Molecular ecology and systematics of blue mussels (genus *Mytilus*) (Mytilidae; Bivalvia; Mollusca) in the Southern hemisphere

Kristen Marie Westfall

A thesis
submitted to Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Marine Biology

2011

Preface

The primary focus of this thesis began as exploring the taxonomic status and molecular ecology of blue mussels in New Zealand. Like most investigations of inquiring Ph.D. minds, the research scope quickly advanced into considerations at multiple spatial and temporal scales. The metamorphosis of ideas as new information brought about more questions than answers has shaped this thesis into a suite of complementary investigations expanded to other Southern hemisphere countries.

The geographical focus of this thesis is mainly restricted to three countries in the Southern hemisphere: Chile, New Zealand and Australia with extra sample locations in the Northern hemisphere, Argentina and the Falkland Islands plus mined Genbank data used for phylogenetic purposes. After consideration of the taxonomic status and evolutionary history of Southern hemisphere blue mussels and hybridisation dynamics between native and non-native individuals, the focus shifts to microevolutionary and population level processes within New Zealand. Although this thesis comprises a spectrum of spatial and temporal scales by using samples from multiple countries and DNA regions of variable mutation rates, the unified theories and conclusions provide information on aspects of Southern hemisphere blue mussel systematics and molecular ecology that have never been put to paper.

Chapter 2 is comprised of two peer-reviewed, published articles: (1) 'An RFLP assay to determine if *Mytilus galloprovincialis* (Mytilidae; Bivalvia) is of Southern or Northern hemisphere origin' and (2) 'Genetic diversity of Southern hemisphere blue mussels (Mytilidae; Bivalvia) and the identification of non-indigenous taxa'. Although the published articles are co-authored by Dr. Jonathan P.A. Gardner (1,2) and Dr. Peter H. Wimberger (1), I performed laboratory work, analyses and manuscript preparation, with manuscript editing contributions from the other authors. Dr. Wimberger was also

responsible for the conception of ideas and preliminary investigation, and aided in laboratory work for article (1).

This thesis contains references to anthropogenically-induced transport of mussels on international and domestic scales. Although invoking an explanation of human-mediated transport may be the easy way to explain anomalies, the transport of mussels as hull-foul, in ballast water and sea chests is a well-documented and common phenomenon. For aquaculture, *M. galloprovincialis* is deliberately transported to and perpetuated in locations not part of their natural distribution. *M. galloprovincialis* from the Northern hemisphere has been introduced to many parts of the world: in some cases the invasion was cryptic (i.e. Northeast Pacific Ocean) or it was a newly documented species for a region (i.e. South Africa and Namibia). In most cases the offending species is *M. galloprovincialis*, therefore, it is reasonable to invoke human-mediated introduction to explain irregularities in the data sets with proper evidence to support the hypothesis.

All DNA sequences are deposited in Genbank with accession numbers provided in the 'Materials and Methods' sections of applicable chapters. Where methods are repeated exactly among chapters they are referred to the chapter in which they were first presented.

Abstract

The *Mytilus edulis* species complex, comprised of *M. edulis*, *M. galloprovincialis* and *M. trossulus*, is antitropically distributed in temperate coastal regions of all oceans and main seas of the world. This genus has been heavily studied in the Northern hemisphere but Southern hemisphere populations have received much less attention. This thesis aims to place Southern hemisphere blue mussels into global evolutionary relationships among *Mytilus* species and investigate aspects of their molecular ecology, including, effects of non-native Northern hemisphere species introductions, biogeography across the Southern hemisphere, regional phylogeographic patterns and population genetics within New Zealand.

Southern hemisphere blue mussel phylogenetic reconstruction resulted in the detection of a monophyletic *M. galloprovincialis* lineage. Two new molecular markers developed with specificity for this lineage and congruence among phylogenetic investigations indicates a panhemispheric distribution of this *M. galloprovincialis* lineage with implications for naming a new sibling species of the *M. edulis* complex. This proposed new species, *M. meridianus*, is distributed in South America, the Kerguelen Islands, New Zealand and Australia at latitudes between ~ 30°S and ~ 55°S.

Non-native *M. galloprovincialis* introduced from the Northern hemisphere have been present in NZ, Australia and Chile for at least ten years and hybridise with native blue mussels. Introgression is observed in New Zealand and Australian but not Chilean hybrid regions. The limited number of introduced mussels in Australia induces hybrid swamping of non-native alleles but an interlineage gender bias towards non-native maternal parents may result in eventual loss of the unique genomic content of native blue mussels in NZ.

Southern hemisphere blue mussels form a monophyletic sister clade to a haplogroup shared by Northern hemisphere *M. edulis* and *M. galloprovincialis*. Although single gene histories are not congruent with respect to evolutionary relationships within the Northern hemisphere due to introgressive hybridisation after speciation, it is clear that Southern hemisphere blue mussels arose from a species native to the northeast Atlantic Ocean after speciation of the three '*M. edulis* complex' members.

Within the Southern hemisphere monophyletic clade lies three reciprocally monophyletic clades restricted to the geographic regions South America/Kerguelen Islands, New Zealand and Australia. Phylogeographic analysis indicates past gene flow between South American/Kerguelen Islands and New Zealand populations that has ceased

at present day and ongoing gene flow between South America and the Kerguelen Islands likely via the West Wind Drift.

Within NZ, population subdivision inferred from mtDNA indicates genetic variation is distributed within an east-west phylogeographic split on the North Island. These populations experienced gene flow in the past that has ceased at present day. Microsatellite allele frequencies indicate a different population subdivision within the northwest North Island that has been narrowed down to a 15 km stretch of coastline in a sheltered bay. The abrupt discontinuity within a small geographic area does not conform to classic population subdivision in this broad-cast spawning species, therefore, further investigation into the genomic content of northwest North Island mussels with respect to introgressed non-native genes is warranted.

Resolving complex phylogenetic patterns from interspecific introgression is key to understanding the evolutionary history of Southern hemisphere *M. galloprovincialis*. Further characterisation of hybrid introgression would increase accuracy of (1) inferences of processes contributing to hybridisation dynamics and (2) population subdivision in NZ. Probing the basis for variation of hybridisation dynamics would help to predict the outcomes of Northern hemisphere *M. galloprovincialis* introductions in other areas of the world.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Jonathan Gardner, for providing the opportunity to do this research and being an excellent mentor in all aspects of the doctoral degree. Thank you Dr. Ann Wood who did so much in the lab and provided sound advice on all aspects of the research. Thank you to Dr. Peter Wimberger for being an inspiration with infectious enthusiasm and ambition.

This research was partially funded by Biosecurity New Zealand and I would like to thank Brendan Gould for providing the opportunity to secure this funding.

There were so many people who collected mussels even before this research was started, thank you to everyone: E. Tarifeno, B. Hayward, F. Smith, M. Francis, A. Wood, M. Curwen, P.H. Wimberger, N. Philips, J. Shima, K. Hill, D. Marshall, B. Bell, C. & H. Hay, R. Williamson, C. Bedford, J. Helson, N. Moltschaniwskyj, D. Innes, L. Westfall, W. Ponder and the Australian National Museum.

The School of Biological Sciences at Victoria University of Wellington has many support and administrative staff without whom life would have never run as smoothly as it did. A special thank you to Sandra Taylor.

Thank you to my fellow students and family for providing support and an ear to listen when the bumps on the road got a little too rough.

Table of Contents

Table of Figures

Table of Tables

6

Chapter 1: Literature Review

1.1 Blue mussels of the genus Mytilus

1.1.1 Taxonomy and ecology

Members of the genus *Mytilus* are antitropically distributed in all oceans, main seas and every continent in the world, except Antarctica. Early blue mussel ecological studies spanning the widespread distribution identified disjunct morphological differences due to environmental variables, which confused the taxonomy of the species. This provided an impetus to review the taxonomic status of one of most cosmopolitan marine invertebrate species in the world. The earliest published taxonomic review (Lamy 1936, reviewed in McDonald et al. 1991) used morphological characteristics to distinguish the species *Mytilus edulis* Linnaeus, 1758, *M. galloprovincialis* Lamarck, 1819, *M. trossulus* Gould, 1850, *M. chilensis* Hupe, 1854, *M. platensis* Orbigny, 1846, *M. planulatus* Lamarck, 1819, and also describe *M. desolationis* for the first time. This was the beginning of a rigorous taxonomic review of the genus *Mytilus*, which is still ongoing today mostly in the molecular realm. In 1955, Soot-Ryen (reviewed in McDonald et al. 1991) published a review concluding species described as distinct (morphologically) by Lamy (1936) were merely all sub-species of *M. edulis*.

Several seminal papers in the 1980's identified the first biochemical (genetic) variation mostly among species in the Northern hemisphere and concluded that *M. edulis*, *M. galloprovincialis*, and *M. trossulus* should be considered as sibling species termed the '*Mytilus edulis* species complex' (Gosling 1984, Koehn et al. 1984, McDonald & Koehn 1988, McDonald et al. 1991).   The interest generated from observed patterns of biochemical variation became the basis for studies also incorporating mitochondrial DNA variation (Edwards & Skibinski 1987, Gardner & Skibinski 1991) and linking morphologically derived with genetically derived taxonomy (Beaumont et al. 1989, Grant

& Cherry 1985, McDonald et al. 1991, Martel et al. 2000, Suchanek et al. 1997). Although this is not an exhaustive list, it gives an impression of the early interest and major developmental stages in blue mussel taxonomy.

Natural allopatric populations of blue mussel species contain pure genetic lineages, but introgressive hybridization occurs in areas where two species are sympatrically distributed. Within hybrid zones the geographic distribution of pure and hybrid individuals has been related to clinal variation, differential habitat selection, and species-specific larval pre and post-settlement processes (Gilg & Hilbish 2000, Kijewski et al. 2006, Quesada et al. 1993,1999). These natural (or 'naturalised', if introduced species have been present for an extended period of time) hybrid zones exhibit limited amounts of nuclear genetic introgression, which have been shown to be asymmetrically distributed based on linkage analysis with several marker loci (Rawson et al. 1996).

1.1.2 Phylogeography

Pacific Ocean *M. trossulus* and the ancestral species of *M. galloprovincialis* and *M. edulis* diverged 3.5 million years ago (mya) coinciding with the opening of the Bering Strait (Riginos et al. 2004). The ancestral species migrated across the Atlantic Ocean and populated European coasts where *M. galloprovincialis* is believed to have diverged from modern *M. edulis* in the Mediterranean Sea approximately 1.0-1.5 mya (Rawson and Hilbish 1995).

In the Northern hemisphere, *Mytilus edulis* is currently distributed along North Atlantic coastlines and *Mytilus trossulus* is naturally distributed in the northeast Pacific, northwest Atlantic and the Baltic Sea (McDonald et al. 1991, Seed 1992, Suchanek et al. 1997, Vainola & Hvilsom 1991). Within the sibling species complex, *M. galloprovincialis* has the largest distribution which spans both hemispheres and includes the Mediterranean

Sea, Atlantic coasts of the southwestern British Isles, southern continental Europe, northern and southern African coasts, parts of the northern Pacific, higher latitudes of the southern Pacific and Atlantic Oceans, and even sub Antarctic waters of the Southern Ocean. The success of invasive *M. galloprovincialis* lineages through human-mediated introductions (e.g., Japan, Hong Kong, South Africa, California, Hawaii) is partially responsible for the markedly extended distribution of this species (in comparison to its sibling taxa) (Zuris et al. 2006).

Modern biogeographical interpretations of *M. galloprovincialis* mitochondrial DNA sequence divergence suggest Southern hemisphere lineages arose through an Atlantic route migration during the Pleistocene (Hilbish et al. 2000).   In addition, *M. edulis*-like genes closely related to Northern hemisphere lineages were detected in the study by Hilbish et al. (2000), in mitochondrial DNA from New Zealand, Australia, the Kerguelen Islands and the Falkland Islands and also in a similar study by Daguin & Borsa (1999) which sampled nuclear DNA from northwest African populations of *M. galloprovincialis*. The presence of Northern hemisphere *M. edulis*-lineage genes suggests possible human-mediated recent introductions (of *M. edulis* or *M. edulis*/*M. galloprovincialis* hybrids).

Recently described populations of *M. galloprovincialis* in Chile, previously believed to be *M. chilensis,* inhabit the highest latitude known for this species (Castilla et al. 2005, Toro et al. 2005). In New Zealand, *M. galloprovincialis* is believed to be the most prevalent species based on mitochondrial molecular characterisation (Hilbish et al. 2000) and morphometric analysis of contemporary, midden and fossil valves (Gardner 2004).   Blue mussels are distributed throughout New Zealand from the Bay of Islands to Campbell Island, a geographic span of approx. 1800 km., but are more common at higher latitudes (Gardner 2004). Shell morphometry and allozyme differentiation have suggested that *M. galloprovincialis* inhabits Australian coastlines (McDonald et al. 1991).   Other

Southern hemisphere populations (The Falkland Islands, The Kerguelen Islands and parts of South America) display characteristics intermediate between *M. edulis* and *M. galloprovincialis* (McDonald et al. 1991).

1.1.3 Population genetics

The long-lived (usually 4-7 wk, but can be longer) planktonic larval stage life history characteristic of smooth-shelled blue mussels enables dispersal over large geographic distances and population genetic structuring may not exist at all (i.e. panmixia) or exhibit only slight differentiation. The dispersal distance of larvae may be increased using delayed metamorphosis (Bayne 1965) and post-larval long monofilament threads to increase drag (Sigurdsson et al. 1976). Even though studies have shown that typical dispersal distance rarely exceeds 50 km (Gilg & Hilbish 2003), the larvae can drift for up to 6 months (Lane et al. 1985) and potentially travel much longer distances.

In New Zealand, the endemic greenshell mussel, *Perna canaliculus*, exhibits several models of population subdivision based on allozyme and mtDNA analysis (Apte & Gardner 2002, Gardner et al. 1996, Smith 1988). MtDNA and RFLP analyses show a pronounced North-South split of *P. canaliculus* population subdivision at approximately 42°S with probable barriers to gene flow being strong currents in Cook Strait and intensive upwelling on both coasts of the South Island (Apte & Gardner 2002, Star et al. 2003). The shared life history characteristics between *Mytilus* sp./spp. and *P. canaliculus* (both species have planktonic larvae but may differ in larval dispersal and local recruitment capabilities) suggest that the pattern of population genetic structuring observed for *P. canaliculus* may also be observed in *M. galloprovincialis*.

1.2 Molecular Methodologies

1.2.1 Mitochondrial DNA – analysis and inference

Phylogenetic inferences are based on homologous DNA sequence divergence measured under varying amounts of selection, ranging from neutral to highly selective models (Campbell et al. 1999), but the use of mitochondrial DNA to infer phylogenetic history has been under considerable debate since the practice started. The notion that mitochondrial DNA is inherited in a strictly neutral or nearly neutral equilibrium manner has been seriously doubted (Avise et al. 1987) and a comprehensive review of the literature by Ballard & Kreitman (2005 and references therein) concluded that the general acceptance of mitochondrial neutral evolution has been brought about by common misconceptions of the neutral model of evolution itself and a lack of studies quantifying the fitness of mitochondrial DNA types within a species. MtDNA type and fitness have been correlated for some species with topics such as human male sperm count (Ruiz-Pesini et al. 1998) and fitness of *Drosophila* (Ballard 2004, James & Ballard 2003). Ballard & Rand (2005) give several plausible reasons why mitochondrial DNA is under highly selective evolution including genetic hitchhiking of somewhat disadvantageous genotypes to highly advantageous mtDNA types. Despite these disputes, mtDNA is valuable for inferring phylogenetic topologies within and among species (Avise et al. 1987, Ballard & Rand 2005) and is widely accepted in the scientific community as such. The highly evolving nature of mitochondrial DNA allows phylogenetic inferences on relatively short evolutionary time scales, and can even be applied to measure gene flow between intraspecific populations assuming there is a significant amount of genetic divergence. When applied with spatial scale, microevolutionary processes due to geographical discontinuities can also be inferred in a phylogeographical context.

Members of the family Mytilidae (blue mussels) and Unionidae (fresh-water mussels) exhibit a unique mode of mitochondrial DNA inheritance termed doubly uniparental inheritance (DUI) where two lineages of mitochondrial genomes are present in male mussels, a maternally inherited unit passed on to male and female offspring, and a paternally inherited unit passed on to male progeny (Skibinski et al. 1994, Zouros et al. 1994).   Sequences of the two genomes diverge by 22% but exhibit similar gene arrangement and have a length difference of only 105 base pairs (Mizi et al. 2005). The paternal mitotype is present in small amounts within the male gonadic tissue and has variable inheritance characteristics in comparison with the female mitotype (Cao et al. 2004, Kenchington et al. 2002). The fidelity of DUI is compromised in certain situations where the male mitotype may be completely absent in male gonadic tissue (Saavedra et al. 1997, Zouros et al. 1994), or where the larval regulation of M mitotype inheritance is disrupted resulting from hybridisation events in natural populations and laboratory experiments (Rawson et al. 1996, Wood et al. 2003). In cases where DUI is disrupted, female mitotypes can be recruited into the paternal lineage and become 'masculanised' (Hoeh et al. 1996, 1997, Curole & Kocher 2002).

Despite complications imposed by the disruption of DUI, both lineages of mitotypes have been incorporated into phylogenetic and phylogeographic studies of *Mytilus* spp. to estimate measures of gene flow within and among populations (Quesada et al. 1999, Riginos et al. 2004), infer past migration routes (Hilbish et al. 2000, Riginos et al. 2004), and estimate divergence times within and among species (Hoeh et al. 1996, 1997, Stewart et al. 1996). Although mitotype-specific patterns of inheritance may skew estimates of divergence or gene exchange across large geographic areas (Riginos et al. 2004), these biases offer challenges and opportunities in the general phylogenetic implications of mitochondrial DNA.

The study by Hilbish et al. (2000) used sequence data from the 16s rRNA female mitotype to infer a trans-Atlantic migration route and timing for the origin of the Southern hemisphere lineage of *M. galloprovincialis*. The detected 1.4% consensus sequence divergence corresponds to a migration timing of approximately 1 million years ago (Hilbish et al. 2000). Molecular clock interpretations from divergence estimates differ between male and female mitotypes (Riginos et al. 2004), but the female mitotype maintains higher fidelity in inheritance patterns and is therefore the most reliable mitotype for '*M. edulis* complex' species phylogenetic inferences (Mizi et al. 2005, Riginos et al. 2004).

A more recent phylogenetic investigation using the cytochrome oxidase subunit I gene indicated that blue mussels may have diverged in the Southern hemisphere before speciation of *M. galloprovincialis* and *M. edulis* in the Northern hemisphere (Gérard et al. 2008). Additionally, reciprocally monophyletic clades associated with Southern hemisphere continental distributions (South America/Kerguelen Islands, New Zealand and Australia) were also detected (Gérard et al. 2008).

1.2.2 Microsatellite analysis and inference

Microsatellites are short tandem repeats of DNA usually located at the centromere or telomere regions of chromosomes (Campbell et al. 1999). Sequence polymorphisms are thought to regularly arise through slipped-strand mispairing events during recombination resulting in repeats ranging from 2 to 6 base pairs each (Goldstein & Schlotterer 1999). Microsatellite loci have high rates of mutation leading to large levels of intraspecific variation, but flanking (priming) regions are generally conserved allowing for amplification by polymerase chain reaction (PCR), although this rule is not always observed (Goldstein & Schlotterer 1999). Genetic diversity can be estimated from

microsatellite polymorphisms in alleles and allelic frequency distributions by quantifying rates of gene flow, identifying gene flow barriers, and estimating divergence times between populations. Data from microsatellite loci often show recent separations in population structure because historical separations are masked by back mutation of alleles, meaning mutations are more likely to produce an allele already present in the population than a novel allele.

Seven polymorphic microsatellite loci have been identified for *Mytilus* species by Presa et al. (2002) to assess the amount of genetic partitioning within and among sibling species. Application of these markers to some European *Mytilus* populations has shown that two of the seven loci exhibit excess homozygosity (Presa et al. 2002). The same microsatellites were successfully applied to populations of *M. galloprovincialis* on the northwest coast of Spain but did not detect significant population structure (Diz & Presa 2009).

Excess homozygosity in *Mytilus* has been detected with allozyme markers (Mitton & Koehn 1985), but microsatellite loci for other bivalves tend to show both excess and deficient homozygosity (Bierne et al. 1998, Launey & Hedgecock 2001). Non-Mendelian segregation of alleles has been detected before in bivalve molluscs (Launey & Hedgecock 2001) so *Mytilus* microsatellite loci could potentially exhibit this pattern as well.

1.3 Mytilus sp. bioinvasions

The introduction of non-native commercially viable cultured organisms has had negative ecological impacts in many parts of the world (Carlton 1992) and is especially damaging to biodiversity and normal ecosystem functioning (Carlton & Geller 1993). For the marine environment, the potential for further introduction from transport and discharge of ballast water in large ships confounds the commercial culture impact (Carlton & Geller

1993, Gardner 1996). Together these invasion vectors may lead to severe consequences for native marine biota (Geller 1999).

Characteristics that dictate the relative 'success' of invasive species are the ability to alter ecosystem physical conditions (engineering), to prey on indigenous species and/or to aggressively compete for space (Carlton 1996). These characteristics have been shown to lead to secondary intergradation between allopatric populations (i.e. alien populations that have been introduced) of '*Mytilus edulis* species complex' members shown by clinal variation in mitochondrial DNA (Quesada et al. 1995, 1999).

The best measure of success for invasive species is measuring how successful they have been invading other parts of the world, and *M. galloprovincialis* has proven to be very successful at invasion in the north-east Pacific (Carlton 1992, Geller 1999, McDonald et al. 1991), South Africa (Griffith et al. 1992), Hong Kong (Lee & Morton 1985), Japan (Wilkins et al. 1993) and possibly Hawaii (Apte et al. 2000). For example, in South Africa, invasive populations of *M. galloprovincialis*, competing with natural *Perna perna*, have used an evolutionary strategy of putting the majority of energy into rapid growth and high reproductive output while putting smaller amounts of energy into byssal thread growth (Zardi et al. 2006). These (possibly adaptive) evolutionary advantages have enabled the species to spread over 2000 kilometres in less than 20 yr (Griffiths et al. 1992).

The spread of *M. galloprovincialis* lineages throughout the world is partially responsible for the extended distribution of this species. *M. galloprovincialis* is highly successful at invading new areas and has reached the status of one of the world's 100 most invasive species, in part due to its physiology as mentioned in the South African case history.

Included with the threat of invasion is the potential for new lineages of conspecifics to be introduced and hybridise with existing lineages. For example, the Southern

hemisphere *M. galloprovincialis* have been evolving separately since their natural migration during the Pleistocene (Hilbish et al. 2000), and lineages from Northern hemisphere *M. galloprovincialis* may be introduced for aquaculture or through ship-mediated vectors and displace some natural Southern hemisphere lineages (if adaptive advantages are present).

The spread of *M. galloprovincialis* throughout the world via human-mediated vectors must be addressed firstly with measures to prevent the transport of larvae or adults, and secondly with measures to help stop the spread of populations once they have been introduced into a new area. Accurate identification (reliable morphological and genetic characteristics) is paramount to reducing the spread of potentially invasive sibling species and alien lineages of conspecifics.

1.4 Research objectives

The first objective of this thesis (Chapter 2) is to characterize the taxonomic identity of blue mussels across their natural Southern hemisphere distribution using a previously developed nuclear marker (Me15/16) (Inoue et al., 1995) that discerns among the Northern hemisphere sibling species *M. edulis*, *M. galloprovincialis* and *M. trossulus* by length polymorphism. Two new RFLP's will be developed as a rapid identification tool to distinguish Southern hemisphere *Mytilus* sp. from Northern hemisphere congeners; one RFLP marker based on (mitochondrial DNA) 16s gene sequence polymorphism and one based on the previously mentioned nuclear marker Me15/16. With cytoplasmic and nuclear taxonomic data available from multiple Southern hemisphere locations (Chile, New Zealand and Australia), the distribution and frequency of non-native mussels will be examined with respect to hybridization dynamics and cytonuclear disequilibria (Chapter 3).

Following the initial assessment of taxonomic status across the Southern hemisphere distribution, mitochondrial 16s rRNA and cytochrome oxidase sub-unit III (herein COIII) sequence data will be used to reconstruct the evolutionary history of *Mytilus* across their global distribution for both single and combined gene data under Maximum Likelihood and Bayesian statistical frameworks (Chapter 4). Representative species sequences from the North Atlantic Ocean (all sibling species), North Pacific Ocean (*M. galloprovincialis* and *M. trossulus*), the Mediterranean Sea (*M. galloprovincialis*) and the Black Sea (*M. galloprovincialis*) will form the basis of phylogenetic reconstructions with known evolutionary histories and the topological placement of Southern hemisphere *Mytilus* sp. samples will help to determine their evolutionary relationships with the above mentioned Northern hemisphere species.

Phylogenetic reconstructions will also be used to estimate the number, timing and likely pathway of transequatorial migration events through topological patterning and dating nodes with both fossil calibrated and coalescent models. The phylogenetic reconstructions from the current data set will be compared with previously published topologies to determine support for alternate single gene evolutionary histories. The main objective is to reconcile a scenario of evolutionary history derived from (1) current mitochondrial DNA sequence data, (2) published mitochondrial DNA phylogenetic reconstructions and (3) previous nuclear allozyme and RFLP marker data.

