allele balance 0.15 - 0.85 (proportion of alternate to reference alleles)	16,407 SNPs
	Mapping quality score 0.8 - 1.2	14,639 SNPs
	Read quality score >20% of read depth	13,103 SNPs
	<5% missing data per variant	11,291 SNPs
	<25% of populations out of HWE	10,673 SNPs
	Best quality SNP per locus	8,179 SNPs
	Linkage Disequilibrium - r^2 <0.8	8,104 SNPs
	Aligned to *T. aduncus* genome (final number of SNPs)	8,081 SNPs
	Adaptive Dataset (RDA candidates)	241 SNPs
	Neutral Dataset (RDA candidates + additional outliers removed)	7,817 SNPs

Table D.ii.2 Environmental, topographic and oceanographic variables used in the genotype-environment association (GEA) analysis.

Variable	Description	Spatial Resolution	Source	Temporal Range
Chlorophyll A (ChloA)	Chlorophyll A concentration at mean depth (mean annual maximum, mean, minimum and standard deviation). Units: mg/mŸ	5 arcmin (9.2 km)	Satellite (Aqua-MODIS) Ocean Color (Feldman and McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/ Accessed through Marine Geospatial Ecology Tools (MGET) in ArcGIS	2002-2009
Salinity	Model - Dissolved salt concentration at mean depth (mean annual maximum, mean, minimum and range). Units: PSS	0.25 arcdegree	Global Ocean Physics Reanalysis ECMWF ORAP5.0 URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE (URL: http://www.bio-oracle.org/downloads-to-email.php) (Tyberghein et al. 2012; Assis et al. 2018)	2000-2014
Sea Surface Temperature (SST)	Temperature in the topmost one metre of water column (mean annual maximum, mean, minimum and range) Units: ^{O}C	5 arcmin (9.2 km)	Satellite (Aqua-MODIS) Ocean Color (Feldman and McClain 2010) URL: http://oceancolor.gsfc.nasa.gov/ Accessed through Bio-ORACLE	2002-2009
Current Velocity (CV)	Model - Sea water velocity at mean depth (mean annual maximum, mean, minimum and range) Units: m/s	0.25 arcdegree	Global Ocean Physics Reanalysis ECMWF ORAP5.0 URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE	2000-2014
Primary Production (PP)	Model - Net primary productivity of carbon at mean depth (mean annual maximum, mean, minimum and range) Units: g/mŸ/day	0.25 arcdegree	Global Ocean Biogeochemistry Non assimilative hindcast (Pisces) URL: http://marine.copernicus.eu/ Accessed through Bio-ORACLE	2000-2014

Bathymetry	Average depth of seafloor Units: Metres (m)	9 arc second (0.0025Ao or ~250m at equator)	Geoscience Australia Australian Bathymetry and Topography Grid, June 2009 URL: https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search?node=srv#/metadata/a05f7892-fae9-7506-e044-00144fdd4fa6 Accessed through data.gov.au portal	2009
Seafloor Rugosity	Surface area of 3x3 cells of the Aust Bathymetry and Topography Grid above	9 arc second (0.0025A° or ~250m at equator)	Geoscience Australia Bathymetry derived topographic rugosity grid – created from the Australian Bathymetry and Topography Grid, June 2009 URL: https://data.gov.au/dataset/bathymetry-derived-topographic-rugosity-grid Accessed through data.gov.au portal	2017
Seafloor Slope	Steepness of the seafloor (degree of incline)	1km^2	Geoscience Australia Australian bathymetry and its derivatives URL: https://data.gov.au/dataset/australian-bathymetry-and-its-derivatives Accessed through data.gov.au portal	2017
Topographic Relief	Amount of topographic change on the seafloor	1km^2	Geoscience Australia Australian bathymetry and its derivatives URL: https://data.gov.au/dataset/australian-bathymetry-and-its-derivatives Accessed through data.gov.au portal	2017

Table D.ii.3 Genomic diversity indices for sampling locations of southern Australian coastal bottlenose dolphins, based on 131 individuals and 7,817 SNPs for the neutral dataset and 241 SNPs for the adaptive dataset. N = sample size before filtering (M:F represents the number of samples that are male or female), %PL = percentage of polymorphic loci, H_O = observed heterozygosity, H_E = expected heterozygosity and F_{IS} = Wright's inbreeding coefficient. Standard deviations are shown in parentheses.

Site	Year Sampled	N (M:F)	Adaptive Dataset			Neutral Dataset			
			H_O	H_E	% PL	H_O	H_E	% PL	F_{IS}
Albany (ALB)	2010	4 (unknown)	0.481 (0.245)	0.420 (0.122)	38.56	0.430 (0.226)	0.393 (0.129)	28.13	-0.094
Esperance (ESP)	2010	10 (5:5)	0.389 (0.175)	0.393 (0.132)	43.24	0.303 (0.185)	0.304 (0.150)	36.14	0.003
St. Francis Island (SFI)	2005	10 (5:5)	0.397 (0.206)	0.393 (0.133)	37.72	0.327 (0.199)	0.331 (0.143)	33.75	0.011
Outer Coffin Bay (CBO)	2014	11 (8:3)	0.371 (0.171)	0.383 (0.126)	45.81	0.290 (0.183)	0.289 (0.151)	36.04	-0.007
Inner Coffin Bay (CBI)	2014	16 (3:13)	0.391 (0.176)	0.361 (0.134)	42.02	0.291 (0.187)	0.275 (0.154)	36.05	-0.057
Pt. Lincoln (PL)	2005	14 (9:5)	0.442 (0.173)	0.421 (0.104)	44.73	0.285 (0.180)	0.273 (0.148)	39.40	-0.043
Northern Spencer Gulf (NSG)	2005	11 (8:3)	0.367 (0.175)	0.411 (0.128)	45.92	0.280 (0.180)	0.299 (0.149)	34.56	0.063
Southeast Spencer Gulf (SESG)	2004	11 (3:8)	0.424 (0.188)	0.429 (0.104)	46.96	0.301 (0.189)	0.305 (0.142)	38.27	0.013
Stansbury (SB)	2015	12 (10:2)	0.374 (0.169)	0.386 (0.134)	42.36	0.281 (0.177)	0.281 (0.150)	35.66	-0.000
Pt. Wakefield (PW)	2015	13 (8:5)	0.354 (0.191)	0.347 (0.141)	44.40	0.280 (0.183)	0.274 (0.156)	35.58	-0.020
Adelaide (ADE)	2013-2014	16 (7:9)	0.434 (0.175)	0.396 (0.123)	43.78	0.276 (0.184)	0.258 (0.151)	40.48	-0.072
Cape Jervis (CJ)	2015	11 (6:5)	0.434 (0.176)	0.406 (0.120)	44.16	0.305 (0.192)	0.281 (0.147)	39.30	-0.085
		Average	0.405 (0.038)	0.395 (0.024)	43.30 (2.81)	0.304 (0.042)	0.297 (0.036)	36.11 (3.26)	0.024 (0.047)