The next research objective is to examine the topological patterning of major Southern hemisphere (phylogenetic) clades within a statistical phylogeographic framework (Chapter 5). Mitochondrial DNA sequence data (COIII from this thesis and cytochrome oxidase subunit I available from Genbank) from *Mytilus* in Chile, the Kerguelen Islands, New Zealand and Australia are used to test for biogeographic boundaries previously identified for intertidal marine invertebrate species with life history characteristics similar

to and different from *Mytilus*. Specific evolutionary processes will be investigated using nested clade phylogeographic analysis with qualitative estimates of gene flow and connectivity on the macrogeographic scale across major Southern hemisphere water bodies and at the microgeographic scale within each region. Furthermore, the timing and pathways of Southern hemisphere colonization will be estimated through phylogenetic reconstruction using the hemispheric divergence estimates from Chapter 4.

The final research objectives are to examine (1) population level genetic structure of blue mussels within New Zealand using multilocus microsatellite allelic variation and (2) correlations among genetic and environmental variation (Chapter 6). Seven polymorphic loci developed for Northern hemisphere *Mytilus* sp. from three independent studies are assayed for 13 New Zealand sample sites distributed from the northern most Bay of Islands region to the southern most offshore Campbell Island. A seascape genetics statistical approach is used to identify correlations between genetic structure and marine environmental variation measured by the National Institute of Water and Atmospheric Research (NIWA). Specific hypotheses of genetic structure and connectivity among populations are tested with Bayesian statistical approaches for determining the number and distribution of discrete populations and the estimation of present day migration rates among populations. Population structure for *Mytilus* within New Zealand is discussed with reference to (1) other intertidal marine invertebrates (i.e. greenshell mussels and Asterinid sea stars) with similar life history characteristics and (2) physical oceanographic features.

1.5 References

Apte S, Gardner JPA (2002) Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology* 11: 1617-1628.

Apte S, Holland BS, Goodwin LS, Gardner JPAG (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecological Systematics* 18: 489-522.

Avise JC (1994) *Molecular Markers, Natural History and Evolution.* Chapman and Hall, New York.

Ballard JWO (2004) Sequential evolution of a symbiont inferred from the host: Wolbachia and *Drosophila simulans*. *Biological Journal of the Linnean Society* 21: 428-442.

Ballard JWO, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker? *Biological Journal of the Linnean Society* 10: 485-488.

Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution and Systematics* 36: 621-642.

Bayne BL (1965) Growth and delay of metamorphosis of the larvae of *Mytilus edulis* (L). Ophelia 2: 1-47.

Bierne N, Launey S, Naciri-Graven Y, Bonhomme F (1998) Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* 148: 1893-1906.

Braby CE, Somero GN (2006) Ecological gradients and relative abundance of native (*Mytilus trossulus*) and invasive (*M. galloprovincialis*) blue mussels in the California hybrid zone. *Marine Biology* 148: 1249-1262.

Campbell NA, Reece JB, Mitchell LG (1999) *Biology 5th Ed.* Benjamin/Cummings, Menlo Park, CA, USA, p. 1572.

Cao L, Kenchington E, Zouros E, Rodakis GC (2004) Evidence that the large noncoding sequence is the main control region of maternally and paternally transmitted mitochondrial genomes of the marine mussel (*Mytilus* spp.). *Genetics* 167: 835-850.

Carlton JT (1992) Introduced marine and estuarine mollusks of North America: an end-of-the-20th-century perspective. *Journal of Shellfish Research* 11: 489-505.

Carlton JT (1996) Pattern, process, and prediction in marine invasion ecology. *Biological Conservation* 78: 97-106.

Carlton JT, Geller JB (1993) Ecological roulette: the transport of non-indigenous marine organisms. *Science* 261: 78-82.

Castilla JC, Uribe M, Bahamonde M, Clarke M, Desqueyroux-Faundez R, Kong I, Moyano H, Rozbaczylo N, Santelices M, Valdovinos C, Zavala P (2005) Down under the southeastern Pacific: marine non-indigenous species in Chile. *Biological Invasions* 7: 213-232.

Coghlan B, Gosling E (2007) Genetic structure of hybrid mussel populations in the west of Ireland: two hypotheses revisited. *Marine Biology* 150: 841-852.

Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper and Row, New York, p. 608.

Curole JP, Kocher TD (2002) Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. *Biological Journal of the Linnean Society* 19: 1323-1328.

Daguin C, Borsa P (1999) Genetic characterization of *Mytilus galloprovincialis* Lmk. In north-west Africa using nuclear DNA markers. *Journal of Experimental Marine Biology and Ecology* 235: 55-65.

Diz A, Presa P (2009) The genetic diversity patter of *Mytilus galloprovincialis* in Galician Rias (NW Iberian estuaries). *Aquaculture* 287: 278-285.

Edwards CA, Skibinski DOF (1987) Genetic variation of mitochondrial DNA in mussel (*Mytilus edulis* and *M. galloprovincialis*) populations from south west England and south Wales. *Marine Biology* 94: 547-556.

Elliott J, Holmes K, Chambers R, Leon K, Wimberger P (2006) Differences in morphology and habitat use among the native mussel *Mytilus trossulus*, the non-native *M. galloprovincialis*, and their hybrids in Puget Sound, Washington. *Marine Biology* 156: 39-53.

Gardner JPA (1996) The *Mytilus edulis* species complex in south west England: effects of hybridization and introgression upon interlocus associations and morphometric variation. *Marine Biology* 125: 385-399.

Gardner JPA (2004) A historical perspective of the genus *Mytilus* (Bivalvia: Mollusca) in New Zealand: multivariate morphometric analysis of fossil, midden and contemporary blue mussels. *Biological Journal of the Linnean Society* 82: 329-344.

Gardner JPA, Skibinski DOF (1991) Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series* 71: 235-243.

Gardner JPA, Pande A, Eyles RF, Wear RG (1996) Biochemical genetic variation among populations of the greenshell mussel, *Perna canaliculus*, from New Zealand: preliminary findings. *Biochemical Systematics and Ecology* 24: 763-774.

Geller JB (1999) Decline of a native mussel masked by sibling species invasion. *Conservation Biology* 13: 661-664.

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

Gilg MR, Hilbish TJ (2000) The relationship between allele frequency and tidal height in a mussel hybrid zone: a test of the differential settlement hypothesis. *Marine Biology* 137: 371-378.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology* 84: 2989-2998.

Goldstein DB, Schlotterer C (Eds.) (1999) *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford, p. 368.

Gosling EM (1984) The systematic status of *Mytilus galloprovincialis* in western Europe: a review. *Malacologia* 25: 551-568.

Grant WS, Cherry MI (1985) *Mytilus galloprovincialis* Lmk. in southern Africa. *Journal of Experimental Marine Biology and Ecology* 90: 179-191.

Griffiths CL, Hockey PAR, van Erkom Shurink C, Le Roux PJ (1992) Marine invasive aliens on South African shores: Implications for community structure and trophic functioning. *South African Journal of Marine Science* 12: 713-722.

Grosholz ED, Ruiz GM (2003) Biological invasions drive size increases in marine and estuarine invertebrates. *Ecological Letters* 6: 700-705.

Heath DD, Rawson PD, Hilbish TJ (1995) PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 2621-2627.

Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration. *Marine Biology* 136: 69-77.

Hoeh WR, Stewart DT, Sutherland GW, Zouros E (1996) Multiple origins of gender-associated mitochondrial DNA lineages in bivalves (Mollusca: Bivalvia). *Evolution* 50: 2276-2286.

Hoeh WR, Stewart DT, Saavedra C, Sutherland BW, Zouros E (1997) Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). *Biological Journal of the Linnean Society* 14: 959-967.

James AC, Ballard JWO (2003) Mitochondrial genotype affects fitness in *Drosophila simulans*. *Genetics* 164: 173-186.

Kenchington E, MacDonald B, Cao L, Tsagkarakis D, Zouros E (2002) Genetics of mother-dependent sex ratio in blue mussels (*Mytilus* spp.) and implications for doubly uniparental inheritance of mitochondrial DNA. *Genetics* 161: 1579-1588.

Kijewski TK, Zbawicka M, Vainola R, Wenne R (2006) Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biology* 149: 1371-1385.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, p. 388.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.

Koehn RK, Hall JG, Innes DJ, Zera AJ (1984) Genetic differentiation of *Mytilus edulis* in eastern North America. *Marine Biology* 79: 117-126.

Lane DJW, Beaumont AR, Hunter JR (1985) Byssus drifting and the drifting threads of the young post-larval mussel *Mytilus edulis*. *Marine Biology* 84: 301-308.

Launey S, Hedgecock D (2001) High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159: 255-265.

Lee SY, Morton B (1985) The introduction of the Mediterranean mussel *Mytilus galloprovincialis* into Hong Kong. *Malacological Review* 18: 107-109.

McDonald JH, Koehn RK (1988) The mussels *Mytilus galloprovincialis* and *M. trossulus* on the Pacific coast of North America. *Marine Biology* 99: 111-118.

McDonald JH, Seed R, Koehn RK (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology* 111: 323-333.

Martel AL, Auffrey LM, Robles CD, Honda BM (2000) Identification of settling and early post-larval stages of mussels (*Mytilus spp.*) from the Pacific coast of North America, using prodissoconch morphology and genomic DNA. *Marine Biology* 137: 811-818.

Mitton JB, Koehn RK (1985) Shell shape variation in the blue mussel, *Mytilus edulis* L., and its association with enzyme heterozygosity. *Journal of Experimental Marine Biology and Ecology* 90: 73-80.

Mizi A, Zouros E, Moschonas N, Rodakis GC (2005) The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: Implications for the doubly uniparental inheritance mode of mtDNA. *Biological Journal of the Linnean Society* 22: 952-967.

Presa P, Perez M, Diz AP (2002) Polymorphic microsatellite markers for blue mussels (*Mytilus* spp.). *Conservation Genetics* 3: 441-443.

Quesada H, Beynon CM, Skibinski DOF (1995) A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* (Lmk.): Pleistocene vicariance biogeography and secondary intergradation. *Biological Journal of the Linnean Society* 12: 521-524.

Quesada H, Wenne R, Skibinski DOF (1999) Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. *Biological Journal of the Linnean Society* 16: 655-665.

Rawson PD, Hilbish TJ (1995) Distribution of male and female mtDNA lineages in populations of blue mussels, *Mytilus trossulus* and *M. galloprovincialis*, along the Pacific coast of North America. *Marine Biology* 124: 245-250.

Rawson PD, Agrawal V, Hilbish TJ (1999) Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Marine Biology* 134: 201-211.

Rawson PD, Secor CL, Hilbish TJ (1996) The effects of natural hybridization on the regulation of doubly uniparental mode of mtDNA inheritance in blue mussels (*Mytilus* spp.). *Genetics* 144: 241-248.

Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* 58: 2438-2451.

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

Ruiz-Pesini E, Diez C, Lapena AC, Perez-Martos A, Montoya J, et al. (1998) Correlation of sperm motility with mitochondrial enzymatic activities. *Clinical Chemistry* 44: 1616-1620.

Saavedra C, Reyero MI, Zouros E (1997) Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus galloprovincialis*. *Genetics* 145: 1073-1082.

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

Sigurdsson JB, Titman CW, Davies PA (1976) The dispersal of young post-larval bivalve molluscs by byssal threads. *Nature* 262: 386-387.

Skibinski DOF, Gallagher C, Beynon CM (1994) Sex limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* 138: 801-809.

Smith PJ (1988) Biochemical-genetic variation in the green-lipped mussel, *Perna canaliculus*, around New Zealand and possible implications for mussel farming. *New Zealand Journal of Marine and Freshwater Research* 22: 85-90.

Star B, Apte S, Gardner JPA (2003) Genetic structuring among populations of the greenshell mussel, *Perna canaliculus,* revealed by randomly amplified polymorphic DNA. *Marine Ecology Progress Series* 249: 171-182.

Stewart DT, Kenchington ER, Singh RK, Zouros E (1996) Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA lineages in the mussel *Mytilus*. *Genetics* 143: 1349-1357.

Suchanek TH, Geller JB, Kreiser BR, Mitton JB (1997) Zoogeographic distributions of the sibling species *Mytilus galloprovincialis* and *M. trossulus* (Bivalvia: Mytilidae) and their hybrids in the North Pacific. *Biological Bulletin* 193: 187-194.

Toro JE (1998) PCR-based nuclear and mtDNA markers and shell morphology as an approach to study the Chilean blue mussel, *Mytilus chilensis* (Bivalvia). *Aquatic Living Resources* 11: 347-353.

Toro JE, Ojeda JA, Vergara AM, Castro GC, Alacapan AC (2005) Molecular characterization of the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) demonstrates evidence for the occurrence of *Mytilus galloprovincialis* in Southern Chile. *Journal of Shellfish Research* 24: 1117-1121.

Vainola R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnean Society*. 43: 127-148.

Wilkins NP, Fujino K, Gosling EM (1983) The Mediterranean mussel *Mytilus galloprovincialis* Lmk. in Japan. *Biological Journal of the Linnean Society* 20: 365-374.

Wood AR, Turner G, Skibinski DOF, Beaumont AR (2003) Disruption of doubly-uniparental inheritance of mitochondrial DNA in hybrid mussels (*Mytilus edulis* x *M. galloprovincialis*). *Heredity* 91: 354-360.

Zardi GI, Nicastro KR, Porri F, McQuaid CD (2006) Sand stress as a non-determinant of habitat segregation of indigenous (*Perna perna*) and invasive (*Mytilus galloprovincialis*) mussels in South Africa. *Marine Biology* 148: 1031-1038.

Zouros E, Ball AO, Saavedra C, Freeman KR (1994) An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences USA* 91: 7463-7467.

Chapter 2 - Part A: An RFLP assay to determine if *Mytilus galloprovincialis* Lmk. (Mytilidae; Bivalvia) is of Northern or Southern hemisphere origin.

2.1A Abstract

Mytilus galloprovincialis is one of three smooth shelled blue mussel species belonging to the *Mytilus edulis* species complex. Naturally occurring and introduced populations of *M. galloprovincialis* are widely distributed throughout many regions of the globe. *M. galloprovincialis* includes morphologically indistinguishable Northern and Southern hemisphere mtDNA lineages that have been separated for ~1 my. To distinguish recently introduced Northern *M. galloprovincialis* from resident Southern *M. galloprovincialis* in New Zealand we developed a 16s rRNA RFLP assay. We compared RFLP assignments of 178 mussels to those generated from a 16s rRNA sequence-estimated phylogeny. All mussels were correctly assigned by the RFLP to their sequence-based phylogenetic placement. This assay allows the rapid identification of Northern and Southern hemisphere *M. galloprovincialis* and will provide an important tool for monitoring human mediated introductions of otherwise cryptic lineages.

2.2A Introduction

Mytilus galloprovincialis is a widely distributed intertidal invertebrate species in many temperate and subtropical regions of the world as the result of natural range expansion and human mediated transport. It is one of three sibling species in the *M. edulis* species complex: *Mytilus edulis* Linne, 1758, *M. galloprovincialis* Lmk, 1819, *M. trossulus* (Gould, 1850) (McDonald et al. 1991). The phylogenetic history of these species suggests that *M. galloprovincialis* diverged from the ancestral *M. edulis* in the Mediterranean Sea approximately 1 to 1.5 mya (Barsotti & Meluzzi 1968; Riginos et al. 2004) and then migrated to the Southern hemisphere via the Atlantic Ocean during the Pleistocene

approximately 0.84 to 1.2 mya (Hilbish et al. 2000, Gérard et al. 2008). *M. galloprovincialis* is now widely established in marine temperate and subtropical waters of both hemispheres as the result of human-mediated introductions (Apte et al. 2000, Wonham 2004).

Rawson & Hilbish (1995) developed an RFLP assay based on a 16s rRNA mtDNA fragment that allows rapid identification of Northern hemisphere lineages within the *M. edulis* species complex. Their assay showed that *M. edulis* and *M. trossulus* have fixed unique haplotypes (A and B, respectively), while *M. galloprovincialis* exhibited roughly equal frequencies of a third unique haplotype (D) and the A haplotype of *M. edulis* (Rawson & Hilbish 1995).

Phylogenetic analysis of the 16s rRNA fragment revealed four distinct lineages (Hilbish et al. 2000). *M. trossulus* is the sister group to the other *Mytilus* taxa (Hilbish et al. 2000). The *M. edulis* mitochondrial lineage includes both *M. edulis* and *M. galloprovincialis* individuals (as identified by nuclear markers - denoted *M. edulis/M. galloprovincialis*). The *M. galloprovincialis* clade is composed of monophyletic Northern and Southern hemisphere sub-clades (Hilbish et al. 2000). The RFLP assay described here distinguishes the Northern and Southern *M. galloprovincialis* sub-clades.

Examination of 16s rRNA DNA sequences (Hilbish et al. 2000) indicated that a triple digest using enzymes *Eco*RV, *Nhe*I and *Spe*I distinguishes between Northern and Southern *M. galloprovincialis* (Figure 2.1). To test the effectiveness of this new assay we performed the triple digest and sequenced 135 mussels of all three taxa from locations around the world.

2.3A Materials and Methods

Total cellular DNA was extracted from posterior adductor muscle tissue (Sokolov 2000). A 527 bp fragment of the 16s rRNA gene was amplified using primers 16sAR/16sBR (Palumbi 1996) in a 25 µL reaction containing 100 ng genomic DNA, 10 pmol each primer, 10 nmol total dNTP, 1X reaction buffer, 1.5 mM MgCl$_2$, and 1U Taq. The amplification conditions were 3 min at 95ºC, 30 cycles at 95ºC for 30 s, 52ºC for 30 s and 72ºC for 45 s, and final extension at 72ºC for 3 min. A 20 µL triple restriction endonuclease digest containing 10U *Eco*RV, 5U *Nhe*I, 5U *Spe*I, 1X NEB Buffer #2, 100 mg BSA and 10 µL PCR product was incubated overnight at 37ºC. PCR and restriction products were scored on 1.5% agarose gels stained with ethidium bromide. Please observe that bands less than 100 bp in length are very faint due to electrophoretic conditions (Figure 2.1), but this does not affect the scoring capabilities as the sizes of the other bands produced are diagnostic for the haplotype groups described here. Nuclear DNA species identities were obtained using the species-specific nuclear-DNA PCR assay Me 15/16 (Inoue et al. 1995).

2.4A Results and Discussion

16s rRNA sequences (Genbank Accession #'s GQ455380 to GQ455405) from RFLP assayed mussels were aligned with ClustalW (Thompson 1994) and a maximum likelihood phylogeny was estimated in PAUP*4.0 (Swofford 2003; model with 500 bootstrap iterations available from K. Westfall). The topology is not reported here, but the major clades corresponded to those reported by Hilbish et al. (2000). Hybrids as determined by Me15/16 were not used in the analysis. Some profiles included a faint band at ~600 bp representing amplification of the male mitotype 16s rRNA gene; this band was ignored for genotyping.
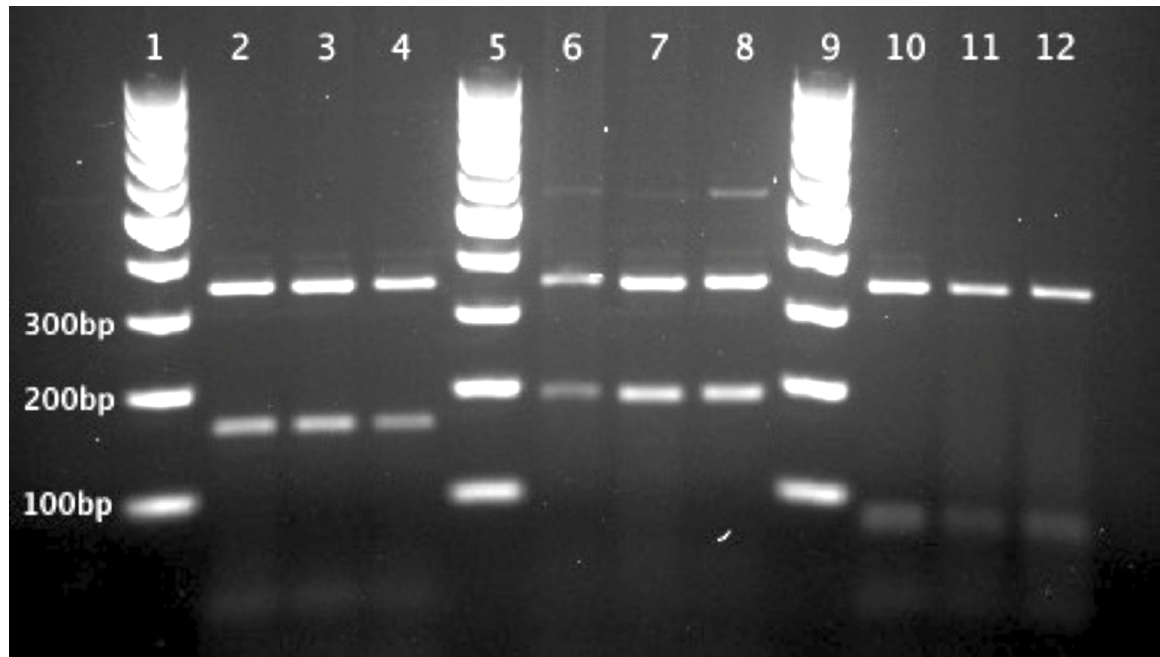
Figure 2.1 - Agarose gel image of restriction fragment profiles.
(1) Southern hemisphere *M. galloprovincialis* (lanes 2, 3, 4) fragments at (342, 167, 28) bp, (2) Northern hemisphere *M. galloprovincialis* (lanes 6, 7, 8) fragments at (342, 195) bp, and (3) Northern hemisphere *M. edulis / M. galloprovincialis* (lanes 10, 11, 12) fragments at (342, 85, 82, 28) bp. Note fragments under 100bp are difficult to visualize, this does not affect scoring. Lanes 1, 5 and 9 are 100bp size ladder. *M. trossulus* profile not visualized, fragments are at (370, 85, 82) bp.

All 178 *M. galloprovincialis* and *M. edulis* RFLP assignments matched their placement in the 16s rRNA sequence-based phylogeny (33 Northern *M. galloprovincialis*, 80 Northern *M. galloprovincialis/M. edulis*, 65 Southern *M. galloprovincialis*). The assay also correctly assigned all 12 *M. trossulus* that were sequenced. Thus, this RFLP assay is a robust and rapid method to distinguish Northern hemisphere *M. galloprovincialis* and *M. edulis* from Southern hemisphere *M. galloprovincialis*. The assay will be useful for monitoring the introduction and possible spread of Northern hemisphere *M. galloprovincialis* in the Southern hemisphere, which can occur through hull fouling, ballast water exchange and aquaculture. Used together with the species-specific nuclear DNA Me15/16 marker, this RFLP assay is a powerful tool to rapidly assess the taxonomic status of members of the *Mytilus edulis* species complex.

2.5A References

Apte S, Holland BS, Godwin LS & Gardner JPA (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Barsotti G, Meluzzi C (1968) Osservazioni su *Mytilus edulis* L. e *Mytilus galloprovincialis* Lamarck (A proposito dei *Mytilus* di Viserbella, Rimini). *Conchiglie* 4: 50-58.

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

Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration. *Marine Biology* 36: 69-77.

Inoue K, Waite JH, Matsouka M, Odo S, Harayama S (1995) Interspecific variations in adhesive protein sequences of *Mytilus edulis*, *M. galloprovincialis*, and *M. trossulus*. *Biological Bulletin* 189: 370-375.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics 2nd edition* (eds. Hillis DM, Morris C, Mabel BK), pp. 205-247. Sinauer Associates, Sunderland, Massachusetts.

Rawson PD, Hilbish TJ (1995) Distribution of male and female mtDNA lineages in populations of blue mussels, *Mytilus trossulus* and *M. galloprovincialis*, along the Pacific coast of North America. *Marine Biology* 124: 245-250.

Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* 58: 2438-2451.

Sokolov EP (2000) An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies* 66: 573-575.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

Wonham MJ (2004) Mini Review: Distribution of the Mediterranean mussel, *Mytilus galloprovincialis* (Bivalvia: Mytilidae) and hybrids in the Northeast Pacific. *Journal of Shellfish Research* 23: 535-543.

Chapter 2.0 Part B: Genetic diversity of Southern hemisphere blue mussels (Bivalvia; Mytilidae) and the identification of non-indigenous taxa.

2.1B Abstract

The taxonomic and evolutionary affinities of Southern hemisphere smooth-shelled blue mussels are unclear, with studies using different marker types having identified different relationships among various geographic regions. Using an existing and a new molecular assay, this study builds on previous work to test the distribution of blue mussels native to and introduced to the Southern hemisphere. Populations of *Mytilus* were sampled from New Zealand, Australia and Chile. The nuclear-DNA marker Me 15/16 was used to identify the taxonomic status of 484 individuals. A new RFLP assay was used to identify the hemisphere of origin for a subset of *M. galloprovincialis*. The Me15/16 marker identified 478 pure *M. galloprovincialis* from Southern hemisphere sites and six *M. edulis/M. galloprovincialis* hybrids from the Auckland Islands (NZ) and Chile. A cytoplasmic RFLP identified Northern hemisphere *M. galloprovincialis* in almost every Southern hemisphere region. The presence of native *M. galloprovincialis* at high latitudes (up to 52°S) has implications for our understanding of environmentally induced selective constraints thought to determine species distributions. Widespread occurrence of invasive Northern hemisphere blue mussels in the Southern hemisphere is documented for the first time. Identification of inter-specific hybrids (*M. edulis* x *M. galloprovincialis*) in Chile and in the Auckland Islands (subantarctic New Zealand) illustrates that environments ranging from international ports to remote protected locations are vulnerable to bioinvasion.