Table D.ii.4 Analysis of molecular variance (AMOVA) based on **A)** 7,817 neutral loci and **B)** 241 putatively adaptive loci testing the amount of genomic variation explained among four populations: 1) Western Australia, SFI and CBO, 2) CBI, 3) SG and 4) GSV, compared to among individual sampling localities of southern Australian coastal bottlenose dolphins. Values significant at 0.001 are marked by an asterisk (*).

Source of Variation	Degrees of Freedom	Sum of Squares	Variance Components	Percentage of Variation
A				
Among populations	3	5,607.25	17.64	4.06*
Among sampling locations within populations	8	5,660.94	14.95	3.44*
Within pops	250	10,0471.97	401.89	92.50*
Total	261	111,740.15	434.48	
B				
Among populations	3	708.16	2.95	12.03*
Among sampling locations within populations	8	374.77	1.30	5.29*
Within pops	250	5,074.78	20.30	82.68*
Total	261	6,157.70	24.55	

Table D.ii.5 Analysis of variance (ANOVA) for testing of genotype-environment association (GEA) using redundancy analysis (RDA) based on 8,081 loci in southern Australian bottlenose dolphins testing the amount of genomic variation explained by the overall model and each of the ecological variables.

	Variable	**Degrees of Freedom**	**Variance Explained**	**F**	***p***
Overall	Model	5	186.1016328	2.6892	0.001
	Residual	123	1,702.4041		
Marginal	Sea surface temperature (SST) minimum	1	31.2681	2.2591	0.001
	Salinity range	1	26.1428	1.8888	0.001
	Current velocity range	1	24.9058	1.7995	0.001
	Current velocity minimum	1	23.921	1.7283	0.001
	Chlorophyll A concentration minimum	1	22.9024	1.6547	0.001
	Residual	123	1,702.4041		

Table D.ii.6 Gene ontology (GO) terms that were significantly over-enriched ($p < 0.05$) in the putatively adaptive dataset as compared to the full (reference) dataset (8,081 SNPs). Determined by a functional enrichment analysis using a Fisher's exact test in Blast2GO.

GO Category	GO ID	GO Term	*p*	Associated genes in: Candidate Dataset (%)	Reference Dataset (%)
Biological Process	GO:0001659	temperature homeostasis	0.0298	0.4167	0.0000
	GO:0002065	columnar/cuboidal epithelial cell differentiation	0.0298	0.4167	0.0000
	GO:0002066	columnar/cuboidal epithelial cell development	0.0298	0.4167	0.0000
	GO:0002067	glandular epithelial cell differentiation	0.0298	0.4167	0.0000
	GO:0002068	glandular epithelial cell development	0.0298	0.4167	0.0000
	GO:0003309	type B pancreatic cell differentiation	0.0298	0.4167	0.0000
	GO:0003323	type B pancreatic cell development	0.0298	0.4167	0.0000
	GO:0006488	dolichol-linked oligosaccharide biosynthetic process	0.0298	0.4167	0.0000
	GO:0006489	dolichyl diphosphate biosynthetic process	0.0298	0.4167	0.0000
	GO:0006490	oligosaccharide-lipid intermediate biosynthetic process	0.0298	0.4167	0.0000
	GO:0007196	adenylate cyclase-inhibiting G protein-coupled glutamate receptor signaling pathway	0.0298	0.4167	0.0000
	GO:0007215	glutamate receptor signaling pathway	0.0298	0.4167	0.0000
	GO:0007216	G protein-coupled glutamate receptor signaling pathway	0.0298	0.4167	0.0000
	GO:0007567	parturition	0.0298	0.4167	0.0000
	GO:0010586	miRNA metabolic process	0.0298	0.4167	0.0000
	GO:0010587	miRNA catabolic process	0.0298	0.4167	0.0000
	GO:0016441	posttranscriptional gene silencing	0.0298	0.4167	0.0000
	GO:0030856	regulation of epithelial cell differentiation	0.0298	0.4167	0.0000
	GO:0031018	endocrine pancreas development	0.0298	0.4167	0.0000
	GO:0031047	gene silencing by RNA	0.0298	0.4167	0.0000
	GO:0031050	dsRNA processing	0.0298	0.4167	0.0000
	GO:0031054	pre-miRNA processing	0.0298	0.4167	0.0000
	GO:0031123	RNA 3'-end processing	0.0298	0.4167	0.0000
	GO:0031640	killing of cells of other organism	0.0298	0.4167	0.0000
	GO:0031958	corticosteroid receptor signaling pathway	0.0298	0.4167	0.0000
	GO:0032008	positive regulation of TOR signaling	0.0298	0.4167	0.0000