2.2B Introduction

Smooth-shelled blue mussels of the genus *Mytilus* (Mytilidae; Bivalvia) are antitropically distributed in all oceans and main seas of the world (McDonald et al. 1991, Seed & Koehn 1991, Seed 1992, Hilbish et al. 2000). Biochemical and molecular

characterisation divided the morphologically classified *Mytilus edulis* Linnaeus 1758 into a complex of three sibling species termed the '*Mytilus edulis* species complex' (McDonald & Koehn, 1988) - *M. edulis*, *M. trossulus* Gould 1850, *M. galloprovincialis* Lamarck 1819. In the Northern hemisphere, *M. edulis* is distributed in the North Atlantic; *M. galloprovincialis* in the Mediterranean Sea, the British Isles including southwest England, Ireland and northern Scotland, continental Europe, North Africa, and parts of the Northern Pacific; and *M. trossulus* in the northeast Pacific, northwest Atlantic, and the Baltic Sea (McDonald et al. 1991, Väinölä & Hvilsom 1991, Seed 1992, Suchanek et al. 1997, Dias et al. 2009, Doherty et al. 2009). Despite their ecological and economic importance (Gosling 1992, Carranza et al. 2009, Dias et al. 2009), and the biosecurity threat that they pose (Lowe et al. 2000, Lewis et al. 2003, Lee & Chown 2007), *Mytilus* taxa are much less studied in the Southern than the Northern hemisphere.

Several studies addressing the taxonomic status and evolutionary history of *Mytilus* have indicated that the genus may be native to the Southern hemisphere (McDonald et al. 1991, Seed 1992, Sanjuan, et al. 1997, Hilbish et al. 2000, Gardner 2004). This suggestion arises from two separate pieces of evidence: *Mytilus*-like valves of Pleistocene origin have been found in Australia and New Zealand (NZ) dating back to ~1.0 M years ago (mya) (Hilbish et al. 2000, Gardner 2004; and references in both papers), and phylogenetic dating places the origin of Southern hemisphere mussels between 0.84 mya (Gérard et al. 2008) and 1.2 mya (Hilbish et al. 2000). However, until recently (Westfall et al. 2010) there has been no way to differentiate between native Southern and native Northern hemisphere conspecifics.

In South America, *Mytilus* is distributed throughout much of Argentina and Chile, reaching a northern limit at ~30°S. McDonald et al. (1991) described mussels from this region as having allozyme allele characteristics most similar to *M. edulis*, and with

morphological traits intermediate between *M. trossulus* and *M. edulis*. Toro (1998) described mtDNA similarities among *M. edulis*, *M. galloprovincialis* and the putative *M. chilensis* from southern Chile. Although his study differentiated between *M. edulis* and *M. chilensis*, it failed to differentiate between *M. chilensis* and *M. galloprovincialis*. More recent allozyme and molecular approaches have not confirmed the existence of *M. chilensis*, but have classified these mussels as *M. galloprovincialis* (Daguin & Borsa 2000, Toro et al. 2004, Cárcamo et al. 2005).

Kerguelen Islands blue mussels (previously classified as *Mytilus desolationis* Lamy 1936) have greatest affinity to North Atlantic *M. edulis* based on their allozyme alleles (Blot et al. 1988, McDonald et al. 1991). Molecular analyses indicate that they (1) exhibit northern hemisphere affinities for mitochondrial lineage at the 16s rRNA gene, possibly as a result of a recent introduction (Hilbish et al. 2000), (2) possess a nuclear genome of mixed *M. edulis* and *M. galloprovincialis* ancestry (Borsa et al. 2007), and (3) exhibit South American affinity for mitochondrial lineage at the COI gene (Gérard et al. 2008).

In Australia, *Mytilus* is distributed throughout Tasmania and along the southern mainland coast, extending as far north as Port Stephens in New South Wales (~32.4°S) and Perth in Western Australia (~32.2°S) (Wallis 1975, Daguin & Borsa 2000, Gérard et al. 2008). In New Zealand, their distribution includes all subantarctic islands, the Chatham Islands, Stewart Island, the South Island and the North Island, spanning ~35°S to ~52°S latitude (1800 km.). Biochemical profiles of Australasian *Mytilus* suggest a species with intermediate characteristics between *M. edulis* and *M. galloprovincialis*, but with more in common with the latter (McDonald et al. 1991). Mitochondrial 16s and COI sequences have subsequently confirmed the presence of *M. galloprovincialis* in Australia and NZ (Hilbish et al. 2000; Gérard et al. 2008).

M. galloprovincialis is a successful invasive species, having been introduced, deliberately or accidentally, to the Pacific coast of North America, Japan, Hong Kong, Chile, South Africa, probably Western Australia and possibly Hawaii (Wilkins et al. 1983, Grant & Cherry 1985, Lee & Morton 1985, Apte et al. 2000, Daguin & Borsa 2000, Gérard et al. 2008). *M. galloprovincialis* was accidentally introduced in the 1970's to western South Africa (Grant & Cherry 1985) and deliberately transported to southern South Africa for aquaculture (Branch & Steffani 2004). It is now found on the southern and western coastline of South Africa (Robinson et al. 2005) and south Namibia (Branch & Steffani 2004). Blue mussels are absent from the fossil record for southern Africa (Grant & Cherry 1985), indicating that this genus has never naturally existed here. Due to the well-characterised nature of this invasion (Hanekom 2008), and because *Mytilus* is not native to the region, the taxonomic status of South African and Namibian populations will not be considered further.

As a successful invader, *M. galloprovincialis* has an extensive history of displacing native biota and/or of interbreeding with native blue mussels, resulting in hybrids or individuals of mixed ancestry detectable only through molecular methods (Geller et al. 1994, Gardner 1997, Geller 1999, Branch & Steffani 2004, Robinson et al. 2005, Elliott et al. 2008, Hanekom 2008, Brannock et al. 2009, Shinen & Morgan 2009). This species therefore poses both an ecological threat at the community level (Robinson et al. 2005, Shinen & Morgan 2009) and a genetic threat because of its ability to interbreed with native mussels (Brannock et al. 2009, Doherty et al. 2009) and highlights the importance of using molecular methods in marine invasion investigations. Much of this invasive ability arises from the environmental tolerance of the species, its moderately long-lived larval dispersal phase (4-7 weeks) and the fact that it is an important aquaculture species and is actively transplanted by human activity (Seed 1992, Heath et al. 1995, Apte et al. 2000, Branch &

Steffani 2004, Elliott et al. 2008). As a consequence of the ease of its historical and contemporary spread, *M. galloprovincialis* is now recognised worldwide as one of the top 100 invasive threats (Lowe et al. 2000).

Our study builds on existing work to address the question of the taxonomic status of native smooth-shelled blue mussels in the Southern hemisphere. Using a combination of an existing species-specific nuclear DNA marker (Inoue et al. 1995, Rawson et al. 1996) with a species and lineage-specific mitochondrial DNA RFLP marker (Westfall et al. 2010), we determined the taxonomic status of native and non-native blue mussels at locations in the Southern hemisphere including major ports, coastal regions away from the presumptive influence of shipping, and remote offshore islands. Our findings provide information about large spatial scale distributions of native blue mussels and the spatial extent of biological invasions.

2.3B Materials and Methods

2.3.1B Sample collection

Blue mussels were collected from Southern hemisphere sample sites between 1994 and 2009 (Table 2.1), including NZ (Tables 2.2 and 2.3, Figure 2.2a,b), Chile (Table 2.4, Figure 2.2c), and Australia (Table 2.4, Figure 2.2d). Fine scale spatial sampling occurred in NZ (n=19 sample sites) to test for the presence of non-native mussels in ports and regions well removed from the putative influence of shipping, and remote offshore islands (Figure 2.2a,b). Several locations have small sample sizes depending on the abundance of the species at those locations, the ease of collection or the fact that samples were provided by colleagues. In Australia, the Wonboyn River and Wonboyn Lake samples were provided by the Australian National Museum. In New Zealand, some Bay of Islands locations (Matauri Bay, Taupiri Bay and Oakura Bay,) Auckland Harbour's Maungamaungaroa

Estuary and Cape Palliser near Wellington all had naturally occurring very small population sizes such that only a few individuals could be collected.

Table 2.1 – Total individuals assayed per country (and New Zealand separated into regions) for the Me15/16 taxonomic marker and the 16s RFLP assay.

COUNTRY	Me 15/16		RFLP Haplotype 16s		
	Mg	Mg/Me	MgS	MgN	MgN/Me
New Zealand – North Island	138	0	39	26	16
New Zealand – South and Offshore Islands	282	2	49	1	4
Chile	25	4	30	4	5
Australia	33	0	36	3	4
TOTAL	*478*	*6*	*154*	*34*	*29*

Note the abbreviations Mg = *M. galloprovincialis* Lamarck, 1819, MgN = Northern hemisphere *M. galloprovincialis* Lamarck, 1819, MgS = Southern hemisphere *M. galloprovincialis* Lamarck, 1819 and MgN/Me = "North Atlantic" haplotype as defined by Hilbish et al. (2000)

Table 2.2 – The geographical coordinates, codes, sample number (n), genotype (Me 15/16) and 16s RFLP haplotype for New Zealand North Island.

LOCATION	LAT.	LONG.	CODE	n	Genotype Me15/16	RFLP Haplotype 16s
Matauri Bay	35º 01.726' S	173º 55.034' E	MB	2	Mg	-
TeTii Bay	35º 08.768' S	174º 00.258' E	TB	7	Mg	MgN
				19	Mg	MgS
				4	Mg	-
Taupiri Bay	35º 16.427' S	174º 17.421' E	TP	5	Mg	-
Waitangi	35º 16.718' S	174º 05.383' E	WA	10	Mg	MgN
				14	Mg	MgS
				6	Mg	-
Oakura Bay	35º 22.975' S	174º 20.890' E	OA	1	Mg	MgN
Waiheke Island	36º 47.272' S	175º 05.264' E	WI	6	Mg	MgN/Me
				7	Mg	MgN
				2	Mg	MgS
				2	Mg	MgN/Me
				1	Mg	MgS
				8	Mg	MgN/Me
				3	Mg	MgS
				1	Mg	MgN
Maungamau-ngaroa Estuary	37º 02.947' S	174º 54.017' E	ME	4	Mg	-
Wellington Harbour	41º 17.000' S	174º 50.000' E	WH	20	Mg	-
				8	Mg	MgS
				1	Mg	MgN
				1	Mg	MgN/Me
Cape Palliser	41º 36.686' S	175º 17.128' E	CP	4	Mg	-
				2	Mg	-

Note the abbreviations Mg = *M. galloprovincialis* Lamarck, 1819, MgN = Northern hemisphere *M. galloprovincialis* Lamarck, 1819, MgS = Southern hemisphere *M. galloprovincialis* Lamarck, 1819 and MgN/Me = "North Atlantic" haplotype as defined by Hilbish et al. (2000). Lat. = latitude, Long. = longitude.

Table 2.3 - The geographical coordinates, codes, sample number (n), genotype (Me 15/16) and 16s RFLP haplotype for New Zealand South Island and offshore islands.

LOCATION	LAT.	LONG.	CODE	n	Genotype Me15/16	RFLP Haplotype 16s
Whakariki Beach	40º 30.223' S	172º 39.592' E	WB	1	Mg	-
				5	Mg	-
				24	Mg	MgS
Mussel Point	41º 43.989' S	174º 16.052' E	MP	1	Mg	MgN
				4	Mg	MgN/Me
				5	Mg	MgS
				20	Mg	-
Kaikoura	42º 24.260' S	173º 41.103' E	KA	1	Mg	-
				9	Mg	-
Point Kean	42º 25.304' S	173º 39.357' E	PK	24	Mg	-
Greymouth	42º 27.457' S	171º 12.357' E	GR	30	Mg	-
Chatham Islands	43º 35.956' S	176º 39.939' W	CI	24	Mg	-
				6	Mg	MgS
Akaroa	43º 40.323' S	172º 57.914' E	AK	30	Mg	-
Ringitingi Beach	46º 53.259' S	167º 59.907' E	RB	30	Mg	-
Auckland Island	50º 29.391' S	166º 16.767' E	AI	2	Mg/Me	-
				4	Mg	MgS
				28	Mg	-
Campbell Island	52º 31.600' S	169º 06.835' E	CB	24	Mg	-
				6	Mg	MgS

Note the abbreviations Mg = *M. galloprovincialis* Lamarck, 1819, MgN = Northern hemisphere *M. galloprovincialis* Lamarck, 1819, MgS = Southern hemisphere *M. galloprovincialis* Lamarck, 1819 and MgN/Me = "North Atlantic" haplotype as defined by Hilbish et al. (2000). Lat. = latitude, Long. = longitude.

Table 2.4 - The geographical coordinates, codes, sample number (n), genotype (Me 15/16) and 16s RFLP haplotype for Chilean and Australian samples.

LOCATION		LAT.	LONG.	CODE	n	Genotype Me15/16	RFLP Haplotype 16s
Chile	Concepçion	36º 44.003' S	73º 07.830' W	CO	10	Mg	MgS
					4	Mg/Me	MgS
					2	Mg	MgN
					3	Mg	MgN/Me
	Colchogue	37º 03.442' S	73º 10.137' W	CL	7	-	MgS
					9	Mg	MgS
					1	-	MgN
					1	Mg	MgN
					2	-	MgN/Me
Australia	Wonboyn River	37º 15.033' S	149º 57.556' E	WR	1	Mg	MgS
					1	Mg	MgN/Me
	Wonboyn Lake	37º 14.673' S	149º 55.825' E	WL	1	Mg	MgS
	Melbourne	37º 53.532' S	144º 54.910' E	MN	2	Mg	MgN
					2	Mg	MgN/Me
					5	Mg	MgS
					1	Mg	MgS
					10	-	MgS
	Port Arthur	43º 08.489' S	147º 51.397' E	PA	2	Mg	MgS
					2	Mg	MgS
					1	Mg	MgN
					1	Mg	MgN/Me
					14	Mg	MgS

Note the abbreviations Mg = *M. galloprovincialis* Lamarck, 1819, MgN = Northern hemisphere *M. galloprovincialis* Lamarck, 1819, MgS = Southern hemisphere *M. galloprovincialis* Lamarck, 1819 and MgN/Me = "North Atlantic" haplotype as defined by Hilbish et al. (2000). Lat. = latitude, Long. = longitude.
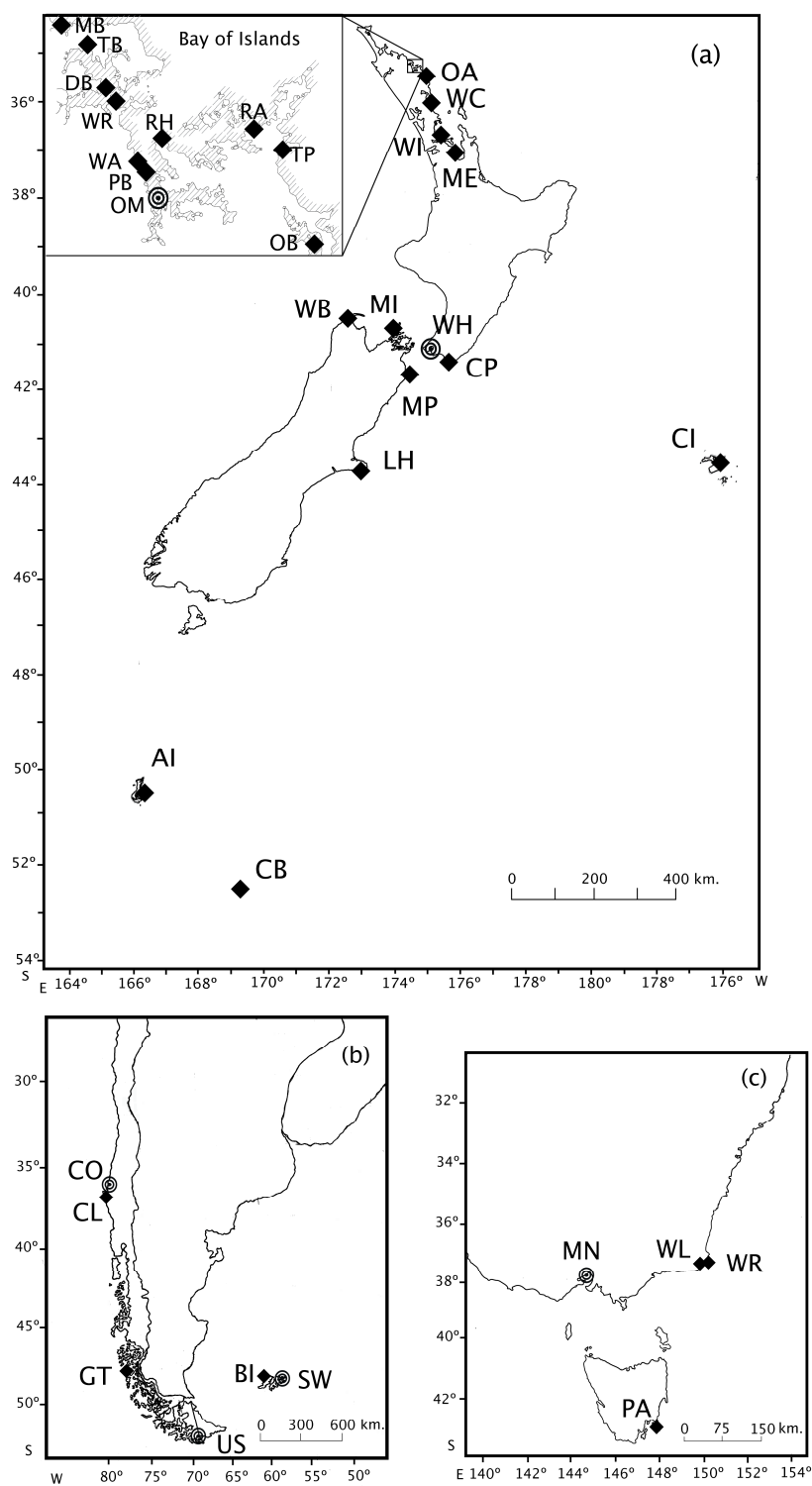
Figure 2.2 - Map of sample locations in current study.
(a) New Zealand (including offshore islands), (b) South America and (c) Australia. The codes for locations are in Tables 2.2 to 2.4. In (a) the inset is a detailed picture of the Bay of Islands. The symbol ◎ denotes current international port sampling sites and the symbol ◆ denotes non-port sampling sites. Scale bars and geographical coordinates are indicated on individual maps.

2.3.2B DNA extraction

Approximately 100 mg of mixed posterior adductor muscle and mantle tissue was dissected and total cellular DNA extracted using a modified (Sokolov 2000) protocol of the DNeasy Extraction Kit (Amersham).

2.3.3B DNA markers

2.3.3.1B Me 15/16

The size of the amplified fragments of the Me 15/16 PCR assay correspond to species-specific taxonomic status (Inoue et al. 1995). PCR conditions were optimised from Inoue et al. (1995) to give: 94°C, 4'; [94°C, 30"; 56°C, 30"; 72°C, 90"]x40; 72°C, 4'; 4°C, ∞ and a $MgCl_2$ concentration of 1.75 mM.

2.3.3.2B 16s RFLP

Westfall et al. (2010) developed an RFLP based on the female mitotype of 16s rRNA gene as determined by 16s rRNA sequences (Hilbish et al. 2000). There are four taxonomic groups identified by this assay: (1) Southern hemisphere *M. galloprovincialis*, (2) Northern hemisphere *M. galloprovincialis,* (3) Northern hemisphere shared *M. edulis/M. galloprovincialis* haplotype, and (4) *M. trossulus*. For samples displaying the shared *M. edulis/M. galloprovincialis* haplotype, the Me15/16 marker was applied to determine species status. To date, the Southern *M. galloprovincialis* haplotype has not been detected in the Northern hemisphere and all Southern mussels have exhibited the same (fixed) mitotype. As far as it is possible to determine based on existing data the assay is reliable and accurate for identifying hemispheric lineages which have been diverging for an estimated 0.84 (Gérard et al. 2008) to 1.2 (Hilbish et al. 2000) million years. This assay was applied to a subset of individuals.

2.4B Results

2.4.1B Me15/16 marker

A total of 484 mussels were genotyped from three Southern hemisphere countries (Table 2.1). Only six individuals (1.2%) exhibited genotypes that were not pure *M. galloprovincialis* based on the Me15/16 marker. For those six individuals, the Me15/16 marker showed two diagnostic bands indicating a *M. galloprovincialis*/*M. edulis* hybrid genotype (Inoue et al. 1995). Two of the hybrids were from the Auckland Islands (Southern Ocean, NZ) and four were from Concepçion (Chile) (Tables 2.3 and 2.4, Figure 2.2b,c, respectively).

2.4.2B 16s RFLP

This assay revealed the extent of Northern hemisphere *M. galloprovincialis* presence in the Southern hemisphere by obtaining haplotypic identity for 217 individuals (Table 2.1). Varying proportions of Northern hemisphere *M. galloprovincialis* genotypes were observed in most countries (Table 2.1).

In the North Island of NZ (Figure 2.2a) Northern hemisphere *M. galloprovincialis* were observed in the northern Bay of Islands region (18 of 51 individuals = 35.3%), at Wellington Harbour (2 of 10 individuals = 20%), and Waiheke Island (24 of 30 individuals = 80%) (Table 2.2). In the South Island of NZ, Northern hemisphere *M. galloprovincialis* were found at Mussel Point, an isolated site in southern Cook Strait. Two NZ offshore island groups (Chatham and Campbell, Figure 2.2b), exhibited only Southern *M. galloprovincialis*, but as noted above the Auckland Islands had two hybrid mussels (*M. galloprovincialis*/*M. edulis*) detected by the Me15/16 marker.

In Australia, the Port Arthur and Melbourne populations had representation of Northern *M. galloprovincialis* at 10% and 20%, respectively (Table 2.4). A single individual from the Wonboyn River also exhibited a Northern hemisphere genotype.

In Chile, the populations at Concepçion and Colchogue exhibited 26.3% and 20% representation of Northern *M. galloprovincialis*, respectively (Table 2.4).

2.5B Discussion

The combined use of the mitochondrial DNA 16s RFLP assay (Westfall et al. 2010) and the nuclear Me15/16 marker has identified native Southern hemisphere *M. galloprovincialis* in all locations sampled, as well as invasive haplotypes from Northern hemisphere congeners, including *M. edulis* as a hybrid genotype (Me15/16 marker) and Northern hemisphere *M. galloprovincialis*.

2.5.1B Taxonomic and genetic diversity of Southern hemisphere blue mussels

The application of a new cytoplasmic (mtDNA) molecular marker (Westfall et al. 2010) has revealed the existence of a native Southern hemisphere lineage of *M. galloprovincialis*, as previously suggested by several authorities (McDonald et al. 1991, Seed 1992, Sanjuan et al. 1997, Hilbish et al. 2000, Gardner 2004). The absence of the Southern hemisphere lineage of *M. galloprovincialis* in the Northern hemisphere (Westfall et al. 2010), and the new information presented here indicates that the Southern hemisphere lineage is the same across sampled countries, spanning the Pacific Ocean and extending into the western part of the South Atlantic Ocean. Sequencing of each 16s fragment to determine within-lineage polymorphism may be informative in terms of the phylogeography of the Southern hemisphere mussels.

The distribution of the Southern hemisphere *M. galloprovincialis* lineage extends from the subtropical (35°S in the Bay of Islands, NZ) to the subpolar (52°S at the Campbell Islands, NZ). This distribution in NZ, which spans a range of sea surface temperatures (SST) from ~5°C to ~25°C (SST data from www.niwa.co.nz/services/sat), is greater than that experienced by *M. galloprovincialis* in other areas, given that this taxon ("the Mediterranean mussel") is a warm temperate species (e.g., McDonald & Koehn 1988, Gardner 1992 and 1994, Seed 1992). This raises interesting questions about the ecophysiology of native Southern hemisphere mussels, when a single lineage is adapted to a wide range of thermal conditions (Gardner 2004). As discussed below, it is also important in the context of interactions (e.g., hybridisation) between the native Southern hemisphere lineage and non-indigenous Northern hemisphere species/lineages in both hemispheres.

2.5.2B Non-indigenous and native species

2.5.2.1B New Zealand

Northern lineage *M. galloprovincialis* were detected mainly in the North Island populations, but also at Mussel Point in the north of the South Island and *M. edulis* alleles were detected in the Auckland Islands (Southern Ocean). The relative prevalence of Northern lineage *M. galloprovincialis* mtDNA on the NZ mainland, especially at international ports (Waiheke Island near Auckland and Wellington Harbour) compared to its almost complete absence on offshore islands, suggests the main mode of introduction is ship-based and via commercial international traffic. Secondary spread within NZ may be via both commercial and personal watercraft traffic.