GO:0034661	ncRNA catabolic process	0.0298	0.4167	0.0000
GO:0035194	posttranscriptional gene silencing by RNA	0.0298	0.4167	0.0000
GO:0035195	gene silencing by miRNA	0.0298	0.4167	0.0000
GO:0035196	production of miRNAs involved in gene silencing by miRNA	0.0298	0.4167	0.0000
GO:0035235	ionotropic glutamate receptor signaling pathway	0.0298	0.4167	0.0000
GO:0035733	hepatic stellate cell activation	0.0298	0.4167	0.0000
GO:0035821	modification of morphology or physiology of other organism	0.0298	0.4167	0.0000
GO:0035883	enteroendocrine cell differentiation	0.0298	0.4167	0.0000
GO:0042634	regulation of hair cycle	0.0298	0.4167	0.0000
GO:0042752	regulation of circadian rhythm	0.0298	0.4167	0.0000
GO:0042753	positive regulation of circadian rhythm	0.0298	0.4167	0.0000
GO:0042921	glucocorticoid receptor signaling pathway	0.0298	0.4167	0.0000
GO:0044364	disruption of cells of other organism	0.0298	0.4167	0.0000
GO:0045844	positive regulation of striated muscle tissue development	0.0298	0.4167	0.0000
GO:0046465	dolichyl diphosphate metabolic process	0.0298	0.4167	0.0000
GO:0048636	positive regulation of muscle organ development	0.0298	0.4167	0.0000
GO:0048643	positive regulation of skeletal muscle tissue development	0.0298	0.4167	0.0000
GO:0051775	response to redox state	0.0298	0.4167	0.0000
GO:0051967	negative regulation of synaptic transmission, glutamatergic	0.0298	0.4167	0.0000
GO:0060078	regulation of postsynaptic membrane potential	0.0298	0.4167	0.0000
GO:0060137	maternal process involved in parturition	0.0298	0.4167	0.0000
GO:0060147	regulation of posttranscriptional gene silencing	0.0298	0.4167	0.0000
GO:0060148	positive regulation of posttranscriptional gene silencing	0.0298	0.4167	0.0000
GO:0060964	regulation of gene silencing by miRNA	0.0298	0.4167	0.0000
GO:0060966	regulation of gene silencing by RNA	0.0298	0.4167	0.0000
GO:0061737	leukotriene signaling pathway	0.0298	0.4167	0.0000
GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	0.0298	0.4167	0.0000
GO:0070918	production of small RNA involved in gene silencing by RNA	0.0298	0.4167	0.0000
GO:0070920	regulation of production of small RNA involved in gene silencing by RNA	0.0298	0.4167	0.0000
GO:0071322	cellular response to carbohydrate stimulus	0.0298	0.4167	0.0000
GO:0071326	cellular response to monosaccharide stimulus	0.0298	0.4167	0.0000

GO:0071331	cellular response to hexose stimulus	0.0298	0.4167	0.0000
GO:0071333	cellular response to glucose stimulus	0.0298	0.4167	0.0000
GO:0072537	fibroblast activation	0.0298	0.4167	0.0000
GO:0090184	positive regulation of kidney development	0.0298	0.4167	0.0000
GO:0090403	oxidative stress-induced premature senescence	0.0298	0.4167	0.0000
GO:0099505	regulation of presynaptic membrane potential	0.0298	0.4167	0.0000
GO:0106106	cold-induced thermogenesis	0.0298	0.4167	0.0000
GO:0120161	regulation of cold-induced thermogenesis	0.0298	0.4167	0.0000
GO:0120163	negative regulation of cold-induced thermogenesis	0.0298	0.4167	0.0000
GO:1901724	positive regulation of cell proliferation involved in kidney development	0.0298	0.4167	0.0000
GO:1901863	positive regulation of muscle tissue development	0.0298	0.4167	0.0000
GO:1901985	positive regulation of protein acetylation	0.0298	0.4167	0.0000
GO:1903798	regulation of production of miRNAs involved in gene silencing by miRNA	0.0298	0.4167	0.0000
GO:1903800	positive regulation of production of miRNAs involved in gene silencing by miRNA	0.0298	0.4167	0.0000
GO:1990845	adaptive thermogenesis	0.0298	0.4167	0.0000
GO:2000074	regulation of type B pancreatic cell development	0.0298	0.4167	0.0000
GO:2000322	regulation of glucocorticoid receptor signaling pathway	0.0298	0.4167	0.0000
GO:2000323	negative regulation of glucocorticoid receptor signaling pathway	0.0298	0.4167	0.0000
GO:2000489	regulation of hepatic stellate cell activation	0.0298	0.4167	0.0000
GO:2000491	positive regulation of hepatic stellate cell activation	0.0298	0.4167	0.0000
GO:2000637	positive regulation of gene silencing by miRNA	0.0298	0.4167	0.0000
GO:2001016	positive regulation of skeletal muscle cell differentiation	0.0298	0.4167	0.0000
GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	0.0340	0.8368	0.1021
GO:0007188	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	0.0278	0.8368	0.0894
GO:0007193	adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	0.0083	0.8368	0.0383
GO:0031929	TOR signaling	0.0051	0.8368	0.0255
GO:0032006	regulation of TOR signaling	0.0051	0.8368	0.0255
GO:0006396	RNA processing	0.0485	3.4335	1.6466
GO:0006401	RNA catabolic process	0.0434	3.4335	1.6071
GO:0006473	protein acetylation	0.0434	3.4335	1.6071
GO:0006644	phospholipid metabolic process	0.0468	3.4335	1.6334

GO:0006959	humoral immune response	0.0388	3.4335	1.5676
GO:0008654	phospholipid biosynthetic process	0.0418	3.4335	1.5939
GO:0010466	negative regulation of peptidase activity	0.0403	3.4335	1.5807
GO:0010951	negative regulation of endopeptidase activity	0.0403	3.4335	1.5807
GO:0019439	aromatic compound catabolic process	0.0485	3.4335	1.6466
GO:0030162	regulation of proteolysis	0.0485	3.4335	1.6466
GO:0034655	nucleobase-containing compound catabolic process	0.0451	3.4335	1.6202
GO:0034660	ncRNA metabolic process	0.0451	3.4335	1.6202
GO:0043543	protein acylation	0.0451	3.4335	1.6202
GO:0045861	negative regulation of proteolysis	0.0418	3.4335	1.5939
GO:0051346	negative regulation of hydrolase activity	0.0485	3.4335	1.6466
GO:0052547	regulation of peptidase activity	0.0451	3.4335	1.6202
GO:0052548	regulation of endopeptidase activity	0.0451	3.4335	1.6202
GO:0007186	G protein-coupled receptor signaling pathway	0.0259	3.8793	1.7257
GO:0009891	positive regulation of biosynthetic process	0.0420	3.8793	1.8976
GO:0031324	negative regulation of cellular metabolic process	0.0497	3.8793	1.9638
GO:0031328	positive regulation of cellular biosynthetic process	0.0420	3.8793	1.8976
GO:0048699	generation of neurons	0.0497	3.8793	1.9638
GO:0051172	negative regulation of nitrogen compound metabolic process	0.0481	3.8793	1.9506
GO:0051336	regulation of hydrolase activity	0.0327	3.8793	1.8050
GO:0006412	translation	0.0340	4.3290	2.1232
GO:0006518	peptide metabolic process	0.0363	4.3290	2.1498
GO:0043043	peptide biosynthetic process	0.0352	4.3290	2.1365
GO:0043603	cellular amide metabolic process	0.0388	4.3290	2.1765
GO:0043604	amide biosynthetic process	0.0352	4.3290	2.1365
GO:0065008	regulation of biological quality	0.0499	4.3290	2.2831
GO:1901565	organonitrogen compound catabolic process	0.0401	4.3290	2.1898
GO:0006351	transcription, DNA-templated	0.0494	4.7826	2.6044
GO:0006355	regulation of transcription, DNA-templated	0.0494	4.7826	2.6044
GO:0006508	proteolysis	0.0249	4.7826	2.3098
GO:0009056	catabolic process	0.0257	4.7826	2.3232

	GO:0009057	macromolecule catabolic process	0.0173	4.7826	2.1765
	GO:0010033	response to organic substance	0.0403	4.7826	2.5105
	GO:0035556	intracellular signal transduction	0.0494	4.7826	2.6044
	GO:0044248	cellular catabolic process	0.0224	4.7826	2.2698
	GO:0044265	cellular macromolecule catabolic process	0.0149	4.7826	2.1232
	GO:0071310	cellular response to organic substance	0.0357	4.7826	2.4569
	GO:1901575	organic substance catabolic process	0.0240	4.7826	2.2965
	GO:0006955	immune response	0.0471	5.7018	3.2530
	GO:0007154	cell communication	0.0483	5.7018	3.2666
	GO:0007165	signal transduction	0.0425	5.7018	3.1986
	GO:0023052	signaling	0.0483	5.7018	3.2666
Cellular Component	GO:0000932	P-body	0.0298	0.4167	0.0000
	GO:0005791	rough endoplasmic reticulum	0.0298	0.4167	0.0000
	GO:0005844	polysome	0.0298	0.4167	0.0000
	GO:0008328	ionotropic glutamate receptor complex	0.0298	0.4167	0.0000
	GO:0009898	cytoplasmic side of plasma membrane	0.0298	0.4167	0.0000
	GO:0019897	extrinsic component of plasma membrane	0.0298	0.4167	0.0000
	GO:0031234	extrinsic component of cytoplasmic side of plasma membrane	0.0298	0.4167	0.0000
	GO:0032983	kainate selective glutamate receptor complex	0.0298	0.4167	0.0000
	GO:0033391	chromatoid body	0.0298	0.4167	0.0000
	GO:0043195	terminal bouton	0.0298	0.4167	0.0000
	GO:0043679	axon terminus	0.0298	0.4167	0.0000
	GO:0044306	neuron projection terminus	0.0298	0.4167	0.0000
	GO:0098562	cytoplasmic side of membrane	0.0298	0.4167	0.0000
	GO:0098839	postsynaptic density membrane	0.0298	0.4167	0.0000
	GO:0098878	neurotransmitter receptor complex	0.0298	0.4167	0.0000
	GO:0099634	postsynaptic specialization membrane	0.0298	0.4167	0.0000
	GO:1904423	dehydrodolichyl diphosphate synthase complex	0.0298	0.4167	0.0000
	GO:0035770	ribonucleoprotein granule	0.0026	0.8368	0.0128
	GO:0036464	cytoplasmic ribonucleoprotein granule	0.0026	0.8368	0.0128
	GO:0005667	transcription factor complex	0.0485	3.4335	1.6466