The subtropical Bay of Islands population (Figure 2.2a) represents the northern-most distribution of *Mytilus* in NZ. It is relatively isolated and represents a disjunctive

distribution for blue mussels in northern NZ (Gardner 2004). Water temperature (SST range of 10-25°C; data from www.niwa.co.nz/services/sat) is considered to be much more suitable for *M. galloprovincialis* than *M. edulis* based on our understanding of (Northern hemisphere) environmental tolerances of these taxa (reviewed by Gardner 1992, Gosling 1992, Seed 1992). In the Bay of Islands the 16s RFLP assay identified a mix of Northern and Southern hemisphere *M. galloprovincialis* (ratio of 6:11), providing evidence of the sympatric occurrence of Northern hemisphere blue mussels with their Southern hemisphere counterparts in northern NZ.

The records of live *M. galloprovincialis* in Maungamaungaroa Estuary and Waiheke Island are new because blue mussels have not previously been reported from this area (Morton & Miller 1968, Inglis et al. 2006), although Fleming (1959) recorded the presence of *Mytilus* spp. fossils at a nearby location, and Gardner (2004) noted the occurrence of blue mussels dating back 200-350 years before present from Motutapu Island, off Auckland. The Waiheke Island mussels are of mixed ancestry and represent the highest percentage of Northern hemisphere *M. galloprovincialis* haplotypes discovered at a single sample location in this study.

The *M. edulis/M. galloprovincialis* hybrids identified from the subantarctic Auckland Islands are rare in the NZ context, but common at this site, with two of 34 Auckland Islands mussels, but only two of a total of 484 individuals in NZ being identified as hybrids. The Auckland Islands group has had several unsuccessful attempts at settlement in the 19th century, has been a base for international whaling activity in the Southern Ocean, and has experienced many vessel visits and also ship wrecks dating back to at least the early 1800's (O'Connor 1999). All of these factors suggest the possibility of accidental introduction of blue mussels via hull fouling to these remote islands. Today, the islands are administered by the NZ Department of Conservation, only a handful of

designated charter companies are allowed access and then under strict biosecurity regulations, and the region was established as a full no-take marine reserve of 484,000 hectares in 2003. The other high latitude island group, Campbell Island at 52°S, exhibited only Southern hemisphere lineage of *M. galloprovincialis*.

The potential for introductions of non-indigenous species to high latitude areas of the Southern Ocean has been investigated by Lewis et al. (2003) and Lee & Chown (2007). These authors cited hull and sea chest fouling assemblages to be of particular concern for this area because of long winter port layover times, slower cruising speeds, and the lack of anti-fouling agents used on ship hulls breaking through sea-ice in the Southern Ocean and polar seas. Lewis et al. (2003) and Lee & Chown (2007) also cite future increases in the number of ships for tourism, fisheries and science as possible vectors for introduction of non-native species to the Southern Ocean and its islands.

2.5.2.2B Australia

Mussels from Melbourne, Wonboyn River and Port Arthur were a mix of Southern and Northern hemisphere *M. galloprovincialis* based on mitochondrial DNA identity. Additionally, at Melbourne and Port Arthur, the 16s RFLP revealed the presence of Northern hemisphere "Mediterranean" and "North Atlantic" *M. galloprovincialis* (Hilbish et al. 2000), reflecting the status of Melbourne as a large and active international port for the last 200 years, and the historical status of Port Arthur as a medium-sized harbour that experienced significant shipping activity until the mid-1900's.. The presence of a non-indigenous haplotype at the Wonboyn River highlights the widespread existence of non-indigenous mussels in Australian waters (e.g., Hewitt 2003). Previously, Borsa et al. (2007) reported high frequencies (>90%) of *M. edulis*-like alleles at the *mac-1* nuclear locus and also high frequencies of *M. galloprovincialis*-like allelomorphs at the nuclear

Glu-3' and *Glu-5'* loci in mussels from six Tasmanian populations. Gérard et al. (2008) reported pronounced differentiation of three Tasmanian mussel populations from other Southern hemisphere populations at the mtDNA COI locus, and a single sample from Nedlands (Western Australia) as clustering with mussels from Dichato (Chile) and Pater Noster Bay (South Africa). Lack of agreement among the three aforementioned studies may be attributable to the situation in Australia with respect to site-specific occurrences of native and introduced taxa, the different powers of resolution of the assays used in all studies to date, preconceived notions of species status (e.g., Borsa et al. 2007), and/or the persistence of ancestral polymorphism due to common ancestry, which is dependent on the evolutionary history of the Southern hemisphere lineage.

2.5.2.3B Chile

At Concepçion and Colchogue the 16s RFLP assay identified a mix of Northern and Southern hemisphere *M. galloprovincialis*, and the Me15/16 marker also identified a small number of hybrid Northern *M. galloprovincialis*/*M. edulis* individuals. The presence of non-indigenous taxa/lineages in Chile is consistent with patterns of shipping activity as Colchogue is a regional fishing port near to Concepçion, which is a major international port. The identification of native Southern hemisphere *M. galloprovincialis* at both sites in Chile is consistent with earlier reports of *M. galloprovincialis* in Chile (Daguin & Borsa 2000, Toro et al. 2004). Most recently, Gérard et al. (2008), using the mtDNA COI gene, identified a "South American" group of mussels (from Patagonia and Maullin, both in Chile) which exhibited close affinity to the Kerguelen Island mussels, but which was distinct from three other groups ("New Zealand", "Tasmanian" and un-named mussels) when the un-named group contained mussels from Nedlands (Western Australia), Dichato (Chile) and Pater Noster Bay (South Africa). Taken together, these separate studies

indicate that blue mussels now found in Chile are a mix of native Southern hemisphere *M. galloprovincialis*, introduced Northern hemisphere *M. galloprovincialis*, introduced *M. edulis*, and their respective hybrids. The aquaculture industry and historical exploitation of natural mussel beds in Chile may have affected the current distribution of native and non-native species as heavy restocking has occurred from Yaldad (island of Chiloe) in the south to northern populations (Carranza et al. 2009). Non-indigenous mussels have not yet established pure wild populations (Castilla & Neill 2009), but the high percentage of non-indigenous haplotypes at Concepçion and Colchogue suggests either leakage from nearby aquaculture farms or port-related introduction.

2.5.3B Conservation implications

The identification of non-native *Mytilus* lineages or taxa at many sites in the Southern hemisphere, including sites well removed from commercial ports, suggests many different points of introduction. The identification of Northern hemisphere *M. galloprovincialis* and also mussels with hybrid-like and/or *M. edulis*-like characteristics supports an interpretation of multiple points of origin from the Northern hemisphere for these invasive species. It is also possible that Southern hemisphere *M. galloprovincialis* are invasive in the Northern hemisphere, and this is clearly an important aspect of marine biosecurity research that also requires attention now that a suitable molecular marker (the 16s RFLP assay) exists to answer this question (Westfall et al. 2010).

Our findings of non-indigenous species of blue mussel taxa in NZ, Australia and Chile reveal that the problem is geographically widespread and more advanced than was previously recognised. An important consideration for blue mussels is not just the introduction of a non-indigenous species which may out-compete and/or replace the native

species (e.g., Geller 1999, Bownes & McQuaid 2006, Elliott et al. 2008), but the fact that wherever two or more taxa co-occur they hybridise (Gardner 1997, Brannock et al. 2009).

Differences in bioengineering capabilities between mussel taxa are not well understood, but can contribute to ecological changes in rocky intertidal habitats. For example, South African research contrasts the well developed three-dimensional mussel bed structure created by introduced *Mytilus* species in areas with high food supply leading to an increase of infaunal species richness and abundance, with the less complex mussel beds of the native *Perna perna* (Robinson et al. 2005). The consequences of lineage or taxon introduction in other Southern hemisphere locations on the hydrodynamics, water quality and habitat alteration in and around mussel beds are yet to be investigated, but elsewhere such introductions have been shown to have significant ecological (e.g. Geller 1999, Elliott et al. 2008) and economic impacts (e.g. Castilla & Neill 2009, Dias et al. 2009).

The frequency and spatial extent of non-native *Mytilus* alleles or taxa in the Southern hemisphere highlight the conservation threat to the genetic integrity of native Southern hemisphere *M. galloprovincialis*. This is perhaps most apparent for remote islands which might be expected to be relatively immune to the introduction of non-native taxa because of their small size and their geographic isolation. However, 200 years of shipping and whaling activity, as well as numerous ship wrecks, and also the modern tourism cruise ship industry have apparently eroded this geographic protection in instances such as the Auckland Islands, although not in others such as the Campbell Islands and the Chatham Islands. Interbreeding between Northern and Southern *M. galloprovincialis* may be limited by the numbers of Northern *M. galloprovincialis* being introduced if that number is very small and/or if there are ecophysiological differences between the lineages which contribute to reduced fitness of the invader (e.g. Gardner 1994 and 1997, Gardner &

Skibinski 1991, Gardner et al. 1993). In regions where two taxa co-occur, hybrid zones or zones of introgression are reported (e.g. Gardner 1997, Elliott et al. 2008, Brannock et al. 2009). Interbreeding may result in disequilibrium between the uniparentally inherited cytoplasmic genome and the biparentally inherited nuclear genome. For example, hybrid zones between *M. galloprovincialis*/*M. edulis* in Europe exhibit an asymmetric mode of mtDNA introgression, with the *M. edulis* genes introgressing into *M. galloprovincialis* but not vice versa (Rawson & Hilbish 1998). Therefore the Northern hemisphere *M. galloprovincialis* that are hybridising with the native Southern lineage may carry genes from previous introgressive hybridisation in their native habitat. At the moment, zones of hybridisation between two populations of pure taxa have not been identified in the Southern hemisphere.

The establishment of a non-indigenous population where blue mussels have not previously existed (e.g., in South Africa) may allow for biological control by complete removal if the introduction is detected in time either fortuitously or by rapid assessment surveys. However, mixed ancestry and hybrid populations elsewhere in the Southern hemisphere mean such measures are not possible. Ongoing protection of Southern hemisphere biota from Northern hemisphere bioinvasions will require a suite of management applications, including rigorous enforcement of hull cleaning and ballast water exchange regulations (e.g. Lewis et al. 2003, Lee & Chown 2007), consideration of vessel path and speed (e.g. Coutts et al. 2010), baseline surveys for rapid identification of non-indigenous species (e.g., Inglis et al. 2006), and the development and application of new management/response strategies (Forrest et al. 2009).

2.6B References

Apte S, Holland BS, Godwin LS, Gardner JPA (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Blot M, Thiriot-Quiévreux C, Soyer J (1988) Genetic relationships among populations of *Mytilus desolationis* from Kerguelen, *M. edulis* from the North Atlantic and *M. galloprovincialis* from the Mediterranean. *Marine Ecology Progress Series* 44: 239-247.

Borsa P, Daguin C, Bierne N (2007) Genomic reticulation indicates mixed ancestry in Southern-hemisphere *Mytilus* spp. mussels. *Biological Journal Of The Linnean Society* 92: 747-754.

Bownes SJ, McQuaid CD (2006). Will the invasive mussel *Mytilus galloprovincialis* Lamarck replace the indigenous *Perna perna* L. on the south coast of South Africa? *Journal of Experimental Marine Biology and Ecology* 338: 140-151.

Brannock PM, Wethey DS, Hilbish TJ (2009) Extensive hybridization with minimal introgression in *Mytilus galloprovincialis* and *M. trossulus* in Hokkaido, Japan. *Marine Ecology Progress Series* 383: 161-171.

Branch GM, Steffani CN (2004) Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology* 300: 189-215.

Carranza A, Defeo O, Beck M, Castilla JC (2009) Linking fisheries management and conservation in bioengineering species: the case of South American mussels. *Reviews in Fish Biology and Fisheries* 19: 349-366.

Cárcamo C, Comesaña AS, Winkler FM, Sanjuan A (2005) Allozyme identification of mussels (Bivalvia: *Mytilus*) on the Pacific coast of South America. *Journal of Shellfish Research* 24: 1101-1115.

Castilla JC, Neill PE (2009) Marine bioinvasions in the southeastern Pacific: status, ecology, economic impacts, conservation and management. Biological Invasions in Marine Ecosystems (ed. by Rilov, G., Crooks, J.A.), pp. 439-457. *Springer Berlin Heidelberg Volume 204*.

Coutts ADM, Piola RF, Hewitt CL, Connell SD, Gardner JPA (2010) Effect of vessel voyage speed on survival of biofouling organisms: implications for translocation of non-indigenous marine species. *Biofouling* 26: 1-13.

Daguin C, Borsa P (2000) Genetic relationships of *Mytilus galloprovincialis* Lmk. populations worldwide: evidence from nuclear-DNA markers. *Bivalve Systematics and Evolution* (ed. by Crame, A., Harper, E., Taylor, J.), pp. 389-397. *Geological Society of London Special Publications volume 177*.

Dias PJ, Bland M, Shanks AM, Beaumont A, Piertney SB, Davies IM, Snow M (2009).

Mytilus species under rope culture in Scotland: implications for management. *Aquaculture International* 17: 437-448.

Doherty SD, Brophy D, Gosling E (2009) Synchronous reproduction may facilitate introgression in a hybrid mussel (*Mytilus*) population. *Journal of Experimental Marine Biology and Ecology* 378: 1-7.

Elliott J, Holmes K, Chambers R, Leon K, Wimberger P (2008) Differences in morphology and habitat use among the native mussel *Mytilus trossulus*, the non-native *M. galloprovincialis*, and their hybrids in Puget Sound, Washington. *Marine Biology* 156: 39-53.

Fleming CA (1959) Notes on New Zealand recent and tertiary mussels (Mytilidae). *Transactions of the Royal Society of New Zealand* 87: 165-178.

Forrest BM, Gardner JPA, Taylor M (2009) Internal borders for the management of invasive marine species. *Journal of Applied Ecology* 46: 46-54.

Gardner JPA (1992) *Mytilus galloprovincialis* (Lamarck) (Bivalvia: Mollusca): the taxonomic status of the Mediterranean mussel. *Ophelia* 35: 219-243.

Gardner JPA (1994) The structure and dynamics of naturally occurring hybrid *Mytilus edulis* Linnaeus, 1758 and *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia, Mollusca) populations: review and interpretation. *Archiv für Hydrobiologie, Monographische Beitrage* 99: 37-71.

Gardner JPA (1996) The *Mytilus edulis* species complex in south-west England: effects of hybridization and introgression upon interlocus associations and morphometric variation. *Marine Biology* 125: 385-399.

Gardner JPA (1997) Hybridisation in the sea. *Advances in Marine Biology* 31: 1-78.

Gardner JPA (2004) A historical perspective of the genus *Mytilus* (Bivalvia: Mollusca) in NZ: multivariate morphometric analysis of fossil, midden and contemporary blue mussels. *Biological Journal of the Linnaean Society* 82: 329-344.

Gardner JPA, Skibinski DOF (1991) Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series* 71: 235-243.

Gardner JPA, Skibinski, DOF, Bajdik CD (1993) Shell growth and viability differences between the marine mussels *Mytilus edulis* (L.), *Mytilus galloprovincialis* (Lmk.) and their hybrids from two sympatric populations in S.W. England. *Biological Bulletin* 185: 405-416.

Geller JB (1999) Decline of a native mussel masked by sibling species invasion. *Conservation Biology* 13: 661-664.

Geller JB, Carlton JT, Power DA (1994) PCR-based detection of mtDNA haplotypes of native and invading mussels on the northeastern Pacific coast: latitudinal pattern of invasion. *Marine Biology* 119: 243-249.

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

Gosling EM (1992) Genetics of *Mytilus*. The Mussel *Mytilus*: Ecology, Physiology, Genetics and Culture (ed. by E.M. Gosling), pp. 309-382. Elsevier, London. *Developments in Aquaculture and Fisheries Science volume 25*.

Grant S, Cherry I (1985) *Mytilus galloprovincialis* Lmk. In Southern Africa. *Journal of Experimental Marine Biology and Ecology* 90: 179-191.

Hanekom N (2008) Invasion of an indigenous *Perna perna* mussel bed on the south coast of South Africa by an alien mussel *Mytilus galloprovincialis* and its effect on the associated fauna. *Biological Invasions* 10: 233-244.

Heath DD, Rawson PD, Hilbish TJ (1995) PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 2621-2627.

Hewitt CL (2003) Marine biosecurity issues in the world oceans: Global activities and Australian directions Ocean Year Book 17: 193-212.

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

Inglis G, Gust N, Fitridge I, Floerl O, Woods C, Hayden B, Fenwick G (2006) Port of Auckland: Baseline survey for non-indigenous marine species (Research Project ZBS 2000/04), pp. 1-91. *Biosecurity New Zealand Technical Paper No: 2005/08*.

Inoue K, Waite J, Matsouka M, Odo S, Harayama S (1995) Interspecific variations in adhesive protein sequences of *Mytilus edulis*, *M. galloprovincialis*, and *M. trossulus*. *Biological Bulletin* 189: 370-375.

Lee JE, Chown SL (2007) *Mytilus* on the move: transport of an invasive bivalve to the Antarctic. *Marine Ecology Progress Series* 339: 307-310.

Lee S, Morton B (1985) The introduction of the Mediterranean mussel *Mytilus galloprovincialis* into Hong Kong. *Malacology Review* 18: 107-109.

Lewis PM, Hewitt CL, Riddle M, McMinn A (2003) Marine introductions in the Southern Ocean: an unrecognized hazard to biodiversity. *Marine Pollution Bulletin* 46: 213-223.

Lowe S, Browne M, Boudjelais S, De Poorter M (2000) 100 of the world's worst invasive alien species: A selection from the global invasive species database. Published by The Invasive Species Specialist Group (ISSG), a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), pp. 12.

McDonald J, Seed R, Koehn R (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology* 111: 323-333.

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

Morton J, Miller M (1968) *The New Zealand Seashore*, pp. 1-638. Collins, London & Auckland.

O'Connor T (1999) *New Zealand's Subantarctic Islands*, pp. 1-104. Reed, Auckland, New Zealand.

Rawson PD, Hilbish TJ (1998) Asymmetric introgression of mitochondrial DNA among European populations of blue mussels (*Mytilus* spp.). *Evolution* 52: 100-108.

Rawson PD, Joyner K, Hilbish TJ (1996) Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity* 77: 599-607.

Robinson TB, Griffiths CL, McQuaid CL, Rius M (2005) Marine alien species of South Africa – status and impacts. *African Journal of Marine Science* 27: 297-306.

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

Santaclara FJ, Espineira M, Cabado AG, Aldasoro A, Gonzalez-Lavin N, Vieites JM (2006) Development of a method for the genetic identification of mussel species belonging to *Mytilus*, *Perna*, *Aulacomya* and other genera. *Journal of Agricultural and Food Chemistry* 54: 8461-8470.

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

Shinen JS, Morgan SG (2009) Mechanisms of invasion resistance: competition among intertidal mussels promotes establishment of invasive species and displacement of native species. *Marine Ecology Progress Series* 383: 187-197.

Sokolov EP (2000) An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies* 66: 573-575.

Suchanek T, Geller J, Kreiser B, Mitton J (1997) Zoogeographic distributions of the sibling species *Mytilus galloprovincialis* and *M. trossulus* (Bivalvia: Mytilidae) and their hybrids in the North Pacific. *Biological Bulletin* 193: 187-194.

Toro JE (1998) PCR-based nuclear and mtDNA markers and shell morphology as an approach to study the taxonomic status of the Chilean blue mussel, *Mytilus chilensis* (Bivalvia). *Aquatic Living Resources* 11: 347-353.

Toro JE, Ojeda JA, Vergara AM, Castro GC, Alcapán AC (2004) Molecular characterization of the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) demonstrates evidence for the occurrence of *Mytilus galloprovincialis* in southern Chile. *Journal of Shellfish Research* 24: 1117-1121.

Väïnölä R, Hvilsom M (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnaean Society* 43: 127-148.

Wallis RL (1975) Thermal tolerance of *Mytilus edulis* of Eastern Australia. *Marine Biology* 30: 183-191.

Westfall KM, Wimberger PH, Gardner JPA (2010) An RFLP mtDNA assay to determine if *Mytilus galloprovincialis* (Mytilidae; Bivalvia) is of Northern or Southern hemisphere origin. *Molecular Ecology Resources* 10: 573-575.

Wilkins N, Fujino K, Gosling E (1983) The Mediterranean mussel *Mytilus galloprovincialis* Lmk. in Japan. *Biological Journal of the Linnaean Society* 20: 365-374.

Chapter 3: Interlineage *Mytilus galloprovincialis* Lmk. 1819 hybridization yields inconsistent genetic outcomes in three Southern hemisphere regions

3.1 Abstract

A unique Southern hemisphere lineage of the blue mussel *Mytilus galloprovincialis* has been diverging allopatrically from Northern hemisphere conspecifics for an estimated 0.84-1.2 million years and human-mediated secondary contact in the Southern hemisphere provides an opportunity to study interlineage hybridization. Cytonuclear disequilibria patterns and zone architecture may reveal regional hybridization dynamics in Chile, New Zealand and Australia between the native Southern hemisphere lineage and the non-native Northern hemisphere lineage of *M. galloprovincialis*. The close taxonomic relationship between hemispheric *M. galloprovincialis* lineages leads to the prediction that introgressive hybridization pervades and threatens the unique genomic content of *M. galloprovincialis* in both hemispheres. A comparative regional approach to hybridization dynamics in the Southern hemisphere is an important step to help predict the genetic compositional effects of future introduction events in other areas of the world. Putatively pure non-native *M. galloprovincialis* and hybrids of native and non-native mussels are present in all studied countries. Introgressive hybridization with native *M. galloprovincialis* is indicated from recombinant cytonuclear genotypes present in New Zealand and Australia. Significant cytonuclear disequilibria exist for endemic homozygotes in all locations except Australia, heterozygotes in Chile and non-native homozygotes in New Zealand. Disequilibria in Australian mussels are non-significant but normalized disequilibria and no detected putatively pure non-native mussels indicate that introductions should be considered as extremely rare events. Continued immigration from one or both parental taxa is the driving factor of disequilibria in New Zealand, whilst directional interlineage crosses with no introgression is the driving factor of disequilibria in Chile. Rare non-native mussel introduction events lead to introgressive swamping of Northern hemisphere alleles in Australia. This is the first account of hybridization, introgression and cytonuclear disequilibrium among native and non-native lineages of *M. galloprovincialis* in either hemisphere. Inter-regional differences of non-native introduction outcomes may be attributed to genetically distinct native-range source populations, local environmental conditions or uncharacterized genetic variation among

native Southern hemisphere populations leading to reproductive incompatibility between lineages. Hybridization dynamics in Chile and Australia do not pose an immediate threat to the perpetuation of Southern lineage genes but the situation is less favourable in New Zealand, where introgression, continued immigration and slight hybridization gender bias towards non-native maternal parents could lead to the eventual regional extirpation of this native lineage.

3.2 Introduction

Differentiation among sibling species *M. galloprovincialis*, *M. edulis* Linn. 1758 and *M. trossulus* Gould 1850, comprising the "*Mytilus edulis* species complex" (McDonald and Koehn, 1988), currently relies on molecular and biochemical methods (McDonald et al., 1991; Seed, 1992). Hybridization readily occurs among sibling species in sympatric areas of distribution resulting in variable patterns of genetic introgression attributed to assortative mating, habitat selection and hybrid unfitness (Rawson et al., 1999; Kijewski et al., 2006).

Cytonuclear disequilibrium, analogous to the measure of linkage disequilibrium in the context of two genes from nuclear and mitochondrial genomes, may be quantified to investigate directional selection, asymmetric migration and assortative fertilization in hybrid zones (Asmussen et al., 1987; Arnold et al., 1988; Asmussen et al., 1989). The cytoplasmic mitochondrial genome is usually inherited uniparentally, even though *Mytilus* spp. exhibit a doubly uniparental mode of mtDNA inheritance (Wood et al., 2003). Well-documented phenomena in *Mytilus* spp. hybrid zones are asymmetric introgression of genes and semi-permeable barriers to gene flow (Rawson and Hilbish, 1998; Rawson et al., 1999; Kijewski et al., 2006).Unique genetic lineages of *M. galloprovincialis* are present in the Northern and Southern hemispheres and have been diverging for approximately 0.84 (Gérard et al., 2008) to 1.2 (Hilbish et al., 2000) million years, although evolutionary relationships among species and lineages are still under investigation. Phylogenetic

investigations have uncovered reciprocally monophyletic alleles of Northern and Southern hemisphere *M. galloprovincialis* (Daguin and Borsa, 2000; Hilbish et al., 2000; Borsa et al., 2007; Gérard et al., 2008), leading to the suggestion that these geographically separated lineages be considered as regional subspecies for the purposes of conservation (Moritz, 1994; Westfall, unpublished data). Moreover, *M. galloprovincialis* (without specific reference to hemisphere, but before the Southern hemisphere was identified as a monophyletic lineage) has been classified as one of the top 100 invasive threats in the world (Lowe et al., 2000) and the Northern hemisphere lineage has now been identified as a non-native species throughout the Southern hemisphere (Westfall et al., 2010 a and b). Non-native *M. galloprovincialis* from the Northern hemisphere have been previously identified in Australia, New Zealand (NZ) and Chile and human-mediated vectors are the probable causes of these introductions (Westfall et al., 2010 a and b). Another population of non-native blue mussels exists in South Africa but will not be considered further here because of the natural absence of native Southern hemisphere blue mussels in this region (Grant and Cherry, 1985).

The introduction of non-native mussels in the Southern hemisphere is a cause of concern for the conservation of the recently identified unique genetic lineage naturally existing there. Close taxonomic affinity between the lineages and historical precedent leads to the prediction of extensive introgression without barriers to gene flow in all hybrid regions . More specifically, cytonuclear disequilibria values are expected to be a signature for the merging of taxa. The following investigation examines the extent of hybridization and introgression between Northern and Southern *M. galloprovincialis* lineages for hybrid regions previously identified in Chile, NZ and Australia (Westfall et al., 2010a). The cytotypes used in this study are derived solely from the 16s female mitotype, which has been shown to have high fidelity during hybridization with *M. edulis* (Wood et al., 2003).

Cytonuclear associations between native and non-native mussels are quantified, specific disequilibria sign patterns and zone architecture are compared to investigate the effects of anthropogenic introduction on the genetic composition of native Southern hemisphere blue mussels. Results are discussed in the context of genetic conservation of the unique Southern hemisphere *M. galloprovincialis* lineage with predicted outcomes for each studied region.

3.3 Materials and Methods

3.3.1 Sampling regime

Three regions in as many Southern hemisphere countries with known non-native mussel presence (Westfall et al., 2010a) were sampled for blue mussels (total of n=190) (Table 3.1). In New Zealand, sample locations ranged from the Bay of Islands (North Island) to the port of Lyttleton Harbour (South Island) (Fig. 1a) (Table 3.1). Two locations were sampled in each of Australia (Fig. 1b) and Chile (Fig. 1c) (Table 3.1).
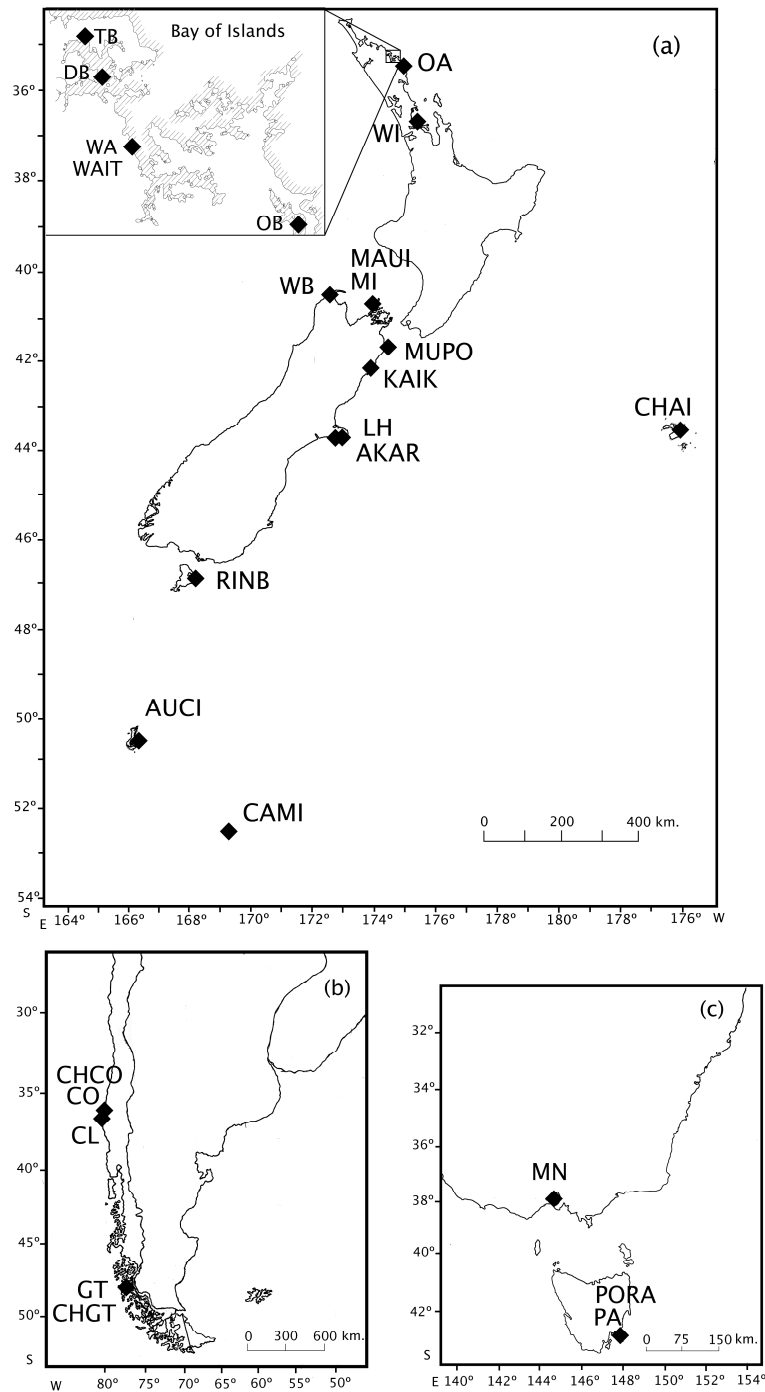
Figure 3.1 - Maps of sample locations in (a) New Zealand, (b) Chile and (c) Australia. In (a) the inset is a detailed picture of the Bay of Islands. Locations are given as 2-letter codes for cytonuclear data (Table 3.2) and 4-letter codes for phylogenetic accuracy data (Table 3.1), note that some locations have both codes provided.

3.3.2 DNA extraction

Total genomic DNA was extracted from mantle and posterior adductor muscle by either the DNeasy Blood and Tissue Extraction kit (QIAGEN) or standard phenol-chloroform protocol (Hillis et al. 1995) in cases where the kit extraction produced poor quality DNA. Both methods were modified by an intermediate salt precipitation step to remove mucopolysaccharides characteristic of molluscan tissues (Sokolov 2000); 50µL saturated KCl solution was added to the Proteinase K digested tissue in lysis buffer, incubated on ice for 10 minutes, centrifuged at 13000rpm for 10 minutes and the aqueous layer transferred to a clean tube for the remaining steps.

3.3.3 Taxonomic status

A mitochondrial 16s RFLP assay (Westfall et al., 2010b) classifies three haplogroups: Southern hemisphere *M. galloprovincialis*, Northern hemisphere *M. galloprovincialis* "Mediterranean" haplogroup (Rawson and Hilbish, 1995; Hilbish et al., 2000) and a shared "North Atlantic" *M. galloprovincialis*/*M. edulis* haplogroup (Rawson and Hilbish, 1995; Hilbish et al., 2000). In the latter group, *M. galloprovincialis* were distinguished from *M. edulis* using the Me15/16 diagnostic marker (Inoue et al., 1995). For further details of assay procedures refer to Westfall et al. (2010b).

The nuclear RFLP assay targets the Me15/16 PCR amplicon (Inoue et al., 1995) and was developed by Santaclara et al. (2006) to differentiate *M. galloprovincialis* and *M. chilensis*. Recent reclassification of *M. chilensis* to *M. galloprovincialis* (Toro et al., 2004) supporting the Southern hemisphere lineage divergence of *M. galloprovincialis* identified by Hilbish et al. (2000) and Gérard et al. (2008) therefore means that the assay actually differentiates between Northern and Southern hemisphere *M. galloprovincialis*. The assay

from Santaclara et al. (2006) was modified as follows: enzymatic digestion in a 25 μL. total reaction volume contained 50 ng DNA, 1U restriction endonuclease AciI, 1X NEB Buffer #3, and 100 mg bovine serum albumin, digested at 37°C for 1 h and heat inactivated at 65°C for 20 min. From Santaclara et al. (2006), the *M. galloprovincialis* Me15/16 amplicon of 126 bp contains a single restriction site resulting in fragments of 69 bp and 57 bp, whereas *M. chilensis* (i.e., Southern hemisphere *M. galloprovincialis*) individuals have a point mutation that causes the cut site for AciI to disappear. In our investigation, two fragments are generated for the Northern *M. galloprovincialis* at 77 bp and 49 bp, while the Southern *M. galloprovincialis* (as identified by the 16s RFLP) Me15/16 amplicon remains uncut at 126 bp. Me 15/16 amplicon sequences of known Southern (n=10) and Northern (n=8) hemisphere *M. galloprovincialis* lineages (based on mtDNA identity and geographic location) were subject to virtual restriction digest to ensure assay integrity with observed and predicted fragment profiles (GenBank accession nos. HQ257459 to HQ257475).

Different fragment sizes obtained in both applications of the same method were investigated. Virtual restriction digests of GenBank sequences submitted by Santaclara et al. (2006) (GenBank accession nos DQ640590 to DQ640610) generate fragments of 53 bp and 49 bp for *M. galloprovincialis* and no cut site for "*M. chilensis*", constituting a total sequence length of 102 bp However, the published sequence length of the Me15/16 amplicon in *M. galloprovincialis* is 126 bp (Inoue et al., 1995) and with the addition of the reverse primer (Inoue et al., 1995), the Me15/16 fragment length matches Inoue et al. (1995) at 126 bp for Southern *M. galloprovincialis*. Virtual restriction digests generate restriction fragments of 77 bp and 49 bp for Northern *M. galloprovincialis*, an exact match to results of our virtual (from sequenced individuals) and laboratory digests. Both assays have fixed alleles for hemispheric lineages of origin.

3.3.4 Phylogenetic accuracy of markers

To ensure the polymorphism within the cytoplasmic and nuclear identified by the RFLP markers accurately reflects the taxonomic status of hemispheric *M. galloprovincialis* lineages, subsets of individuals were sequenced for all taxonomic variants. Taxonomic differentiation is recognised between *Mytilus* sp. taxa in Northern and Southern hemispheres through mitochondrial and nuclear sequence variation (Kenchington et al., 1995; Hilbish et al., 2000; Gerard et al., 2008), allozymes (McDonald et al., 1991; Sanjuan et al., 1997) and shell morphometric variation (McDonald et al., 1991). More recently, the endemic *Mytilus* taxon distributed throughout South America and Australasia has been recognised as a genetically unique lineage of *M. galloprovincialis* (Westfall et al., 2010; Westfall & Gardner, 2010). In order to infer a high degree of confidence in the nuclear and cytoplasmic markers to differentiate between hemispheric taxa, phylogenies were reconstructed for subsets of mussels whose initial taxonomic designation was assigned from geographic placement (Figure 3.1 and Table 3.1). In the Southern hemisphere, mussels were sampled from regions well-removed from hybrid zones (Figure 3.1 and Table 3.1). The goal of phylogenetic reconstructions is to establish taxonomic differentiation between Northern and Southern hemisphere individuals that is captured by both nuclear and cytoplasmic RFLP assays.

Table 3.1 – Collection information and Genbank Accession numbers for 16s and Me15/16 sequences used for phylogenetic analyses.

Data Set	Country	Site	Code	n	Tax ID	Latitude	Longitude	Genbank Acc. #
16s	NZ	Ringitingi B.	RINB	4	MgS	46º 53.259' S	167º 59.907' E	GQ455387
↓	NZ	Chatham Is.	CHAI	2	MgS	43º 35.956' S	176º 39.939' W	GQ455383
	NZ	Auckland Is.	AUCI	1	MgS	50º 29.391' S	166º 16.767' E	GQ455395
	NZ	Lyttleton H.	LYTH	2	MgS	43º 37.152' S	172º 43.008' E	GQ455386
	NZ	Maud Is.	MAUI	1	MgS	41º 01.505' S	173º 53.764' E	GQ455382
	NZ	Maud Is.	MAUI	1	MgS	41º 01.505' S	173º 53.764' E	GQ455385
	Chile	Golfo Trinidad	CHGT	3	MgS	49º 58.557' S	075º 12.085' W	GQ455394
	Chile	Concepcion	CHCO	1	MgS	36º 44.003' S	073º 07.830' W	GQ455388
	Russia	Black Sea	n/a	2	MgN	44º 31.668' N	037º 42.300' W	GQ455398
	England	Mundesley	n/a	7	Mg/Me	52º 52.776' N	001º 26.208' W	GQ455399
	England	Lowestoft	n/a	3	Mg/Me	52º 28.404' N	001º 45.546' W	GQ455405
	Canada	Chemainus	n/a	1	Mg/Me	48º 55.560' N	123º 43.134' E	GQ455397
	Canada	Comox	n/a	1	Mt	49º 40.332' N	124º 56.568' E	GQ455400
	Canada	Comox	n/a	1	Mt	49º 40.332' N	124º 56.568' E	GQ455401
	Canada	Comox	n/a	1	Mt	49º 40.332' N	124º 56.568' E	GQ455403
Me15/16	NZ	Akaroa	AKAR	1	MgS	43º 40.323' S	172º 57.914' E	HQ275459
↓	NZ	Mussel Point	MUPO	1	MgS	41º 43.989' S	174º 16.052' E	HQ275462
	NZ	Waitangi	WAIT	1	MgS	35º 16.718' S	174º 05.383' E	HQ275465
	NZ	Waitangi	WAIT	1	MgS	35º 16.718' S	174º 05.383' E	HQ275466
	NZ	Kaikoura	KAIK	1	MgS	42º 24.260' S	173º 41.103' E	HQ275468
	NZ	Campbell Is.	CAMI	1	MgS	52º 31.600' S	169º 06.835' E	HQ275467
	Chile	Golfo Trinidad	CHGT	1	MgS	49º 58.557' S	075º 12.085' W	HQ275460
	Chile	Concepcion	CHCO	1	MgS	36º 44.003' S	073º 07.830' W	HQ275461
	Australia	Port Arthur	PORA	1	MgS	43º 08.489' S	147º 51.397' E	HQ275463
	Australia	Port Arthur	PORA	1	MgS	43º 08.489' S	147º 51.397' E	HQ275464
	NZ	Kaikoura	KAIK	1	MgN	42º 24.260' S	173º 41.103' E	HQ275468
	Australia	Port Arthur	PORA	1	MgN	43º 08.489' S	147º 51.397' E	HQ275472
	Chile	Concepcion	CHCO	1	MgN	36º 44.003' S	073º 07.830' W	HQ275469
	Turkey	Izmir	n/a	1	MgN	38º 25.242' N	027º 07.656' W	HQ257470
	Turkey	Izmir	n/a	1	MgN	38º 25.242' N	027º 07.656' W	HQ257473
	Turkey	Izmir	n/a	1	MgN	38º 25.242' N	027º 07.656' W	HQ257474
	Russia	Black Sea	n/a	1	MgN	44º 31.668' N	037º 42.300' W	HQ257475
	Russia	Black Sea	n/a	1	MgN	44º 31.668' N	037º 42.300' W	HQ275476

The following taxonomic codes apply: MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern, Mt = *M. trossulus*. Site codes include B. = Beach, Is. = Island and H. = Harbour. Site codes can be found in Figure 3.1, n/a indicates there is no map shown. Country code NZ = New Zealand.

3.3.4.1 16s RFLP

A 527 bp fragment of the 16s rRNA gene was amplified by PCR using the universal primers 16sAR/16s BR (Palumbi 1996) under the following cycling conditions: 95ºC, 3'; [95ºC, 30"; 52ºC, 30"; 72ºC, 45"] x 30; 72ºC, 3'. The mantle and adductor muscle tissue types used for DNA extraction contain both male and female mitotypes

involved in the doubly uniparental mode of mitochondrial DNA inheritance (Stewart et al. 1995). Because this primer pair amplifies both male and female mitotypes, pair-wise uncorrected p-distances between individual sequences within each species/haplogroup were analyzed to identify male genes at an average of 10-20% difference. Female genes may have been "highjacked" into the male mitochondrial genome and subject to male heritability patterns whilst appearing to be female with less than 10% genetic distance compared to other female genes. Identification of high-jacked genes was not a priority as the male mitochondrial genome undergoes relaxed selection and has a significantly higher rate of mutation due to elevated meiotic events during spermatogenesis, therefore, less than 10% genetic distance indicates the gene was very recently high-jacked and resulting phylogenetic constructions would be minimally affected. After sequencing, a reliable 420 bp fragment of the 16s rRNA gene was used for phylogenetic reconstruction. An unrooted maximum parsimony tree was generated from 16s rRNA gene *M. galloprovincialis* sequences in MEGA v4.0.2 (Kimura et al., 1995) and underwent 500 bootstrap iterations.

3.3.4.2 Me15/16 RFLP

Northern (n=11) and Southern (n=10) hemisphere sequences described by Santaclara et al. (1996) (GenBank Accession nos DQ640590 to DQ640610) were aligned with a reference *M. galloprovincialis* genome sequence (Genbank Accession no. D63778). Additional Northern (n=8) and Southern (n=10) hemisphere *M. galloprovincialis* (as defined by geographic placement) individuals were directly sequenced from Me15/16 PCR products (Inoue et al., 1995) and simply aligned with Genbank sequences. An unrooted maximum parsimony tree was generated from Genbank and newly obtained polyphenolic adhesive foot protein gene *M. galloprovincialis* sequences in MEGA v4.0.2 (Kimura et al., 1995) and underwent 500 bootstrap iterations.

3.3.5 Cytonuclear disequilibria statistics

The program CNDd (Asmussen et al., 1987; Asmussen and Basten, 1994) was used to test for cytonuclear disequilibrium (D) and normalized disequilibrium (D' – within minimal and maximal marginal frequencies specific to each region) for genotypic and allelic combinations in samples across the Southern hemisphere grouped by country. The program requires coding the resulting 16s RFLP's three haplogroups into two (Southern and Northern lineages) while conserving hemispheric lineage identification in order to maintain the framework of a 2-locus 2-allele system in the 2 X 3 contingency table. Nuclear alleles were coded as *A* (Southern lineage) and *a* (Northern lineage) and mitochondrial alleles as *M* and *m*, respectively. The composite genotype and haplotype will be referred to as the cytonuclear genotype, individuals typed as Southern lineage will be *AA/M* and Northern lineage *aa/m*. Genotypic disequilibria are defined as the difference between the cytonuclear frequency and the product of genotypic and allelic frequencies. Equations used are described in (Asmussen et al., 1987) and the notation D, D_1, D_2 and D_3 will be used henceforth in this study.

$$D_1 = \text{freq. } (AA/M) - \text{freq. } (AA) \text{ freq. } (M)$$

$$D_2 = \text{freq. } (Aa/M) - \text{freq. } (Aa) \text{ freq. } (M)$$

$$D_3 = \text{freq. } (aa/M) - \text{freq. } (aa) \text{ freq. } (M)$$

$$D = \text{freq. } (A/M) - \text{freq. } (A) \text{ freq. } (M)$$

where freq. means frequency.

Fisher's exact tests (Basten and Asmussen, 1997) firstly quantify associations of nuclear genotypes with cytotypes throughout the whole data set and secondly test for associations among specific nuclear genotype and cytotype combinations. Allelic and genotypic frequencies were also calculated.

The contingency method is used to generate expected values of genotypic and allelic combinations without assuming linkage disequilibrium between the markers. Chi-square tests were performed for single cells (each cytonuclear genotypic combination) and Fisher's exact tests were performed within each sample country across all cytonuclear combinations.

3.4 Results

3.4.1 Taxonomic status

A total of 190 individuals typed by two RFLP assays (Table 3.2) into three genotypes (*AA, Aa, aa*) and two haplotypes (*M* and *m*) produced six unique genotypic and haplotypic combinations: all alleles were fixed for each taxon (Table 3.2). The nuclear Northern *M. galloprovincialis* allele frequency ranged from 8.4% in NZ to 34.5% in Chile (Table 3.4). The cytoplasmic Northern allele frequency ranged from 20% in Australia to 42.7% in NZ (Table 3.4). Examination of the Australia region indicated a pattern of low nuclear (11.7%) and mitochondrial (20%) Northern lineage *M. galloprovincialis* allele frequencies (Table 3.4), whereas the same pattern (frequencies of both alleles at one end of the spectrum) was not observed in other regions. Evidence for hybridization was observed in all country-specific hybrid regions. The existence of recombinant *Aa* genotypes with either maternal haplotype (*M* or *m*) indicated at least a single initial hybridization event with parental Northern or Southern identity. Backcrossing was identified in NZ and Australia with the identification of recombinant homozygous Southern genotypes with a Northern haplotype and vice versa. Although there were no individuals identified in Chile that fit the criterion for introgression described above, there is a caveat that genotypes with "pure" taxon identification have a probability of being an F2 hybrid. A cross of two F1 hybrids with maternal Southern parents would produce 25% progeny with the putatively

pure cytonuclear genotype if alleles segregate in a Mendelian fashion. It will be assumed with this caveat that no introgressed individuals have been identified in Chile, although further research using one or more additional nuclear loci would provide greater resolution of introgression in hybrids.

Table 3.2 - Sampling and taxonomic status listed by country and collection site.

COUNTRY	SITE	LAT.	LONG.	CODE	n	Me15/16 Genotype	16s Haplotype
Chile	Concepcion	36º 44.003' S	73º 07.830' W	CO	14	*AA*	*M*
					5	*aa*	*m*
	Colchogue	37º 03.442' S	73º 10.137' W	CL	9	*Aa*	*M*
					1	*Aa*	*m*
New Zealand	TeTii Bay	35º 08.768' S	174º 00.258' E	TB	7	*AA*	*m*
					19	*AA*	*M*
	Dove's Bay	35º 11.789' S	174º 01.860' E	DB	1	*AA*	*m*
					3	*AA*	*M*
	Waitangi	35º 16.718' S	174º 05.383' E	WA	10	*AA*	*m*
					14	*AA*	*M*
	Okura Bay Road, (Totara North)	35º 20.945' S	174º 22.523' E	OB	3	*AA*	*m*
					2	*AA*	*M*
	Oakura Bay	35º 22.975' S	174º 20.890' E	OA	1	*AA*	*m*
	Waiheke Is.	36º 47.272' S	175º 05.264' E	WI	13	*AA*	*m*
					2	*AA*	*M*
					2	*aa*	*m*
					1	*aa*	*M*
					9	*Aa*	*m*
					3	*Aa*	*M*
	Whakariki Beach	40º 30.223' S	172º 39.592' E	WB	24	*AA*	*M*
	Maud Is.	41º 01.505' S	173º 53.764' E	MI	3	*Aa*	*m*
					1	*Aa*	*M*
					1	*AA*	*m*
					1	*AA*	*M*
	Mussel Point	41º 43.989' S	174º 16.052' E	MP	5	*AA*	*m*
					5	*AA*	*M*
	Lyttleton H.	43º 37.150' S	172º 43.008' E	LH	1	*AA*	*m*
Australia	Melbourne	37º 53.532' S	144º 54.910' E	MN	4	*AA*	*m*
					5	*AA*	*M*
					1	*Aa*	*M*
	Port Arthur	43º 08.489' S	147º 51.397' E	PA	2	*Aa*	*M*
					2	*aa*	*M*
					2	*AA*	*m*
					14	*AA*	*M*

Collection site codes, latitude and longitude for New Zealand, Chile and Australia listed in North to South direction for each country. Cytonuclear genotype information for each collection site. Note the following abbreviations: *A/M* – Southern hemisphere *M. galloprovincialis* (Lamark, 1819), *a/m* – Northern hemisphere *M. galloprovincialis* (Lamark, 1819). Is. = Island and H. = Harbour.

Table 3.3 - Observed and expected numbers of individual genotypes and haplotypes.

Genotype (Me15/16)	Haplotype (16s)	New Zealand†	Chile†	Australia	TOTAL
AA	m	42 (47.9)	0 (2.9)	6 (5.0)	48
AA	M	70 (64.1)	14 (11.0)	19 (20.0)	103
Aa	m	12 (6.8)	1 (2.1)	0 (0.6)	13
Aa	M	4 (9.2)	9 (7.9)	3 (2.4)	16
aa	m	2 (1.3)	5 (1.0)*	0 (2.4)	7
aa	M	1 (1.7)	0 (4.0)	2 (1.6)	3
TOTAL		131	29	30	190

Genotypic and haplotypic observed and expected (parentheses) numbers of individuals for sampled areas. *M. galloprovincialis* genotype and haplotype codes are *AA/M* for Southern lineage and *aa/m* for Northern lineage, respectively. Mitotypes were doubled for allelic/genotypic counts to comprise homozygous diploid genotypes (Basten & Asmussen, 1997). Single cell significance (χ^2) indicated by * and column significance (Fisher's Exact test) indicated by † on column headers. Bonferroni correction column-wide for both significance estimates $p < 0.008$.

Table 3.4 - Genotypic and allelic frequencies for regional (grouped by country) samples.