	GO:0034702	ion channel complex	0.0451	3.4335	1.6202
	GO:1902495	transmembrane transporter complex	0.0468	3.4335	1.6334
	GO:1990351	transporter complex	0.0468	3.4335	1.6334
	GO:0005694	chromosome	0.0392	3.8793	1.8711
	GO:1990904	ribonucleoprotein complex	0.0137	4.7826	2.0966
Molecular Function	GO:0001640	adenylate cyclase inhibiting G protein-coupled glutamate receptor activity	0.0298	0.4167	0.0000
	GO:0002094	polyprenyltransferase activity	0.0298	0.4167	0.0000
	GO:0002151	G-quadruplex RNA binding	0.0298	0.4167	0.0000
	GO:0004659	prenyltransferase activity	0.0298	0.4167	0.0000
	GO:0004712	protein serine/threonine/tyrosine kinase activity	0.0298	0.4167	0.0000
	GO:0004869	cysteine-type endopeptidase inhibitor activity	0.0298	0.4167	0.0000
	GO:0004970	ionotropic glutamate receptor activity	0.0298	0.4167	0.0000
	GO:0005230	extracellular ligand-gated ion channel activity	0.0298	0.4167	0.0000
	GO:0008066	glutamate receptor activity	0.0298	0.4167	0.0000
	GO:0015277	kainate selective glutamate receptor activity	0.0298	0.4167	0.0000
	GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	0.0298	0.4167	0.0000
	GO:0017162	aryl hydrocarbon receptor binding	0.0298	0.4167	0.0000
	GO:0022824	transmitter-gated ion channel activity	0.0298	0.4167	0.0000
	GO:0022835	transmitter-gated channel activity	0.0298	0.4167	0.0000
	GO:0031490	chromatin DNA binding	0.0298	0.4167	0.0000
	GO:0031491	nucleosome binding	0.0298	0.4167	0.0000
	GO:0031492	nucleosomal DNA binding	0.0298	0.4167	0.0000
	GO:0035198	miRNA binding	0.0298	0.4167	0.0000
	GO:0043027	cysteine-type endopeptidase inhibitor activity involved in apoptotic process	0.0298	0.4167	0.0000
	GO:0043028	cysteine-type endopeptidase regulator activity involved in apoptotic process	0.0298	0.4167	0.0000
	GO:0045547	dehydrodolichyl diphosphate synthase activity	0.0298	0.4167	0.0000
	GO:0061980	regulatory RNA binding	0.0298	0.4167	0.0000
	GO:0070491	repressing transcription factor binding	0.0298	0.4167	0.0000
	GO:0070888	E-box binding	0.0298	0.4167	0.0000
	GO:0098960	postsynaptic neurotransmitter receptor activity	0.0298	0.4167	0.0000
	GO:0098988	G protein-coupled glutamate receptor activity	0.0298	0.4167	0.0000
	GO:0099507	ligand-gated ion channel activity involved in regulation of presynaptic membrane potential	0.0298	0.4167	0.0000

GO:0099529	neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential	0.0298	0.4167	0.0000
GO:1904315	transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	0.0298	0.4167	0.0000
GO:1905538	polysome binding	0.0298	0.4167	0.0000
GO:1990825	sequence-specific mRNA binding	0.0298	0.4167	0.0000
GO:0000978	RNA polymerase II proximal promoter sequence-specific DNA binding	0.0485	3.4335	1.6466
GO:0000987	proximal promoter sequence-specific DNA binding	0.0485	3.4335	1.6466
GO:0003924	GTPase activity	0.0418	3.4335	1.5939
GO:0004857	enzyme inhibitor activity	0.0434	3.4335	1.6071
GO:0004866	endopeptidase inhibitor activity	0.0388	3.4335	1.5676
GO:0022836	gated channel activity	0.0485	3.4335	1.6466
GO:0022839	ion gated channel activity	0.0468	3.4335	1.6334
GO:0030414	peptidase inhibitor activity	0.0388	3.4335	1.5676
GO:0030594	neurotransmitter receptor activity	0.0388	3.4335	1.5676
GO:0030695	GTPase regulator activity	0.0434	3.4335	1.6071
GO:0060589	nucleoside-triphosphatase regulator activity	0.0451	3.4335	1.6202
GO:0061134	peptidase regulator activity	0.0388	3.4335	1.5676
GO:0061135	endopeptidase regulator activity	0.0388	3.4335	1.5676
GO:0030234	enzyme regulator activity	0.0270	3.8793	1.7389
GO:0003700	DNA-binding transcription factor activity	0.0469	4.3290	2.2564
GO:0003723	RNA binding	0.0455	4.3290	2.2431
GO:0005102	signaling receptor binding	0.0469	4.3290	2.2564
GO:0008270	zinc ion binding	0.0352	4.3290	2.1365
GO:0015075	ion transmembrane transporter activity	0.0414	4.3290	2.2031
GO:0015318	inorganic molecular entity transmembrane transporter activity	0.0414	4.3290	2.2031
GO:0016462	pyrophosphatase activity	0.0328	4.3290	2.1099
GO:0016817	hydrolase activity, acting on acid anhydrides	0.0328	4.3290	2.1099
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	0.0328	4.3290	2.1099
GO:0017111	nucleoside-triphosphatase activity	0.0328	4.3290	2.1099
GO:0022857	transmembrane transporter activity	0.0499	4.3290	2.2831
GO:0043565	sequence-specific DNA binding	0.0455	4.3290	2.2431
GO:0046914	transition metal ion binding	0.0414	4.3290	2.2031
GO:0003677	DNA binding	0.0335	4.7826	2.4301

GO:0098772	molecular function regulator	0.0257	4.7826	2.3232

Table D.ii.7 Functional annotation and gene ontology (GO) terms associated with candidate loci identified by genotype-environment association (GEA) analysis to be moderately to highly correlated with salinity range or minimum sea surface temperature (SST).