Sample	Cytotype	TOTAL	Genotype AA	Aa	aa	Nuclear Allele A	a
New	M	0.573	0.534	0.031	0.00	0.550	0.023
Zealand	m	0.427	0.321	0.092	0.015	0.366	0.061
	TOTAL		0.855	0.122	0.023	0.916(p)	0.084(p)
Chile	M	0.793	0.483	0.310	0.000	0.638	0.155
	m	0.207	0.000	0.034	0.172	0.017	0.190
	TOTAL		0.483	0.345	0.172	0.655(p)	0.34(p)
Australia	M	0.800	0.633	0.100	0.067	0.683	0.117
	m	0.200	0.200	0.000	0.000	0.200	0.000
	TOTAL		0.833	0.100	0.067	0.883(p)	0.117(p)

3.4.2 Phylogenetic accuracy of markers

3.4.2.1 16s

There were 16 unique haplotypes identified from 31 mussels distributed across their global distribution, eight unique haplotypes discovered in 15 Southern hemisphere mussels and eight unique haplotypes discovered in 16 Northern hemisphere mussels. The

four haplogroups identified by the 16s RFLP marker can be identified in monophyletic clades of the neighbour-joining phylogenetic tree (Figure 3.2). Bootstrap support values are high for separation of the shared *M. edulis/M. galloprovincialis* North Atlantic haplogroup and above 50% for the separation between Mediterranean and Southern hemisphere *M. galloprovincialis*.

○ – *M. trossulus*
● – *M. galloprovincialis* and *M. edulis* 'North Atlantic'
◆ – *M. galloprovincialis* 'Mediterranean'
■ – *M. galloprovincialis* 'Southern hemisphere'
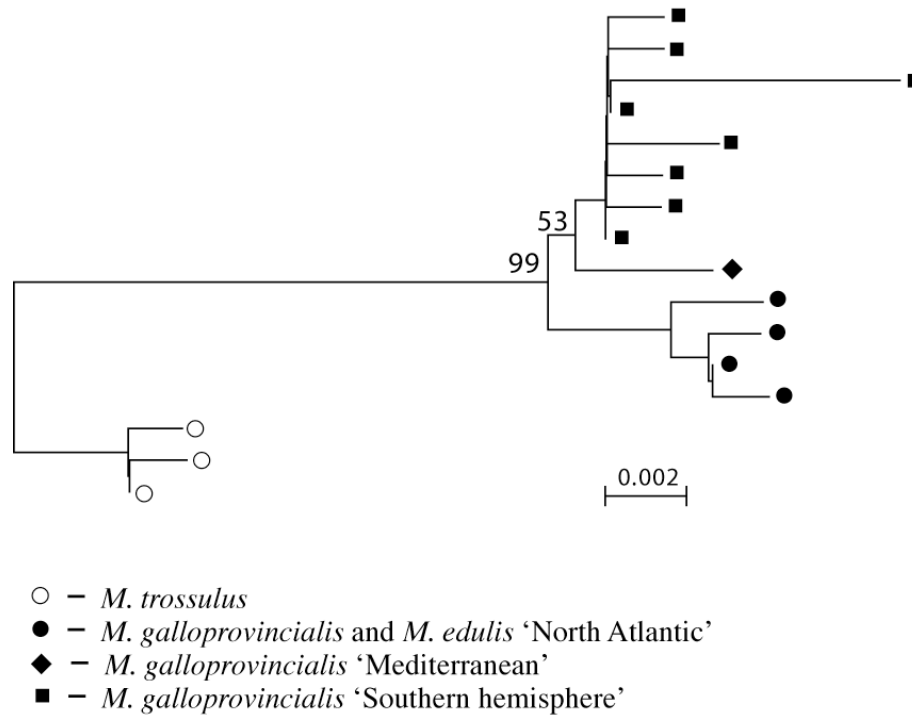
Figure 3.2 – Neighbour-joining tree of 16s gene for all *Mytilus edulis* species complex members. Codes for species and haplogroups included and numbers at nodes are bootstrap support values.

3.4.2.2 Me15/16

A single point mutation in the *M. galloprovincialis* Me15/16 amplicon (Inoue et al., 1995) causes the cut-site for the restriction enzyme AciI to disappear. Current and previously reported phylogenies (Daguin & Borsa, 2000) both indicate that highly supported and significant differentiation exists between hemispheric *M. galloprovincialis* lineages as shown by monophyly of lineages (Figure 3.3). In this case, hemispheric taxonomic differentiation, inferred from a single parsimoniously informative site within

the Me15/16 gene, is the direct target of restriction enzyme AciI and is therefore accurately captured by the Me15/16 RFLP assay. It is important to note that not all expected taxonomic designations were attained for sequence data. For example, Table 3.1 shows that one individual from each of Kaikoura (New Zealand), Port Arthur (Australia) and Concepcion (Chile) had a Northern hemisphere *M. galloprovincialis* genotype (when Southern hemisphere was expected) and resolved onto the Northern portion of the tree.
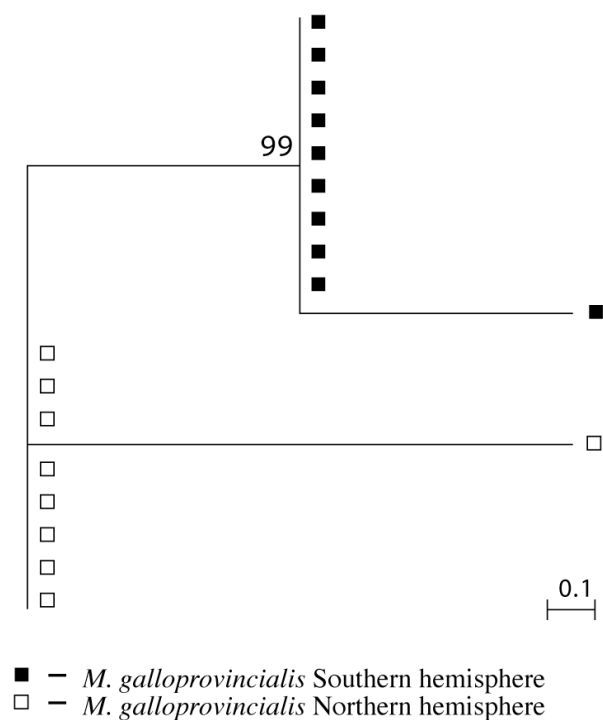
Figure 3.3 – Maximum parsimony tree of me15/16 gene for *M. galloprovincialis.* Codes for species lineages included and numbers at nodes are bootstrap support values.

3.4.3 Cytonuclear disequilibria statistics

Genotypic and allelic cytonuclear disequilibria values (D, D_1, D_2, D_3) describe differences of observed cytonuclear frequencies from expected values under Hardy-Weinberg equilibrium (Table 3.5). Allelic disequilibrium is defined here as any significant deviation from the expected frequency of cytonuclear haplotypes predicted by the product

of each allele frequency. Genotypic disequilibria that are significantly different from the null hypothesis of no cytonuclear disequilibrium occurred for NZ and Chilean Southern lineage homozygotes D_1 (Table 3.5). Hybrids (D_2) from NZ and Northern lineage homozygotes (D_3) from Chile also had significant disequilibrium interactions. Standard error estimates reported in Table 3.5 show a deficiency in sample sizes for all Northern hemisphere homozygotes (except Chile) and heterozygotes from Australia. D' values were at their maximal and minimal bounds for all genotypic and allelic combinations in Chile and Australia, with the exception of heterozygotes in Chile. In cases where sample size is not sufficiently large, normalized D' values become an important tool for interpretation of cytonuclear disequilibria (Avise, 2000).Gametic phase (allelic) disequilibrium (D) is significantly different from the null hypothesis of no association for NZ and Chilean samples. Maximum and minimum disequilibria bounds were present in Chile and Australia for D_1 and D_3 and have opposing signs for each lineage although the polarity changes between the two sample sets (Table 3.5).

Non-significant disequilibria were identified in Australia (Table 3.5) but normalized disequilibria values were at maximal and minimal boundaries for all cytonuclear genotypic combinations. The normalized D' values at maximal and minimal boundaries may be even more biologically significant when associated with non-significant D values (Asmussen and Basten, 1996). Polarity of normalized disequilibria changed between native (D_1) and non-native (D_3) nuclear homozygotes and heterozygote (D_2) polarity aligned with non-native nuclear genotypes (D_3) (Table 3.5).

Table 3.5 - Cytonuclear disequilibria in samples (grouped by country) for each genotypic and allelic combination.

Sample Location		D_1	D_2	D_3	D
New Zealand	D	0.0449±0.0159*	-0.0394±0.0142*	-0.0055±0.0069	0.0252±0.0097*
	D'	0.5404	-0.5633	-0.4178	0.5236
Chile	D	0.0999±0.0316*	0.0369±0.0358	-0.1367±0.0284*	0.1183±0.0352*
	D'	1.000	0.5167	-1.000	1.000
Australia	D	-0.0333±0.0272	0.0200±0.0219	0.0133±0.0182	- 0.0233±0.0204
	D'	-1.000	1.000	1.000	- 1.000

D range –0.25,0.25 and D' range –1,1. *p<0.017 (Bonferroni correction for genotypic and allelic classes at each location).

3.5 Discussion

Dispersal, selection and assortative mating are driving forces that maintain and shape the spatial genetic composition of hybrid zones (Barton and Hewitt 1989; Arnold 1997; Gardner 1994a, 1997; Mallet 2007 and references therein), but the additive effects of genetic drift and asymmetric migration into hybrid regions may produce permanent cytonuclear disequilibrium in the unique case of human-mediated introduction events (REFS). The frequency and distribution of Northern hemisphere *M. galloprovincialis* at the Southern hemisphere sample locations investigated in this study is minimal and patchy, and hence does not conform to the typical structure of a hybrid region flanked by pure subpopulations of contributing taxa or to hybrid mosaic architecture closed to outside recruitment of pure taxa. The frequency and distribution of Northern hemisphere mussels throughout the sample sites is directly attributable to three main causes; (1) the stochasticity of introduction event mode, magnitude and frequency through domestic and/or international shipping and personal watercraft, (2) the chance cultivationof non-natives or hybrids through aquaculture practise, and (3) the natural dispersal of non-natives or hybrids from the original point of introduction.

Investigations into cytonuclear disequilibria of Northern hemisphere *M. trossulus/M. edulis* hybridization in and around the Baltic Sea have uncovered significant results when samples across sites are pooled but not when individual sites or groups of related populations have been tested (Smietkana et al., 2004; Kijewski et al., 2006). In this case, significant cytonuclear associations are thought to reflect covariation of allelic and haplotypic frequencies across a geographic range (Kijewski et al., 2006) and be influenced more by geographic factors than taxonomy (Kijewski et al., 2010).

Mytilus recombinant cytonuclear genotypes in NZ and Australia imply there are no intrinsic barriers to gene flow between lineages, while normalized disequilibria at maximal and minimal levels across most genotypic and haplotypic combinations in Chile and Australia indicate that lineage boundaries and taxonomic identities are not being compromised. Disequilibria magnitude and sign pattern variation among sample countries leads to possible combinations of or differences in processes affecting local hybridization dynamics. Disequilibria sign patterns may characterize the effects of continued migration into the hybrid region from both or either parental taxa (Arnold, 1993) and create permanent disequilibria. This is observed in NZ and Chile samples where $D_1 > 0$, $D_3 < 0$, and D_2 is allowed to vary (D is assumed to be at a steady state) (Avise et al., 1990; Arnold, 1993). The observation of this sign pattern in NZ and Chile suggests that disequilibria in these populations are permanent, at a steady state and that a model with migration alone can explain the data. Australia had a sign pattern opposite to that observed at other locations ($D_1 < 0$, $D_3 > 0$), indicating that continued migration into this hybrid region may not be a determining factor in (non-significant) disequilibria.

Sign opposition for D_2, positive in Chile and negative in NZ, is a curious departure from the hypothesis of continued migration into both hybrid regions and warrants further attention because the sign pattern is expected to be constant between countries. The

additive effect of assortative fertilization with or without gender bias can also produce the same sign pattern as continued migration while the sign of D_2 infers directionality (Arnold, 1993). The observation of no introgressed individuals and normalized disequilibria at their minimum and maximum values for Northern and Southern homozygotes, respectively, supports the hypothesis of high fidelity within the Southern hemisphere lineage due to either assortative fertilization or genetic drift from unequal population sizes. Accepting a more complicated model with migration and assortative fertilization is not as parsimonious as accepting the model with only migration, when sign patterns from both models are the same, but the differences between observed and expected number of cytonuclear genotypes in Chile (Table 3.3) indicates gender bias associated with assortative fertilization (Avise, 2000). Cytonuclear genotypes *AA/M*, *Aa/M* and *aa/m* exhibit greater observed than expected numbers of individuals whilst the opposite occurs for *Aa/m*. There are also no detected introgressed genotypes (*AA/m* or *aa/M*) and non-zero disequilibria values. The number of observed hybrids with *M* cytotype being greater than expected and hybrids with *m* mitotype less than expected (although both deviations are non-significant) in addition to the sign pattern of D, D_1, D_2, $D_3 \neq 0$ indicates that hybridization among taxa includes a Southern lineage maternal parent and Northern lineage paternal parent for the majority of crosses. The application of the classic definition of mate choice is not appropriate in the case of broadcast-spawning mussels, rather, temporal variation of gamete release and local hydrographic features usually determine fertilization dynamics among conspecifics (e.g., Gardner & Skibinski 1990) and reproductive barriers (intrinsic or extrinsic) may determine fertilization dynamics within Northern hemisphere congener hybrid regions (Rawson & Hilbish 1995). Spatial genetic composition within Northern hemisphere hybrid zones may be determined by differential selection associated with salinity and exposure (Gilg & Hilbish, 2003; Smietkana et al., 2004). Gender-biased hybridization could be caused by

directional selection of pre-settlement larvae due to habitat related causes or a bias of more males in the immigrating population(s) of Northern lineage individuals. Furthermore, hybrid sterility, reduced fecundity or temporal asynchronicity of gamete release in immigrating females with respect to reproductive success (interbreeding) with native Southern hemisphere males may also contribute to the observed disequilibria pattern.

Considering life history characteristics of blue mussels and our sampling strategy based entirely on adults (no larvae were sampled), the effects of assortative fertilization at the stage of gamete release and the effects of post-zygotic selection until larval settlement cannot be considered separately. Whether the effects of assortative fertilization/pre-settlement selection are acting alone or in addition to the effects of continued migration in creating and maintaining cytonuclear disequilibria in Chile is an avenue to be considered for future research.

In contrast, constant D' values across genotypic combinations in NZ that do not approach minimal or maximal values coupled with the observation of introgressed individuals indicate that assortative fertilization and/or pre-settlement selection factors are not as pronounced here as in Chile and continued migration is most likely the cause of the observed disequilibria. The significant difference between the number of observed and expected cytonuclear genotypes in NZ (Table 3.5) is most pronounced in heterozygotes with an excess number of maternal Northern parents and a deficit of Southern maternal parents. Assortative fertilization cannot be invoked in this case because the specific pattern of differences between observed and expected numbers of individuals does not match with the expected pattern (Avise, 2000) and first and later generation hybrid crosses include more Northern than Southern hemisphere maternal parents. Clear observations of introgressive hybridization in NZ and a somewhat specific directionality to include more Northern than Southern maternal parents indicate the hybridization dynamics in this

country are very different to what is observed in Chile. The lack of interlineage barriers to gene flow, observed introgressive hybridization and non-marginal normalized disequilibria indicates introductions of non-native mussels in NZ pose a greater threat to the genetic integrity of native mussels here than in the other studied countries (Australia is detailed below). Continued immigration of non-native mussels and large-scale potential dispersal capability of pelagic larvae could perpetuate genes throughout the national distribution and threaten the unique genetic identity of native mussels in NZ.

Australian sample sites cover two regions within the country, Southern Australia (Melbourne, n=10) and Tasmania (Port Arthur, n=20), and the lack of observed putative pure Northern *M. galloprovincialis* could indicate that these populations have not experienced recent non-native introductions. There are several possible reasons for this observation; (1) low rates of introduction, possibly associated with Australia's stringent national policies on marine biosecurity, (2) unfavourable environmental conditions (however, this seems most unlikely given the distribution of native blue mussels) or (3) a substantial amount of time has elapsed since the last introduction and pure non-native mussels have largely died out (a reasonable possibility). When one species or taxon is rare in a hybrid setting, it runs the risk of being swamped by introgressive hybridization and eventually losing out to the more common taxon by becoming extinct (Rhymer and Simberloff, 1996). Initially, the genetic make-up of the prevailing taxon will be different, but over time, the fate of non-specific alleles is controlled by drift and/or selection. In this specific case, normalized D' values at maximal and minimal limits associated with non-significant disequilibria reveal that drift and/or selection dynamics are maintaining the taxonomic boundaries between lineages. All recombinant cytonuclear genotypes are present in frequencies higher than expected (except *aa/M* at Melbourne frequency = 0, Table 3.3) leading to the conclusion that extensive introgressive hybridization has

occurred. This is followed with a caveat: Northern alleles in introgressed hybrids could be eventually lost if introductions of non-native mussels are rare enough, essentially preserving the genomic "purity" of native mussels. Australian sites contain hybrids with only maternal Southern parents (Table 3.3) indicating a gender bias in the initial hybridization event towards females of Southern origin, a bias that was also detected in Chile. Conversely, gender bias is not indicated in backcrossing hybridization events because individual frequencies of both backcrossed cytonuclear genotypes are greater than expected (non-significant). The lack of Northern cytonuclear genotypes observed at two locations in Australia (one is an active international port) points to a deficiency of pure Northern hemisphere migrants. Allelic and genotypic combinations take on a negative association for this region, indicating opposing effects from those observed in NZ and Chile.

3.5.1 Conclusions

Significant cytonuclear associations in Southern hemisphere native/non-native M. galloprovincialis hybrid regions are the first reported for *Mytilus* spp. within a single population (Smietkana et al., 2004; Kijewski et al., 2006 and 2010). The mode of introduction providing a unique character to these hybrid regions is not conducive to a simple interpretation of cytonuclear disequilibria beyond the effects of continued immigration into hybrid regions. Inconsistent genetic outcomes of hybridization across several contact regions point to variation in selective forces, assortative fertilization and/or genetic drift resulting either from genetically distinct native-range source populations in the Northern hemisphere or from habitat variation among geographic regions., This point emphasises the extent of the biosecurity threat posed by international shipping, and the extreme difficulty of predicting and preventing unwanted translocations. Perhaps the

most perplexing aspect to the patterns of disequilibria is that introgression was detected in NZ and Australia but not in Chile, although we are confident that non-native mussels have inhabited areas close to the Chilean collection sites for at least ten years. Further classification of introgressed individuals using one or more additional nuclear loci would benefit future investigations in order to determine with greater probability that introgression is not occurring. Lack of introgression points to barriers to gene flow that have previously been observed among members of the *M. edulis* species complex in hybrid zones (Rawson et al., 1999), but not between divergent populations (or lineages) of a single species. If the source collection of introduced mussels is genetically homogeneous across study regions (we recognise that this may well not be the case), then barriers to gene flow are only functioning within the Chilean population, indicating that this specific population may contain uncharacterized genetic variation that leads to reproductive isolation between itself and the Northern hemisphere *M. galloprovincialis* lineage. Across Southern hemisphere sample regions, introduced Northern hemisphere *M. galloprovincialis* exhibit ancestry from both "Mediterranean" and "North Atlantic" haplogroups (Hilbish et al., 2000; Westfall et al., 2010a and b) and could, therefore, potentially be sourced from a very large geographical range (Europe, North America, Asia or South Africa), and the genetic composition of non-native mussels may be a factor in the patterns of disequilibria across these studied hybrid regions.

The genetic swamping of Northern lineage *M. galloprovincialis* alleles in Australia is an example of how extremely low frequencies of non-native alleles can affect the genetic make-up of native populations through introgression. A gender bias in hybrid crosses (first generation and beyond) imposes directionality of introgression and preserves Southern mitochondrial genomic identity within the population. The gender bias of first generation crosses coupled with no evidence of introgression in Chile is a desired outcome when

considering the potential of non-native establishment to alter the genomic content of native populations through introgressive hybridization. In Australia and Chile, the potential exists for minimal impact on the native population from the introduction of Northern hemisphere *M. galloprovincialis*. The outcome in NZ is not as optimistic because introgression, continued immigration of non-native mussels, slight gender bias to interlineage crosses for more Northern than Southern maternal parents (i.e. Southern mtDNA is not preserved) and no detected directional selection all lead to continued cytonuclear disequilibria and genetic mixing of lineages.

In terms of a better understanding of the dynamics of the interaction between introduced Northern mussels with native Southern mussels there is a clear need for laboratory-based experiments and field observations to address these points (e.g., Gardner 1994b, 1995, 1996; Gardner & Skibinski 1990, 1991; Gardner et al. 1993; Gardner & Thompson 2001; Wood et al. 2003). This work will have immediate relevance to the Southern hemisphere situation as described here, but will also be important in terms of invasive Southern hemisphere *M. galloprovincialis* that have yet to be reported in the Northern hemisphere. Given the propensity of *M. galloprovincialis* generally to be a successful invader and the now multiple opportunities for translocation afforded by international shipping (e.g., Apte et al. 2000; Coutts et al. 2010; Hopkins et al. 2010), it seems inevitable that Southern lineage *M. galloprovincialis* will be identified in the Northern hemisphere in the near future. Investigating the factors that have lead to inconsistent genetic consequences of non-native introduction events observed in hybrid regions of three Southern hemisphere countries is an important avenue for future research and could help manage and predict the effects of Northern lineage *M. galloprovincialis* introductions in other areas of the world.

3.6 References

Apte S, Holland BS, Godwin LS & Gardner JPA (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Arnold J (1993) Cytonuclear disequilibria in hybrid zones. *Annual Review of Ecology and Systematics* 24, 521-554.

Arnold ML (1997) *Natural hybridization and evolution*. Oxford University Press, Oxford, UK.

Asmussen M, Basten C (1994) Sampling theory for cytonuclear disequilibria. *Genetics* 138: 351-1363.

Asmussen M, Basten C, (1996) Constraints and normalised measure for cytonuclear disequilibria. *Heredity* 76 201-214.

Asmussen M, Arnold J, Avise JC (1987) Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* 115: 775-768.

Avise JC (2000) Cytonuclear genetic signatures of hybridisation phenomena: Rationale, utility and empirical examples from fishes and other aquatic animals. *Reviews of Fish Biology and Fisheries* 10: 253-261.

Avise JC, Nelson WS, Arnold J, Koehn RK, Williams GC, Thorsteinsson V (1990) The evolutionary genetic status of Icelandic eels. *Evolution* 44: 1254-1262.

Barton NH & Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature* 341: 497-503.

Basten C, Asmussen M (1997) The exact test for cytonuclear disequilibria. *Genetics* 146: 1165-1171.

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

Coutts ADM, Piola RF, Hewitt CL, Connell SD, Gardner JPA (2010) Effect of vessel voyage speed on survival of biofouling organisms: implications for translocation of non-indigenous marine species. *Biofouling* 26: 1-13.

Daguin C, Borsa P (2000) Genetic relationships of *Mytilus galloprovincialis* Lmk. populations worldwide: evidence from nuclear-DNA markers. In: *Bivalve Systematics and Evolution* (A. Crame, E. Harper & J. Taylor, eds), pp. 389-397. Geological Society of London Special Publications volume 177, London.

Gardner JPA (1994a) The structure and dynamics of naturally occurring hybrid *Mytilus edulis* Linnaeus, 1758 and *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia,

Mollusca) populations: review and interpretation. *Archiv für Hydrobiologie, Monographische Beitrage* 99: 37-71.

Gardner JPA (1994b) The *Mytilus edulis* species complex in southwest England: multi-locus heterozygosity, background genotype and a fitness correlate. *Biochemical Systematics & Ecology* 22: 1-11.

Gardner JPA (1995) Developmental stability is not disrupted by extensive hybridization and introgression among populations of the marine bivalve molluscs *Mytilus edulis* (L.) and *M. galloprovincialis* (Lmk.) from southwest England. *Biological Journal of the Linnean Society* 54: 71-86.

Gardner JPA (1996) The *Mytilus edulis* species complex in southwest England: the extent of hybridization and introgression and their effects upon interlocus associations and morphometric variation. *Marine Biology* 125: 385-399.

Gardner JPA (1997) Hybridization in the Sea. *Advances in Marine Biology* 31: 1-78.

Gardner JPA (2004) A historical perspective of the genus *Mytilus* (Bivalvia: Mollusca) in New Zealand: multi-variate morphometric analyses of fossil, midden and contemporary blue mussels. *Biological Journal of the Linnean Society* 82: 329-344.

Gardner JPA, Skibinski DOF (1990) Genotype-dependent fecundity and temporal variation of spawning in hybrid mussel populations. *Marine Biology* 105: 153-162.

Gardner JPA, Skibinski DOF (1991) Biological and physical factors influencing genotype-dependent mortality in hybrid mussel populations. *Marine Ecology Progress Series* 71: 235-243.

Gardner JPA, Skibinski, DOF, Bajdik CD (1993) Shell growth and viability differences between the marine mussels *Mytilus edulis* (L.), *Mytilus galloprovincialis* (Lmk.) and their hybrids from two sympatric populations in S.W. England. *Biological Bulletin* 185: 405-416.

Gardner JPA, Thompson RJ (2001) The effects of coastal and estuarine conditions on the physiology and survivorship of the mussels *Mytilus edulis*, *M. trossulus* and their hybrids. *Journal of Experimental Marine Biology & Ecology* 265: 119-140.

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

Grant S, Cherry I (1985) *Mytilus galloprovincialis* Lmk. In Southern Africa. *Journal of Experimental Marine Biology and Ecology* 90: 179-191.

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

Hopkins GA, Forrest BM, Jiang W, Gardner JPA (2010) Successful eradication of a non-indigenous marine bivalve from a subtidal soft sediment environment. *Journal of Applied Ecology*, in press.

Inoue K, Waite J, Matsouka M, Odo S, Harayama S (1995) Interspecific variations in adhesive protein sequences of *Mytilus edulis*, *M. galloprovincialis*, and *M. trossulus*. *Biological Bulletin* 189: 370-375.

Kijewski T, Zbawicka M, Väinölä R, Wenne R (2006) Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biology* 149: 1371-1385.

Lowe S, Browne M, Boudjelais S, De Poorter M (2000) 100 of the world's worst invasive alien species: A selection from the global invasive species database. Published by The Invasive Species Specialist Group (ISSG), a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), pp. 12.

Mallet J (2007) Hybrid speciation. *Nature* 446: 279-283.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London Series B* 363: 2971-2986.

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

McDonald J, Seed R, Koehn R (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology* 111: 323-333.

Rawson PD, Hilbish TJ (1995) Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution* 12: 893-901.

Rawson PD, Hilbish TJ (1998) Asymmetric introgression of mitochondrial DNA among European populations of blue mussels (*Mytilus* spp.). *Evolution* 52: 100-108.

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

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systemtaics* 27: 83-109.

Santaclara FJ, Espineira M, Cabado AG, Aldasoro A, Gonzalez-Lavin N, Vieites JM (2006) Development of a method for the genetic identification of mussel species belonging to *Mytilus*, *Perna*, *Aulacomya* and other genera. *Journal of Agricultural and Food Chemistry* 54: 8461-8470.

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

Toro JE (1998) PCR-based nuclear and mtDNA markers and shell morphology as an approach to study the taxonomic status of the Chilean blue mussel, *Mytilus chilensis* (Bivalvia). *Aquatic Living Resources* 11: 347-353.

Toro JE, Ojeda JA, Vergara AM, Castro GC, Alcapán AC (2004) Molecular characterization of the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) demonstrates evidence for the occurrence of *Mytilus galloprovincialis* in southern Chile. *Journal of Shellfish Research* 24: 1117-1121.

Westfall KM, Gardner JPA (2010) Genetic diversity of Southern hemisphere blue mussels (Mytilidae; Bivalvia) and the identification of non-indigenous taxa. *Biological Journal of the Linnean Society* 101: 898-909.

Westfall KM, Wimberger PH, Gardner JPA (2010) An RFLP mtDNA assay to determine if *Mytilus galloprovincialis* (Mytilidae; Bivalvia) is of Northern or Southern hemisphere origin. *Molecular Ecology Resources* 10: 573-575.

Wood AR, Turner G, Skibinski DOF, Beaumont AR (2003) Disruption of doubly-uniparental inheritance of mitochondrial DNA in hybrid mussels (*Mytilus edulis* x *M. galloprovincialis*). *Heredity* 91: 354-360.

Chapter 4: Phylogenetics of the *Mytilus edulis* species complex (Mytilidae; Bivalvia): on the origin and divergence time of Southern hemisphere members.

4.1 Abstract

Attention on the evolutionary history and divergence time of *Mytilus edulis* complex species in the Southern hemisphere is moderate in comparison to Northern hemisphere congeners. Allozyme and nuclear evidence underlies published hypotheses of multiple migration events leading to introgressive hybridisation through secondary contact and reticulated evolution, whilst mtDNA evidence underlies conflicting hypotheses on branching patterns within the complex with respect to genealogical relationships among three lineages of *M. galloprovincialis*. Incongruent single gene phylogenetic trees highlight the importance of inferring species relationships from single gene histories, and the modes of reproduction of *M. edulis* complex species, including introgressive hybridisation and mitochondrial recombination, may undermine classic phylogenetic methods. It is therefore important to unify all available evidence into a single theory. The use of phylogenetic methods is nonetheless important for inferring general evolutionary patterns and estimating species divergence times, and congruence among methodologies and other single gene histories allows greater confidence of correct species histories. The current phylogenetic investigation relies on two mtDNA genes with different mutational properties (16s and COIII) used singly and in conjunction to present, for the first time, multiple gene phylogenetic reconstructions under mixed evolutionary models and the Bayesian inference framework. A maximum likelihood framework was also used to provide direct comparisons with previously published topologies. Branching patterns between single gene topologies were incongruent based on a weak phylogenetic signal producing basal polytomies for the 16s gene. COIII single and 16s/COIII combined topologies were congruent with the exception of *M. trossulus*, which appeared as a monophyletic clade in the combined but not single COIII topology. The most supported topologies indicate that Southern hemisphere *M. galloprovincialis* and the Northern hemisphere *M. galloprovincialis*/*M. edulis* haplogroups are sister clades that diverged after the Northern hemisphere *M. galloprovincialis* haplogroup, contrasting with all published topologies to date. Although strict branching patterns were not congruent with previous results, the hypothesis of an early *M. galloprovincialis* introgressed with *M. edulis* genes

ancestor of Southern hemisphere *M. galloprovincialis* is in line with much of the previously described allozyme, nuclear and mtDNA evidence. Coalescent based dating is preferred to a strict Yule process of speciation because of reliance on fossil calibration is undermined by uncertain time estimates of biotic exchange through the Bering Strait. Coalescent based dating places the divergence of the Southern hemisphere *M. galloprovincialis* lineage during the late Pliocene at 2.04 to 2.66 mya, much earlier than previous estimates. Additionally, the Southern hemisphere populations are separated into geographic origin clades, a result congruent with a previously published topology. Reciprocal monophyly and reproductive isolation underlie the suggestion to classify Southern hemisphere *Mytilus* into a separate species or subspecies of the *M. edulis* complex. The addition of nuclear gene sequences and/or a break from the bifurcating nature of phylogenetic inference to include a network type of evolutionary inference may go further to discern phylogenetic relationships among *M. edulis* complex taxa.

4.2 Introduction

The concept of phylogeny has been interpreted in many different ways, inferences from tree topology range from the population level to the deep evolutionary relationships within a wide taxonomic breadth. Population geneticists often view a phylogeny as a model of probability of interbreeding (Maddison 1997), branches represent populations that have undergone reproductive isolation, and the topology is a history of the interrelationships among interbreeding probabilities even if the cause cannot be determined. This is part of the interface between gene and species trees, and the researcher is often confronted with having to infer the evolutionary history of lineages and species from single gene trees, which, more often than not, exhibit incongruence.

Systematic error from non-phylogenetic signals can lead to incongruence among single gene trees and adding genes is an attractive way to solve this problem. Sources of non-phylogenetic signals include nucleotide composition heterogeneity among species, substitution rate variation among lineages and within site rate variation. Well-supported

but incorrect topologies may result from the bias associated with such systematic errors as the non-phylogenetic signal takes over the true phylogenetic signal. The systematic error of non-phylogenetic signals does not even out with the addition of more data, rather, the bias can become stronger and lead to greater incongruence or higher statistical nodal support.

The addition of taxa or genes has been a topic of debate for increasing the accuracy of phylogenies. Both methods have inherent problems; the addition of taxa whilst using one or a few genes can be under the influence of non-phylogenetic signals as mentioned before, and the addition of genes while limiting the number of taxa can be sensitive to homoplasy. Public access to a large number of genes for non-model organisms is limited and a common approach now is to use between tens and hundreds of taxa with one or a few genes. Phylogenetic accuracy may have a greater increase with the addition of genes than with taxa (Rokas & Carroll 2005) but each phylogenetic problem is different and may warrant another approach.

4.2.1 DNA Data Partitioning

Increased data availability for phylogenetic analyses has created debate around the function and application of data partitioning. Multiple genetic markers, and even the addition of morphological characters, are open to a wide range of current and historical evolutionary processes of which little is usually known for the taxonomic group under study. Nucleotide properties of DNA sequences show the downstream effects of evolutionary processes on that single marker of which parametric models like maximum likelihood and Bayesian inference are subject to. When using multiple markers in a single parametric analysis it is important to firstly determine if the separate data sets contain conflicting phylogenetic signals (Bull et al. 1993). Although complete agreement between

trees is not expected due to sampling error, before combining data it is necessary to determine if the level of difference between separate data sets is greater than what would be expected by chance (Bull et al. 1993, Huelsenbeck et al. 1996). If the conflicting phylogenetic signal is greater than what is expected by chance, partitioning the data into single markers and allowing each partition to operate under a separate set of model parameters will greatly enhance the accuracy of reconstruction.

Historically, methods to get around data partitioning focus on topological comparisons derived from different markers in separate analyses. Topological comparisons will be employed in this study and are useful when inferring evolutionary histories for separate markers or enforcing the point that phylogenies derived from single markers aren't necessarily indicative of species history, but combining multiple markers into a single phylogenetic analysis is an important step in the direction of inferring the evolutionary history of the taxonomic group under consideration.

4.2.2 *Mytilus edulis* species complex taxonomy

Members of the *Mytilus edulis* species complex have been intensely studied since identification of the complex over 20 years ago. Seminal research (McDonald & Koehn 1988, McDonald et al. 1991) combined biochemical variation (allozymes) with morphological characteristics to define a complex composed of three sibling species: *M. edulis* Linnaeus 1758, *M. galloprovincialis* Lamarck 1819 and *M. trossulus* Gould 1850. The impetus for defining a complex of sibling species was due to widespread discontinuity of taxonomic delineation among morphologically and biochemically variable worldwide populations of what were considered regional sub-species of *M. edulis* from the mid-1950's onwards (Soot-Ryen 1955). In the Southern hemisphere these included *M. edulis aoteanus* Powell (based on a sampled from Campbell Island, New Zealand) and *M. edulis*

planulatus Lmk. (based on a sampled from Tasmania, Australia) (Fleming 1959) and, until recently, Chilean *Mytilus* spp. have been considered a separate species, *Mytilus chilensis* (Toro 1998 & 2005, Krapivka et al. 2007). A greater understanding of molecular variation among Southern hemisphere populations has reclassified all of the above-mentioned previously recognised species to the Southern hemisphere lineage of *M. galloprovincialis* (Daguin & Borsa 2000, Hilbish et al. 2000, Westfall & Gardner in press).

The presence of *Mytilus* spp. in rocky intertidal habitats is widespread throughout anti-tropical regions of the world due to natural and anthropogenic assisted migration, and the pathways of human-mediated introduction are well documented (see Wonham 2004 for a review of *M. galloprovincialis* introductions). The natural distribution of *Mytilus edulis* complex species encompasses all major oceans and seas: *M. trossulus* in the North Pacific, NW Atlantic and Baltic Sea and *M. edulis* in the North Atlantic (McDonald et al. 1991, Väinölä & Hvilsom 1991, Seed 1992, Suchanek et al. 1997). *M. galloprovincialis* is the most widely distributed species being found naturally in both hemispheres: in the Northern hemisphere *M. galloprovincialis* is native to the Atlantic coasts of the Southern British Isles, continental Europe and northern Africa (McDonald et al. 1991, Väinölä & Hvilsom 1991, Seed 1992, Suchanek et al. 1997), and in the Southern hemisphere it is native to the Atlantic and Pacific coasts of Southern South America, New Zealand (NZ), Australia, the Falkland Islands and the Kerguelen Islands (McDonald et al. 1991, Daguin & Borsa 2000, Hilbish et al. 2000, Gardner 2004, Borsa et al. 2007). Human-mediated introductions of *M. galloprovincialis* extend the Southern hemisphere range to include South Africa and the Northern hemisphere range to the NE Pacific, Japan (Wilkins et al. 1983), Hong Kong (Lee & Morton 1985) and possibly Hawaii (Apte et al. 2000). Extensive hybridisation occurs in areas of sympatric distribution in both natural and anthropogenic influenced ranges although there has previously been no record of hybridisation in the Southern hemisphere.

Areas of hybridisation maintain species boundaries through local adaptation to environmental variation (Riginos et al. 2002).

4.2.3 Fossil calibrated timing of speciation events

The current thinking on the divergence of species in the Northern hemisphere is thought to have occurred several million years ago with *M. trossulus* diverging in the Pacific Ocean and the ancestor to *M. edulis* and *M. galloprovincialis* migrating to the Atlantic Ocean after the Bering Strait opened. The estimated time of the opening of the Bering Strait at 3.5 mya (Vermeij 1991) has been used in many attempts to date the divergence of major clades within the *M. edulis* species complex tree (Hilbish et al. 2000, Wares & Cunningham 2001, Riginos et al. 2004, Gérard et al. 2008), but other evidence suggests this timing is underestimated. Marinkovich & Gladenkov (1998) place the opening of the Bering Strait between 4.8 and 7.4 mya, the former based on diatom biochronology and the latter based on stratigraphic evidence. With the earliest estimate of the first opening of the Bering Strait to be 7.4 mya (Marinkovich & Galdenkov 1998), it would be remiss to assume that *Mytilus* spp. could not have migrated to the Atlantic Ocean earlier than 3.5 mya.

Although major colonisation events may have occurred at 3.5 mya as represented by a rich fossil record, reproductive isolation across the Arctic dated from a conservative 4.8 mya may also have contributed to the divergence of *M. edulis*. Marinkovich & Gladenkov (1998) attribute the Pliocene appearance of molluscs and other marine biota in the Atlantic to a change in water flow from a Southerly to a Northerly direction through the Bering Straight 3.5 mya possibly in response to the creation of the Isthmus of Panama. Even with a northerly flow through the Bering Strait from 3.5 mya, trans-Arctic biotic exchange occurred in both directions but with the majority of taxa migrating from the

Pacific to the Atlantic. Vermeij (1991) stated that Pacific taxa could theoretically have appeared in the Chesapeake Bay region and North Sea basin in the Miocene (>5.3 million years ago [mya]) because of similar ecological opportunities if a marine connection between the Pacific and Atlantic Oceans existed at that time. Vermeij (1991) relied on the assumption that there was no marine connection during the late Miocene and therefore took the limited fossil record at face value, but newer evidence suggests there was a marine connection between ocean basins during the late Miocene, which belies the assumptions of Vermeij (1991) for the timing of the trans-Atlantic biotic interchange. Early Pliocene fossil assemblages are lacking in the cool-temperate NW Atlantic and Vermeij (1991) was careful to make a provisional status of the timing of colonisation for bivalves of that region, indicating a possibility of a late-Miocene/early-Pliocene presence of *Mytilus* spp. in the NW Atlantic.

M. galloprovincialis and *M. edulis* are considered sister taxa and *M. galloprovincialis* is believed to have diverged in the Mediterranean Sea 1-2 mya (Barsotti & Meluzzi 1968, Rawson & Hilbish 1995 and 1998, Quesada et al. 1995a and b). Mitochondrial DNA differentiation reveals "Mediterranean" and "North Atlantic" lineages of *M. galloprovincialis* in Europe separated by the Almeria-Oran oceanographic front off the coast of southern Spain to the east of the Strait of Gibraltar, a hydrographic boundary between Atlantic and Mediterranean surface waters (Quesada et al. 1995a and b). Phylogenetic reconstruction from mtDNA consistently differentiates among haplogroups based on geographic and species barriers; *M. edulis* and *M. galloprovincialis* distributed in the Atlantic Ocean are closely related due to introgressive hybridisation through secondary contact and form a single haplogroup (Rawson & Hilbish 1995, Hilbish et al. 2000). Other haplogroups include *M. trossulus*, *M. galloprovincialis* "Mediterranean lineage" and *M. galloprovincialis* "Southern hemisphere lineage" (Westfall et al. 2010).

4.2.4 Southern hemisphere *Mytilus* sp.

Although the Northern hemisphere populations of *Mytilus edulis* species complex members have been well studied, there are opposing viewpoints on the taxonomic status and evolutionary relationships of Southern hemisphere *Mytilus* spp. McDonald et al. (1991) classifies Southern hemisphere *Mytilus* spp. into two main groups; (1) mussels from the Kerguelen Islands, the Falkland Islands, Chile and Argentina possess allozyme alleles most closely related to *M. edulis* from the Northern hemisphere which were 'morphologically intermediate' between *M. edulis* and *M. trossulus* and were therefore temporarily ascribed to *M. edulis*, and (2) mussels from New Zealand and Australia which exhibit morphological and allozyme similarity to Northern hemisphere *M. galloprovincialis*. There are fossil records of *Mytilus* spp. in New Zealand (McDonald et al. 1991, Gardner 2004), Australia (Hope et al. 1977) and Argentina (Genta-Iturrería et al. 2009) indicating smooth-shelled blue mussels are native to these areas. Soot-Ryen (1955) commented that *Mytilus* spp. are relatively recent in the geological time scale with no definitive shells dating from before the Pliocene (5.332 mya). Fossil records of Mytilid mussels are patchy in New Zealand owing to a lack of rocky shore facies (Fleming 1959). A record of *M. edulis aoteanus* Powell from a shell bed in Devonport, Auckland (New Zealand), suggests this species had a different distribution in historical context than present day (Fleming 1959). New Zealand *Mytilus* spp. fossils date to just over 1 mya (Gardner et al. 2004 and references therein).

Since the seminal research of McDonald et al. (1991) newly developed nuclear and mitochondrial molecular markers have attempted to decipher the Southern hemisphere populations' evolutionary relationships with the worldwide distribution of this species complex (Daguin & Borsa 2000, Hilbish et al. 2000, Borsa et al. 2007, Gérard et al. 2008).

Daguin & Borsa (2000) utilised seven allozyme loci from several studies to produce a phylogenetic tree that places the two Northern *M. galloprovincialis* lineages as sister groups and infers Southern hemisphere *M. galloprovincialis* diverged before the two lineages of Northern *M. galloprovincialis*. They also concluded that the ancestor to Southern hemisphere *M. galloprovincialis* derived from a proto-*M. galloprovincialis* population introgressed with *M. edulis*-like genes (Daguin & Borsa 2000). Borsa et al. (2007) went on to describe Kerguelen Island mussels as having a mixed *M. edulis*/*M. galloprovincialis* nuclear genome and Tasmanian populations as possessing a *M. galloprovincialis* genome with *M. edulis* introgression from secondary contact sometime after the initial settlement of *M. galloprovincialis* (as one nuclear *M. edulis* allele was fixed in the population). The genomic reticulation observed in these populations was attributed to hybridisation followed by genetic drift and pseudo-selection for *M. edulis* genes (Borsa et al. 2007).

Hilbish et al. (2000) and Gérard et al. (2008) have put forward mitochondrial-DNA driven hypotheses on Southern hemisphere *Mytilus* spp. evolution. The 16s rDNA is a slowly evolving gene aptly used to decipher species relationships and the cytochrome-oxidase 1 gene (COI) is a faster evolving gene that can be applied to look at genealogical relationships among recently diverged lineages or populations. Hilbish et al. (2000) used 16s to hypothesise that Southern hemisphere and the "Mediterranean" *M. galloprovincialis* lineages are sister taxa that diverged after the "North Atlantic" *M. galloprovincialis*/*M. edulis* shared haplogroup. The COI single gene paints a different picture whereby Southern *M. galloprovincialis* diverged before the two Northern hemisphere *M. galloprovincialis* lineages (Gérard et al. 2008). Both authors date the hemispheric divergence using a fossil calibration point of 3.5 mya as the divergence between *M. edulis* and *M. trossulus* (to coincide with the opening of the Bering Strait) and congruent results

placed the migration from the Northern to the Southern hemisphere in the Pleistocene (Hilbish et al. 2000, Gérard et al. 2008) at 0.84 mya (Gérard et al. 2008). Additionally, Hilbish et al. (2000) hypothesised an Atlantic route of migration for *Mytilus* spp.

From a morphological standpoint using contemporary, midden and fossil valves, Gardner (2004) classified the majority of New Zealand mussels as *M. galloprovincialis* with a provisional *M. edulis*/*M. galloprovincialis* ancestry for a special Bay of Islands population. The finding of pre-European arrival *M. galloprovincialis* in populations across the range of this species in NZ is consistent with the molecular data and interpretations of Hilbish et al. (2000) and Gérard et al. (2008)

The hypothesis of *M. galloprovincialis* (both Southern and Northern lineages) diverging around the same time from *M. edulis* has never been considered valid to explain the presence of *M. edulis* alleles in the Southern *M. galloprovincialis* genomic background. By chance, some alleles that we consider as diagnostic for species discrimination of *M. edulis* in the Northern hemisphere could have become fixed in populations considered as *M. galloprovincialis* due to common ancestry. The issue comes down to the fact that our taxonomic thinking on the *M. edulis* species complex has been derived solely from Northern hemisphere populations. Size variation in nuclear gene loci has been used to describe Northern hemisphere species. To extend the use of these diagnostic markers into the Southern hemisphere without previous knowledge of the taxonomic identity or evolutionary history of these populations may lead to misleading results and interpretations. An accurate phylogenetic history of *Mytilus edulis* complex evolution is required before we can ascribe "diagnostic" nuclear alleles to populations in the Southern hemisphere. This thinking is based on fundamental questions such as: (1) are *M. edulis* alleles observed in Southern hemisphere populations because of persistent ancestral

polymorphism due to common ancestry and (2) is there enough variation in Southern hemisphere populations to warrant a new species status?

In the view of population genetics, gene trees based on the COI mtDNA gene indicate strongly supported separation of clades within the geographical context for the Southern hemisphere lineage of *M. galloprovincialis* (Gérard et al. 2008). The taxa of these clades are not viewed within the biological species concept, rather, within the intraspecific probability of interbreeding concept as the branch tips represent individuals which are currently viewed as belonging to the same lineage of species. The evidence provided in this paper for the taxonomic differentiation of the Southern hemisphere *M. galloprovincialis* lineage as a separate species suggests these clades be viewed within the classic species concept. The inferences of evolutionary history derived from such a tree encompasses multiple levels; reproductively isolated populations of the same lineage of a single species, monophyletic sister groups representing lineages of the same species, and other polyphyletic groups representing the single gene history of multiple species subject to varying degrees of reproductive isolation (i.e. Western Europe) within the complex.

Horizontal gene transfer can occur between recently diverged species or lineages through hybridisation, and this is a definite possibility between hemispheric lineages of *M. galloprovincialis*. Although separated by uninhabitable equatorial waters through which natural migration events are unlikely to occur with sufficient frequency to prevent reproductive isolation, *M. galloprovincialis* is easily transported via human-mediated events and has the potential to hybridise with native mussels. This transfer is theoretically possible through northerly and southerly directions across the equator. Introgression of mitochondrial genes through hybridisation is variable, although common, in the Northern hemisphere (Rawson & Hilbish 1995, Kijewski et al. 2006). The transport of non-indigenous mussels and the resultant hybridisation between lineages leads to incongruence

in phylogenies within the geographical context and this is taken into consideration for inference of genetic differentiation in *Mytilus* spp.

4.2.5 Investigative objectives

The 16s rDNA gene is one of the slowest evolving genes in the mitochondrial genome (Moritz et al. 1987) and estimates of divergence based on this gene tend to be conservative (Rawson & Hilbish 1995). Conversely, the cytochrome *c* oxidase complex has shown accelerated rates of evolution in *Mytilus* spp. with an estimated rate of 3.67 x 10^{-4} substitutions/nucleotide site/my (million years) (Hoeh et al. 1996). The current study uses the cytochrome oxidase subunit III (COIII) gene separately and in combination with the 16s rDNA gene to reconstruct the phylogenetic history of the *M. edulis* complex from a Southern hemisphere perspective. Phylogenetic reconstructions centre on Bayesian and maximum likelihood methods and congruencies among these methods are also examined. The main purpose of this study is to build on the existing phylogenetic work of Hilbish et al. (2000) and Gérard et al. (2008) in an attempt to reconcile a framework of evolutionary history, migration pathways and genealogical relationships for the Southern hemisphere populations of smooth-shelled blue mussels.

Elements in this study that add to previously published phylogenies are; (1) the addition of a new single gene (COIII) phylogeny derived from mussels sampled across the natural distribution in the Southern hemisphere and all taxa in the Northern hemisphere, (2) an attempt to combine two mitochondrial genes with different evolutionary properties into a single mixed model analysis and (3) reconciliation of the current data and previously published hypotheses. Single gene topologies from Hilbish et al. (2000) and Gérard et al. (2008) are rigorously tested within Bayesian and maximum likelihood frameworks against the single COIII and 16s gene genealogies derived from this study to infer confidence

levels for these new phylogenies and test hypotheses on Southern hemisphere *Mytilus* spp. evolution.

Attempts to date the hemispheric divergence of *M. galloprovincialis* in this study will, for the first time, invoke coalescent theory (Kingman 1982, Nielson & Wakely 2001) and not rely on fossil calibration. For purposes of direct comparisons with previously published divergence estimates, traditional dating techniques relying on a birth-death model of speciation (Yule 1925) and fossil calibration points will be employed using the 3.5 mya split for *M. trossulus* and *M. edulis* used in previous estimates and the newly established conservative 4.8 mya split for the same taxa.