Candidate Locus	Correlated Variable	Correlation Coefficient	e-value	% Identity	Candidate Gene	Protein	GO terms	
							Molecular Function	Biological Process
6022	Salinity range	-0.541	2E-39	100.00	*KCNT2*	Potassium channel subfamily T member 2	ATP binding; outward rectifier potassium channel activity; intracellular sodium activated potassium channel activity; chloride-activated potassium channel activity	regulation of membrane potential; potassium ion export across plasma membrane
5274	Salinity range	-0.520	9E-38	98.88	*TTC21B*	Tetratricopeptide repeat protein 21B	None given	forebrain dorsal/ventral pattern formation; intraciliary retrograde transport; ventricular system development; protein localization to cilium
6042	Salinity range	-0.512	4E-36	97.75	*RYR2*	Ryanodine receptor 2 (also found in Chapter 2)	calcium channel activity; calcium ion binding; ion channel binding; enzyme binding; calmodulin binding	calcium ion transport; calcium-mediated signaling; cardiac muscle contraction; cardiac muscle hypertrophy; positive regulation of heart rate; regulation of cardiac muscle contraction; regulation of heart rate; response to hypoxia, response to muscle activity
4929	Salinity range	-0.461	2E-35	97.75	*SLC22A18*	Solute carrier family 22 member 18	transmembrane transporter activity; symporter activity; ubiquitin protein ligase binding	drug transport; excretion; organic cation transport; xenobiotic transport; xenobiotic detoxification by transmembrane export across the plasma membrane
6124	Salinity range	-0.458	2E-39	100.00	*GRIK3*	Glutamate receptor ionotropic, kainate 3	glutamate receptor activity; transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential; ligand-gated ion channel activity involved in regulation of presynaptic membrane potential	glutamate receptor signaling pathway; modulation of chemical synaptic transmission; regulation of membrane potential
8081	Salinity range	-0.456	9E-38	98.88	*TULP4*	Tubby-related protein 4	None given	post-translational protein modification; protein ubiquitination

2827	Salinity range	-0.453	9E-38	98.88	*SMOC1*	SPARC-related modular calcium-binding protein 1	calcium ion binding; extracellular matrix binding	cell differentiation; eye development; limb development; regulation of osteoblast differentiation
1664	Salinity range	-0.435	2E-39	100.00	*RNF130*	E3 ubiquitin-protein ligase RNF130	metal ion binding; ubiquitin protein ligase activity; ubiquitin-protein transferase activity	apoptotic process; programmed cell death; ubiquitin-dependent protein catabolic process
2449	Salinity range	-0.414	2E-39	100.00	*PFDN4*	Prefoldin subunit 4	unfolded protein binding; chaperone binding	protein folding
1135	Salinity range	-0.413	4E-36	97.75	*IQCA1*	Dynein regulatory complex protein 11	ATP binding	None given
7911	Salinity range	-0.411	1E-36	98.85	*SEC14L1*	SEC14-like protein 1 (also found in Chapter 2)	molecular function regulator; RIG-I binding	choline transport; innate immune response; negative regulation of RIG-I signaling pathway
6624	Salinity range	-0.408	7E-39	100.00	*FGD4*	FYVE, RhoGEF and PH domain-containing protein 4	actin binding; metal ion binding; small GTPase binding; Rho guanyl-nucleotide exchange factor activity	regulation of cell shape; regulation of GTPase activity; cytoskeleton organisation; filopodium assembly; positive regulation of apoptotic process
3799	Salinity range	-0.402	7E-39	100.00	*CHFR*	E3 ubiquitin-protein ligase CHFR	metal ion binding; nucleotide binding; ubiquitin protein ligase activity	cell division; mitotic cell cycle checkpoint; protein destabilization; modification-dependent protein catabolic process
7149	Salinity range	-0.400	3E-37	98.86	*DLX6*	Homeobox protein DLX-6	sequence-specific DNA binding; DNA-binding transcription factor activity	cell differentiation; head development; inner ear morphogenesis; roof of mouth development; epithelial cell differentiation; embryonic limb morphogenesis; anatomical structure formation involved in morphogenesis
137	Salinity range	0.445	2E-39	100.00	*DLG2*	Disks large homolog 2	kinase binding; structural constituent of postsynaptic density; ionotropic glutamate receptor binding; guanylate kinase activity	anterograde axonal protein transport; cell-cell adhesion; cellular response to potassium ion; chemical synaptic transmission; establishment or maintenance of epithelial cell apical/basal polarity; sensory perception of pain; retrograde axonal protein transport

3349	SST min	-0.613	2E-39	100.00	*KLF10*	Krueppel-like factor 10	metal ion binding; DNA-binding transcription factor activity; core promoter sequence-specific DNA binding	bone mineralization; cellular response to starvation; circadian rhythm; skeletal system development; cell-cell signaling, cellular response to peptide; positive regulation of osteoclast differentiation; somatic stem cell population maintenance
6455	SST min	-0.583	2E-39	100.00	*CFAP73*	Cilia- and flagella-associated protein 73	dynein complex binding	cilium movement; regulation of microtubule motor activity; inner dynein arm assembly
6447	SST min	-0.534	1E-36	98.85	*CMKLR1*	Chemokine-like receptor 1	chemokine receptor activity; signaling receptor activity; G protein-coupled receptor activity	chemotaxis; immune response; inflammatory response; positive regulation of cold-induced thermogenesis; positive regulation of fat cell differentiation; positive regulation of macrophage chemotaxis; skeletal system development
5547	SST min	-0.533	2E-39	100.00	*NPHP4*	Nephrocystin-4	structural molecular activity	actin cytoskeleton organization; cell-cell adhesion; ciliary basal body-plasma membrane docking; flagellated sperm motility; hippo signaling; photoreceptor cell maintenance; positive regulation of bicellular tight junction assembly; retina development in camera-type eye; signal transduction; visual behaviour
3067	SST min	-0.530	2E-39	100.00	*PSTPIP2*	Proline-serine-threonine phosphatase-interacting protein 2	actin filament binding; cytoskeleton protein binding	cell migration; actin filament polymerization
3772	SST min	-0.479	6E-35	97.70	*GNAQ*	Guanine nucleotide-binding protein G(q) subunit alpha	GTP binding; GTPase activity; metal ion binding; type 2A serotonin receptor binding; G protein-coupled receptor binding	action potential; phototransduction, visible light; blood coagulation; entrainment of circadian clock; protein stabilization; platelet activation; negative regulation of protein kinase activity

4912	SST min	-0.465	9E-38	98.88	*NELL1*	Protein kinase C-binding protein NELL1	calcium ion binding; heparin binding; protein kinase C binding	cell differentiation; negative regulation of osteoblast proliferation; negative regulation of cellular protein catabolic process; positive regulation of bone mineralization; positive regulation of osteoblast differentiation; nervous system development
4754	SST min	-0.427	7E-39	100.00	*CAPSL*	Calcyphosin-like protein	calcium ion binding	None given
7758	SST min	-0.426	2E-39	100.00	*CACNA1A*	Voltage-dependent P/Q-type calcium channel subunit alpha-1A (CACNA1B found in Chapter 2)	metal ion binding; syntaxin binding; voltage-gated calcium channel activity; high voltage-gated calcium channel activity	calcium ion import; calcium ion transmembrane transport; calcium ion transport; cell death; chemical synaptic transmission; membrane depolarization; modulation of chemical synaptic transmission; positive regulation of cytosolic calcium ion concentration; regulation of insulin secretion; regulation of ion transmembrane transport
5365	SST min	-0.425	7E-39	100.00	*GALNT2*	Polypeptide N-acetylgalactosaminyltransferase 2	carbohydrate binding; manganese ion binding	immunoglobulin biosynthetic process; protein O-linked glycosylation; O-glycan processing
336	SST min	-0.417	2E-39	100.00	*ARRB2*	Beta-arrestin-2	angiotensin receptor binding; arrestin family protein binding; enzyme binding; follicle-stimulating hormone receptor binding; molecular adaptor activity; signaling receptor binding; type 2A serotonin receptor binding; platelet activating factor receptor binding	adult walking behaviour; brain development; cell chemotaxis; detection of temperature stimulus involved in sensory perception of pain; dopamine receptor signaling pathway; excitatory postsynaptic potential; follicle-stimulating hormone signaling pathway; membrane organization; negative regulation of smooth muscle cell apoptotic process, platelet activation; positive regulation of calcium ion transport; positive regulation of cardiac muscle cell differentiation; protein transport; negative regulation of toll-like receptor signaling pathway

4885	SST min	-0.412	2E-39	100.00	*BHLHE40*	Class E basic helix-loop-helix protein 40	E-box binding; MRF binding; protein domain specific binding; transcription corepressor activity; DNA-binding transcription factor activity; RNA polymerase II distal enhancer sequencer-specific DNA binding	anterior/posterior pattern specification; cell differentiation; circadian regulation of cell expression; entrainment of circadian clock by photoperiod; negative regulation of DNA-binding transcription factor activity; regulation of circadian rhythm; regulation of neurogenesis; regulation of transcription, DNA-templated
5280	SST min	0.417	2E-39	100.00	*DDX31*	Probable ATP-dependent RNA helicase DDX31	ATP binding; helicase activity; RNA binding	ribosome biogenesis
998	SST min	0.433	4E-36	97.75	*TMEM163*	Transmembrane protein 163	zinc ion binding	zinc ion import into synaptic vesicle
2332	SST min	0.435	2E-39	100.00	*MCM10*	Protein MCM10 homolog	DNA replication origin binding; double-stranded DNA binding; enzyme binding; metal ion binding; single-stranded DNA binding; identical protein binding	cell population proliferation; cellular response to DNA damage stimulus; DNA replication; DNA replication initiation; G1/S transition of mitotic cell cycle
6910	SST min	0.436	2E-39	100.00	*KIAA0556*	Protein KIAA0556	None given	cerebrospinal fluid circulation
7158	SST min	0.442	2E-39	100.00	*LOC101334788*	LOW QUALITY PROTEIN: acyl-protein thioesterase 2-like	hydrolase activity	None given
5329	SST min	0.458	4E-36	97.75	*TEAD1*	Transcriptional enhancer factor TEF-1	DNA binding; DNA-binding transcription factor activity; protein heterodimerization activity; proximal promoter sequence-specific DNA binding; sequence-specific DNA binding; transcription coactivator binding; transcription regulator recruiting activity	hippo signaling; positive regulation of transcription, DNA-templated; positive regulation of transcription by RNA polymerase II, protein-containing complex assembly
7421	SST min	0.489	7E-39	100.00	*PACRG*	Parkin coregulated gene protein	actin binding; alpha-tubulin binding; beta-tubulin binding; chaperone binding; heat shock protein binding; ubiquitin protein ligase binding; G protein-coupled receptor binding, Hsp70 protein binding, Hsp90 protein binding	cellular response to unfolded protein; negative regulation of cell death; spermatid development

3676	SST min	0.508	2E-39	100.00	*TOX*	Thymocyte selection-associated high mobility group box protein TOX	DNA binding	lymph node development; lymphocyte differentiation; Peyer's patch development; positive regulation of natural killer cell differentiation
1502	SST min	0.515	7E-39	100.00	*DHX57*	Putative ATP-dependent RNA helicase DHX57	ATP binding; metal ion binding; RNA binding; ATP-dependent 3'-5' RNA helicase activity	None given
2894	SST min	0.527	1E-36	98.85	*VPS13D*	Vacuolar protein sorting-associated protein 13D	None given	mitochondrion organization; positive regulation of mitophagy; protein retention in Golgi apparatus; protein targeting to vacuole
662	SST min	0.540	2E-39	100.00	*SORCS3*	VPS10 domain-containing receptor SorCS3	neuropeptide receptor activity	learning, memory, neuropeptide signaling pathway, regulation of long-term synaptic depression

D.iii. *Methods*

D.iii.1 Custom BASH scripts for filtering raw ddRAD sequences of southern Australian bottlenose dolphins (SABD)

```
Remove high missing data:
vcftools --vcf TotalRawSNPs.vcf --missing-indv
#create a document with the names of individuals with >20% missing data
nano missing.data.indiv
vcftools --vcf TotalRawSNPs.vcf --remove missing.data.indiv --recode --recode-INFO-all --out FinalInd

Remove loci with general missing data >20% and a minor allele frequency <0.03:
vcftools --vcf FinalInd.recode.vcf --max-missing 0.8 --min-alleles 2 --max-alleles 2 --out FinalInd80 --non-ref-af 0.001 --max-non-ref-af 0.9999 --mac 1 --maf 0.03 --minQ 30 --recode --recode-INFO-all &>VCFtools80.log
#check how many loci left
mawk '!/#/' FinalInd80.recode.vcf | wc -l

Remove complex haplotypes and INDELs:
vcfallelicprimitives FinalInd80.recode.vcf --keep-info --keep-geno > FinalInd80_1.vcf
mawk '!/#/' FinalInd80_1.vcf | wc –l
vcftools --vcf FinalInd80_1.vcf --remove-indels --recode --recode-INFO-all --out FinalInd80_2

Calculating mean coverage per locus:
#first calculate mean coverage per locus
vcftools --vcf FinalInd80_2.recode.vcf --site-depth --out FinalInd80_2
cut -f3 FinalInd80_2.ldepth > AverageDepthSite2
#x= sample size
mawk '!/D/' AverageDepthSite2| mawk -v x=137 '{print $1/x}' > meandepthpersite2
#calculate mean coverage plus 2 SD to use as a threshold
mawk '{ sum += $1; sumsq += ($1)^2; n++ } END { if (n > 0) print sum / n, sqrt((sumsq-sum^2 / n) / n); }' meandepthpersite2
#x = average + 2 SD
vcftools --vcf FinalInd80_2.recode.vcf --recode-INFO-all --out FinalInd80_3 --max-meanDP 85 --recode

Remove loci with allelic balance <0.15 or <0.85:
vcffilter -s -f "AB > 0.15 & AB < 0.85 | AB < 0.01" FinalInd80_3.recode.vcf > FinalInd80_AB.recode.vcf
mawk '!/#/' FinalInd80_AB.recode.vcf | wc –l

Remove loci with mapping quality score <0.8 or <1.2:
vcffilter -f "MQM / MQMR > 0.8 & MQM / MQMR <1.2" FinalInd80_AB.recode.vcf > FinalInd80_AB2.recode.vcf
mawk '!/#/' FinalInd80_AB2.recode.vcf | wc –l
```

Removing any locus that has a quality score below 0.2 (1/5) of the depth:

```
vcffilter -f "QUAL / DP > 0.20" FinalInd80_AB2.recode.vcf > FinalInd80_AB3.recode.vcf
mawk '!/#/' FinalInd80_AB3.recode.vcf | wc -l
```

Removing SNPs with >5% missing data:

```
#(sample size x 0.05=max missing count)
vcftools --vcf FinalInd80_AB3_RR.recode.vcf --max-missing-count 7 --recode --recode-INFO-all --out FinalInd80_AB4
```

Removing loci outside of Hardy-Weinberg Equilibrium:

```
#using a HWE (-h) value of 0.05 and a –c value (being the percentage of populations out of HWE) of the default of 25%
perl ~/Scripts/HWfilteringVCF.pl -v FinalInd80_AB4.recode.vcf -p ./popmap1.txt -h 0.05 -c 0.25 -o FinalInd80_HWE
#using HWfilteringVCF.pl script (custom script)
mawk '!/#/' FinalInd80_HWE.recode.vcf | wc -l
```

Extracting one locus per site:

```
perl ~/Scripts/ExtractBQFRLM_Elly.pl FinalInd80_HWE.recode.vcf BQ.vcf Frst.vcf Rdm.vcf LM.vcf
# using ExtractBQFRLM.pl script (custom script)
```

Removing loci in linkage disequilibrium:

```
vcftools --vcf BQ.vcf --interchrom-geno-r2 --min-r2 0.8
#using the out.interchrom.geno.ld output file extract the list of loci out of LD in R (LD_exclude_list.txt)
vcftools --vcf BQ.vcf --exclude-positions LD_exclude_list.txt --recode --recode-INFO-all --out BQ_LD
```

Turn final SNPs vcf file into fasta file:

```
perl ~/Scripts/ExtSeq.pl ./reference.fasta ./BQ_LD.recode.vcf ./BQ_LD.fasta
```

Aligning to the *T. aduncus* genome:

```
bowtie2 -x ~/Aduncus_seascape/GenomeAssembly/Tadunc_genome -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 -f ./BQ_LD.fasta -S ./Taust_to_adunc_bowtie2_assembly --no-sq
#list of SNPs that did not align
nano Nonaligning_SNPs.txt
#removing SNPs that did not align
awk 'FNR == NR { h[$1]; next }; !($1 in h)' Nonaligning_SNPs ./BQ_LD.recode.vcf > Mapped_SNPs_australis.vcf
mawk '!/#/' Mapped_SNPs_australis.vcf | wc -l
```

E. Supplementary References

Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O 2018, 'Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling', Global Ecology and Biogeography, vol. 27, no. 3, pp. 277-284.

Boyer TP, Antonov JI, Baranova OK, Coleman C, Garcia HE, Grodsky A, Johnson DR, Locarnini RA, Mishonov AV, O'Brien TD (2013) World Ocean Database 2013. National Oceanic and Atmospheric Administration, NOAA Atlas NESDIS.

Feldman G, McClain C (2010) Ocean Color Web, SeaWiFS Reprocessing, NASA Goddard Space Flight Center [Online]. Available: http://oceancolor.gsfc.nasa.gov].

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, De Clerck O 2012, 'Bio-ORACLE: A global environmental dataset for marine species distribution modelling', Global Ecology and Biogeography, vol. 21, no. 2, pp. 272-281.