4.3 Materials and Methods

4.3.1 Sample collection

A total of 308 samples were collected from 41 sites throughout the world (Figure 4.1, Table 4.1) between 2000 and 2009 and stored in 70% ethanol until DNA extraction. Representatives of *M. edulis* complex taxa were targeted from diverse locations to incorporate the entire distribution of species with an emphasis on sampling as much of the *Mytilus* spp. distribution in the Southern hemisphere as possible. Sampling coverage in the Southern hemisphere was excellent with the exception of the Kerguelen Islands and South Africa. The population of *M. galloprovincialis* in South Africa is purely non-native, being introduced through human-mediated vectors (Grant & Cherry 1985, Branch & Steffani 2006). South American sample locations include the entire distribution of *Mytilus* in Chile with three locations sampled (n=43) and a single location in Ushuaia, Argentina (n=9). The geographically proximate Falkland Islands were sampled on both Bense Island (West Falkland) (n=2) and Port Stanley (East Falkland) (n=2). New Zealand sampling included the entire distribution throughout the North (10 sites, n=33) and South Islands (7 sites,

n=46) and all offshore islands from which it is recorded; Campbell (n=6), Auckland (n=10) and the Chathams (n=13). Additional samples were collected from an offshore oil rig (The Ocean Patriot) positioned in the Taranaki Bight (n=8). In Australia, ethanol preserved specimens from the estuarine Wonboyn River (n=2) and Wonboyn Lake (n=1) (province of New South Wales) (Figure 4.1c) were obtained from the Australian National Museum, the small sample size is due to the relative rarity of blue mussels at these locations. Port specimens from Port Arthur (n=20) (historical port, Tasmania) and Port Phillip Bay (n=20), Melbourne (Victoria) were also collected (Figure 4.1c).   In the Northern hemisphere, North American *M. trossulus* samples were collected from British Columbia (3 sites, n=14) (Pacific) and Newfoundland (2 sites, n=8) (Atlantic), Canada (Table 4.1). European *M. galloprovincialis* samples were collected from Russia (n=13) (Black Sea), the port of Izmir, Turkey (n=17) (Mediterranean Sea), and Galicia, Spain (n=10) (Atlantic Ocean) (Table 4.1).   European *M. edulis* samples were collected from Oosterschelde, Netherlands (n=19), and Lowestoft (n=12) and Mundesley (n=9), England (all North Sea) (Table 4.1).   North American *M. edulis* samples were collected from Newfoundland, Canada (n=8) (Table 4.1).
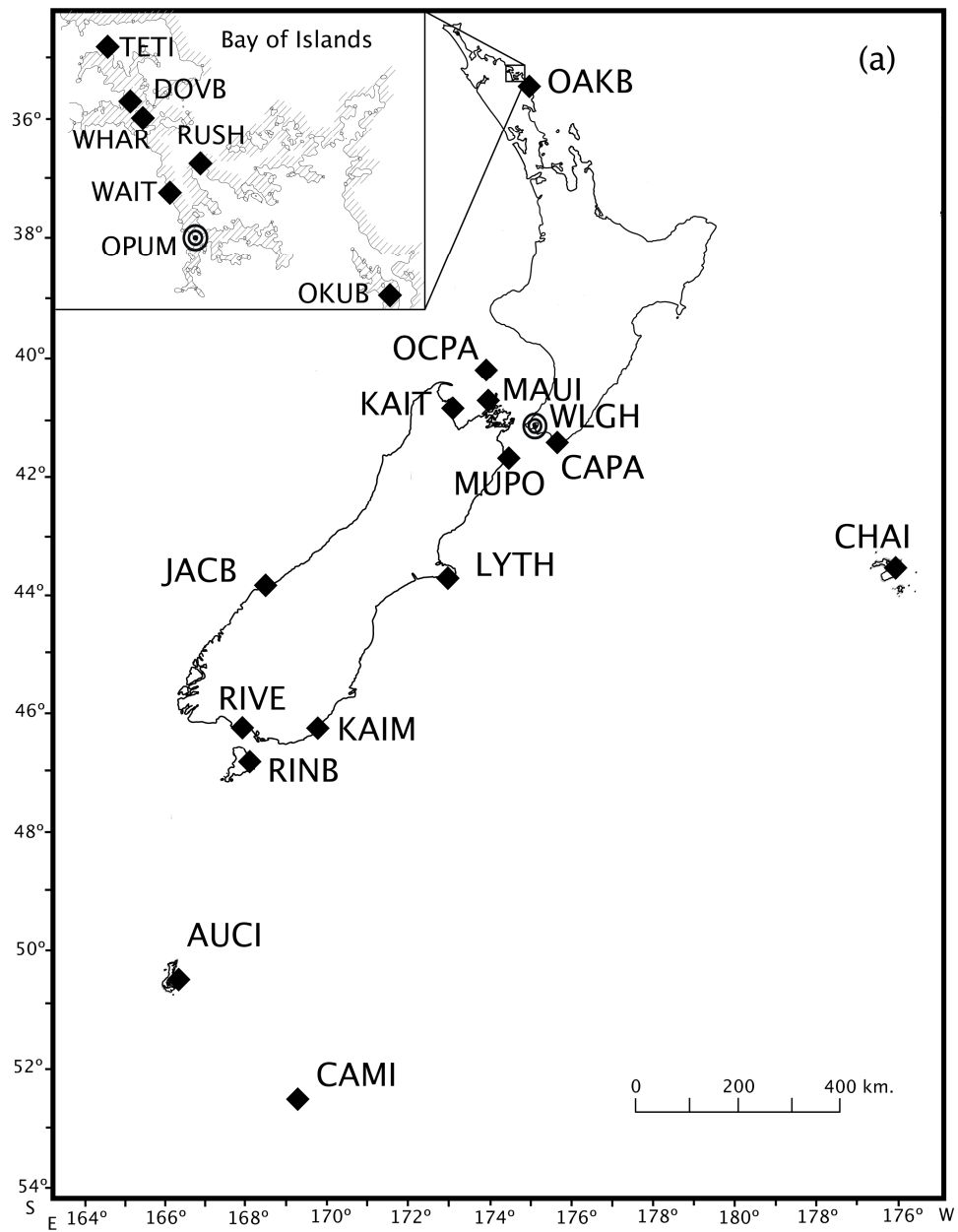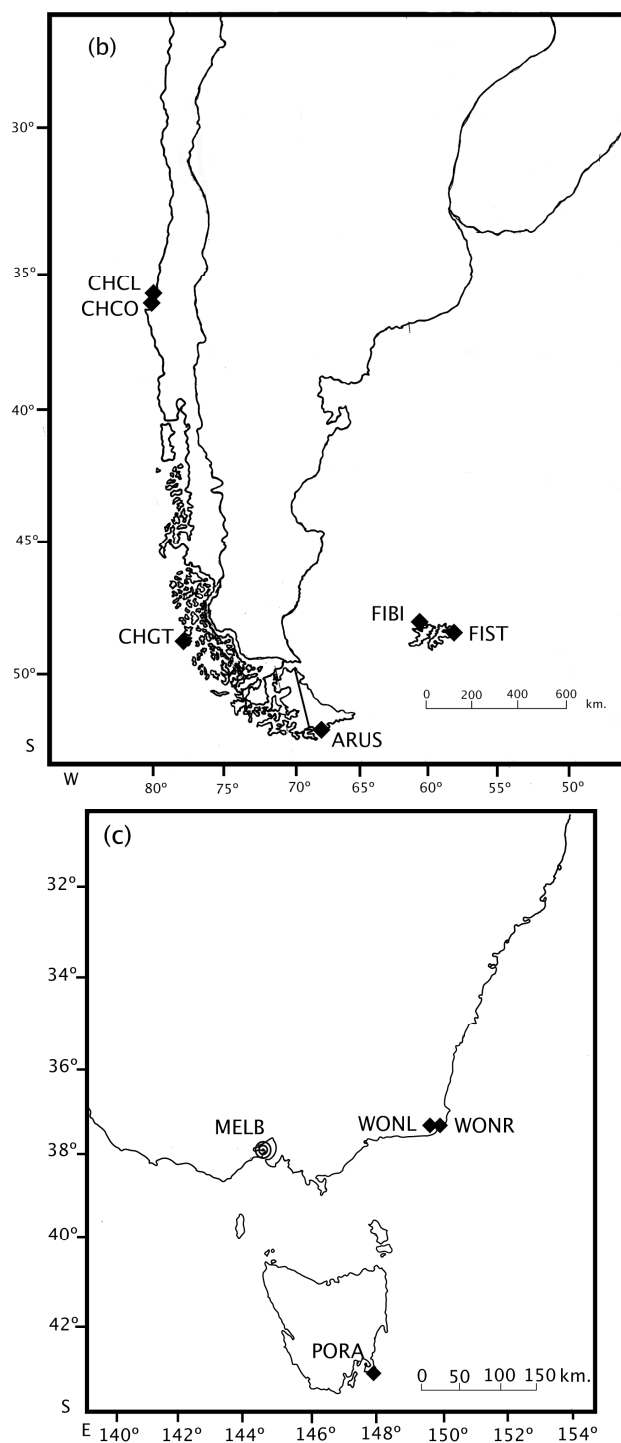
Figure 4.1 – *Mytilus* spp. Southern hemisphere sample locations.
(a) New Zealand including offshore islands, the inset is the Bay of Islands, (b) South America and (c) Australia. Codes and full place names can be found in Table 4.1. Symbol ◆ denotes non-port sample location and ◉ denotes port sample location,

Table 4.1 – Site information for samples from the current study.

COUNTRY	LOCATION	CODE	DATE (MM.YY)	SAMPLE SIZE	LAT.	LONG.
	TeTii Bay	TETI	10.05	3	35º 08.768' S	174º 00.258' E
	Dove's Bay	DOVB	10.05	2	35º 11.789' S	174º 01.860' E
	Wharau Road	WHAR	10.05	1	35° 12.382' S	174º 02.721' E
	Russell Harbour	RUSH	10.05	1	35º 15.571' S	174º 07.195' E
	Waitangi	WAIT	10.05	1	35° 16.718' S	174° 05.383' E
	Opua Marina	OPUM	12.05	1	35º 18.668' S	174º 07.208' E
	Oakura Bay	OAKB	10.05	1	35º 22.975' S	174º 20.890' E
	Okura Bay	OKUB	10.05	3	35° 02.449' S	173º 43.769' E
	Ocean Patriot	OCPA	02.08	8	40º 25.497' S	175º 06.335' E
	Maud Island	MAUI	03.00	6	41º 01.505' S	173º 53.764' E
New	Kaiteriteri	KAIT	02.01	1	41° 02.460' S	173º 01.074' E
Zealand	Wellington	WLGH	10.07	10	41º 27.000' S	174º 50.000' E
	Cape Palliser	CAPA	04.03	2	41º 36.686' S	175º 17.128' E
	Mussel Point	MUPO	01.08	10	41º 43.989' S	174º 16.052' E
	Lyttleton	LYTH	02.01	3	43º 37.152' S	172° 43.008' E
	Jackson Bay	JACB	02.03	1	43º 58.410' S	168º 36.738' E
	Chatham Islands	CHAI	12.02	13	43º 35.956' S	176º 39.939' W
	Kaimataitai	KAIM	01.04	1	46º 26.598' S	169º 48.414' E
	Riverton	RIVE	02.08	8	46° 21.833' S	168° 01.092' E
	Ringitingi Beach	RINB	01.07	16	46º 53.259' S	167º 59.907' E
	Auckland Island	AUCI	04.04	10	50º 29.391' S	166º 16.767' E
	Campbell Island	CAMI	03.04	6	52º 31.600' S	169º 06.835' E
	Concepcion	CHCO	05.03	20	36º 44.003' S	073º 07.830' W
Chile	Colchogue	CHCL	03.07	17	37º 03.442' S	073º 10.137' W
	Golfo Trinidad	CHGT	02.02	7	49º 58.557' S	075º 12.085' W
Argentina	Ushuaia	ARUS	01.02	9	54º 48.136' S	068º 15.483' W
Falkland	Bense Island	MYBI	02.02	2	51º 28.998' S	060º 30.000' W
Islands	Stanley Wharf	MYST	02.03	2	51º 41.573' S	057º 49.283' W
	Wonboyn Lake	WONL	11.04	1	37º 14.673' S	149º 55.825' E
Australia	Wonboyn River	WONR	11.04	2	37º 15.033' S	149º 57.556' E
	Melbourne	MELB	07.07	20	37º 53.532' S	144º 54.910' E
	Port Arthur	PORA	01.09	20	43º 08.489' S	147º 51.397' E
	Comox	COMO	03.06	6	49º 40.332' N	124º 56.568' E
	Otter Point	OTTP	03.06	1	48º 21.774' N	123º 47.310' E
Canada	Chemainus	CHEM	03.06	7	48º 55.560' N	123º 43.134' E
	Placentia Bay	PBAY	03.09	10	47º 38.682' N	054º 09.330' E
	Newfoundland	MYCA	02.98	6	47º 24.348' N	053º 53.600' E
Spain	Galicia	GASP	02.08	10	42º 21.510' N	008º 48.000' W
Russia	Black Sea	BLSE	04.06	13	44º 31.668' N	037º 42.300' W
Turkey	Izmir	IZMI	07.03	17	38º 25.242' N	027º 07.656' W
Great	Mundesley	GBMU	09.03	9	52º 52.776' N	001º 26.208' W
Britain	Lowestoft	GBLO	09.03	12	52º 28.404' N	001º 45.546' W
Holland	Oosterschelde	NLOO	08.04	19	51º 38.502' N	003º 50.682' W

Population codes, sample date and geographical coordinates of all locations sampled.

4.3.2 Taxonomic identification

Taxonomic status was determined for all sampled individuals either through previously published studies, the Me15/16 nuclear marker assay (Inoue et al. 1995), Me15/16 RFLP assays and/or a diagnostic 16s RFLP (Westfall et al. 2010). Most individuals were assayed with the Me 15/16 marker (241 of 308 total individuals or 78%), except *M. galloprovincialis* from Spain and Turkey and *M. edulis* from England and Holland. The location of Galicia, Spain is outside of the known *M. edulis*/*M. galloprovincialis* Western Europe hybrid zone: these mussels are assumed to be Northern *M. galloprovincialis*. Two areas of hybridisation were sampled in Canada; *M. trossulus* and the introduced *M. galloprovincialis* hybridise in SW British Columbia and *M. edulis* hybridises with *M. trossulus* in Newfoundland, and introgression is common in these areas (Rawson et al. 1999, Riginos et al. 2004). *M. trossulus* was identified by the Me15/16 marker (Inoue et al. 1995) (see Tables 4.4 and 4.6). The Atlantic Canada samples originated from two studies; the Newfoundland samples from Apte et al. (2000) were identified by the Glu-5' marker (Rawson et al. 1996) and the Placentia Bay samples provided by D. Innes (unpublished data) were identified by the Glu-5' (Rawson et al. 1996) and ITS (Heath et al. 1995) markers either as pure *M. edulis*, *M. trossulus* or hybrids. Since *M. trossulus* was only sampled in areas of known hybridisation, a pure genetic background could not be guaranteed even if classified by one or several markers, meaning *M. trossulus* sequences could not be used as a closely related monophyletic outgroup in phylogenetic analysis. What is henceforth referred to as *M. trossulus* is identified as such based on the Me15/16, Glu-5' and/or ITS marker. The *M. edulis* complex has been previously partitioned into four mitochondrial DNA lineages, herein referred to as haplogroups (Rawson & Hilbish 1998, Westfall et al. 2010), including *M. trossulus*, *M. galloprovincialis* Northern hemisphere, *M. galloprovincialis* Southern

hemisphere and a shared *M. edulis*/*M. galloprovincialis* haplogroup from the Northern hemisphere. The 16s rRNA gene RFLP designed by Westfall et al. (2010) was applied to all samples except Galicia, Oosterchelde and Placentia Bay. The taxonomic status of sibling species in the Northern hemisphere is fairly well defined but nuclear (allozyme) and mitochondrial DNA data can produce discordant phylogenetic topologies and extensive introgressive hybridisation often blurs the taxonomic lines. This study is not constrained by exact species designations only, rather, haplogroup nomenclature will be used in addition to nuclear marker identification (where possible) that is more suitable to present mitochondrial data sets. The development and distributions of haplogroups can be fully explored in Rawson & Hilbish (1995) and Westfall et al. (2010).

4.3.3 Molecular methods

Total genomic DNA was extracted from mantle and posterior adductor muscle by either the DNeasy Blood and Tissue Extraction kit (QIAGEN) or standard phenol-chloroform protocol (Hillis et al. 1995) in cases where the kit extraction produced poor quality DNA. Both methods were modified by an intermediate salt precipitation step to remove mucopolysaccharides characteristic of molluscan tissues (Sokolov 2000); 50μL saturated KCl solution was added to the Proteinase K digested tissue in lysis buffer, incubated on ice for 10 minutes, centrifuged at 13000 rpm for 10 minutes and then the aqueous layer transferred to a clean tube. A 527 bp fragment of the 16s rRNA gene was amplified by PCR using the universal primers 16sAR/16s BR (Palumbi 1996) under the following cycling conditions: 95ºC, 3'; [95ºC, 30"; 52ºC, 30"; 72ºC, 45"] x 30; 72ºC, 3'. The mantle and adductor muscle tissue types used for DNA extraction contain both male and female mitotypes involved in the doubly uniparental mode of mitochondrial DNA inheritance (Stewart et al. 1995). Because this primer pair amplifies both male and female

mitotypes, pair-wise uncorrected p-distances between individual sequences within each species/haplogroup were analyzed to identify male genes at an average of 10-20% difference. Female genes may have been "highjacked" into the male mitochondrial genome and subject to male heritability patterns whilst appearing to be female with less than 10% genetic distance compared to other female genes. Identification of high-jacked genes was not a priority as the male mitochondrial genome undergoes relaxed selection and has a significantly higher rate of mutation due to elevated meiotic events during spermatogenesis, therefore, less than 10% genetic distance indicates the gene was very recently high-jacked and resulting phylogenetic constructions would be minimally affected. After sequencing, a reliable 420 bp fragment of the 16s rRNA gene was used for phylogenetic reconstruction.

An 860 bp fragment of the cytochrome oxidase subunit III (COIII) mtDNA gene was amplified for female mitotypes using the mitotype specific primers FOR-2/REV-1 (Stewart et al. 1995) corresponding to positions 865-884 and 1326-1301, respectively, of segment 5 of the *M. edulis* female mitochondrial genome described in Hoffman et al. (1992). The PCR cycling conditions were: 95ºC, 3'; [95ºC, 30"; 56ºC 30"; 72ºC, 1'] x 30; 72º, 5'. Here the problem of highjacking was minimal, as mutations in the primer regions would accumulate relatively quickly and only very recently masculinised genes would be amplified. A reliable 669 bp fragment of the COIII gene was used for phylogenetic reconstruction.

PCR products of both genes were purified to remove excess primers and unincorporated nucleotides with ExoSAP-IT. Automatic sequencing PCR reactions were carried out with the Applied Biosystems PRISM BigDye Terminator Version 3.1 Ready Reaction Cycle Sequencing Kit using the forward primer for all individuals and an additional reverse primer sequencing reaction for a random 10% subset of individuals to

ensure accuracy and also for problem sequences that did not return the full nucleotide content specified for each gene. Sequencing reactions were run on an ABI PRISM 3730 DNA Analyzer by staff at the Allan Wilson Centre Genome Service in Palmerston North, New Zealand (http://awcmee.massey.ac.nz/genome-service.htm). Sequence chromatograms were edited with the software 4Peaks Version 1.7.2 (A. Griekspoor and Tom Groothuis, www.mekentosj.com).

16s and COIII sequences were aligned with ClustalX v2.0 (Larkin et al. 2007) using default parameters. The 16s sequences contain one gap present in the *M. trossulus* haplogroup at nucleotide 220. The COIII sequences contain no gaps or missing characters. The outgroup species *M. californianus* (collected at Otter Point, British Columbia, Canada in March 2006) was sequenced in both directions for both genes in this study (Genbank Accession HQ336052). The 16s and COIII data sets were reduced to unique haplotypes identified with the program TCS Version 1.21 (Clement et al. 2000).

To draw on the large amount of 16s data already available for *M. edulis* complex members, additional 16s sequences derived from studies done by Hilbish et al. (2000) and Gérard et al. (2008) were mined from Genbank (Table 4.2). 16s sequences from Gérard et al. (2008) were trimmed to include the first 420 bp in the alignment, this removed up to 43 bp from the 3' end of the sequences and created up to 43 bp of missing nucleotides at the 5' end. A third data set was constructed combining 16s and COIII sequences for the subset of individuals for which both data were available. Mesquite v2.73 (Maddison & Maddison 2010) was used to produce NEXUS formatted files for the 16s, COIII and combined data sets for subsequent analyses.

Table 4.2 – 16s sequences from previous studies used in the current study

Haplotype	Sample Location	Haplogroup	Genbank Accession #	Source
AU1, 2	Australia	MgS	AF179448, AF179461	Hilbish et al. 2000
NZ1, 4	New Zealand	MgS	AF179452, AF179459	Hilbish et al. 2000
CH1, 2, 3, 5, 6	Chile	MgS	AF179445, AF179443, AF179447, AF179455, AF179456	Hilbish et al. 2000
K1, 2	Kerguelen	Mg/Me	AF179449, AF179450	Hilbish et al. 2000
SH1, 2, 3	Southern hemisphere	Mg/Me and MgS	AF179444, AF179454, AF179458	Hilbish et al. 2000
MAU3	Maulin, Chile	MgS	AM905182	Gérard et al. 2008
CHL2, 3	Dichato, Chile	MgN and/or Mg/Me	AM905175, AM905176	Gérard et al. 2008
PAT3, 4	Patagonia, Chile	MgS	AM905189, AM905190	Gérard et al. 2008
KER 2	Kerguelen	MgS	AM905189	Gérard et al. 2008
AFR1, 2	Pater Noster Bay, South Africa	MgN and/or Mg/Me	AM905216, AM905217	Gérard et al. 2008
AUS2, 4	Nedland, Australia	MgN and/or Mg/Me	AM905213	Gérard et al. 2008
CBL2	Cloudy Bay Lagoon,	MgS	AM905162	Gérard et al. 2008
SIM3	Simpson's Bay, Tasmania	MgS	AM905167	Gérard et al. 2008
GRS2, 3	George Sound, NZ	MgS	AM905156, AM905157	Gérard et al. 2008
WHB3	Wellington Harbour, NZ	Unknown	AM905151	Gérard et al. 2008
DUN2	Dunedin, NZ	MgS	AM905147	Gérard et al. 2008
CHG1, 2, 3	Chioggia, Italy	MgN and/or Mg/Me	AM905220, AM905221, AM905222	Gérard et al. 2008

Haplogroup assignment based on placement in each phylogenetic tree from corresponding study. All haplotypes are *M. galloprovincialis*.

4.3.4 Partitioning

Tests of character incongruence are employed before inferring phylogenetic relationships from a combined 16s and COIII data set to evaluate the relative phylogenetic homogeneity of a partitioned data set. The incongruence length difference test (ILD, also called the partition homogeneity test) (Cunningham 1997) implemented in PAUP* v4 beta (Swofford 1998) determines if the variation of evolutionary properties of sequences between genes is greater than what would be expected by chance. Although a positive result of the ILD test indicates that data sets should not be combined and considered under a single model of nucleotide substitution, many programs overcome this by offering the use of multiple models or unlinking of parameter estimates applied to partitions within a single data set.

4.3.5 Bayesian phylogenetic inference

The program MrBayes v3.1.2 (Huelsenbeck & Ronquist 2001) is used to analyse single 16s, COIII and combined data sets in two simultaneous independent runs (for each data set) after optimisation of MC^3 (Metropolis-coupled Markov Chain Monte Carlo) operators. The model of nucleotide substitution is determined for each single gene data set (ModelTest v3.7, Posada & Crandall 1998) under the Bayesian Information Criterion (BIC) (Schwarz 1978) assumed *a priori* as the model with the greatest posterior probability. The combined data set is partitioned into two parts; the most complex model (corresponding to the BIC model determined for the COIII data set) with model parameters unlinked across partitions. Default values of uninformative (flat) prior distributions are specified *a priori*.

Convergence is assessed in two ways: (1) within each run by the effective sample size (ESS) of each parameter of interest using Tracer v1.4.1 (Rambaut & Drummond 2003) (minimum ESS=300) and, (2) between runs with the average standard deviation of clade probabilities (referred to as split frequencies) tending towards zero and the potential scale reduction factor of continuous variables tending towards 1 after convergence. After several initial runs to achieve convergence on the posterior distribution, the following MC^3 operators are applied to subsequent analyses:

	COIII	16s	Combined
Number of generations	10,000,000	5,000,000	10,000,000
Number of chains (number of heated chains is n-1)	8	8	8
Number of swaps (per generation)	2	2	2
Chain Temperature	0.05	1.0	0.05

Independent run results are summarised into a maximum clade credibility tree (total tree probability > 0.95) with nodal support (posterior probability) estimated for each clade.

4.3.6 Maximum likelihood phylogenetic inference

Only single gene data sets are analysed under the maximum likelihood criteria using the program GARLI v0.96 (Zwickl 2006) because the program does not allow data partitioning, the use of multiple models or unlinking of model parameters and the results of the ILD test indicate there is more variation in the evolutionary pattern between genes than is expected by chance. This program uses a stochastic algorithm to optimize topology, branch lengths and model parameters by applying mutations during each generation and assessing the relative "fitness" of individuals (populations of solutions) with those mutations. Individuals with the highest fitness are then used as parents for the next generation. Some parametrer values corresponding to the predetermined model (i.e. the instantaneous rate matrix) are estimated and optimised during each generation.

The *a priori* assumption of the second order corrected Akaike Information Criterion (AICc) (Sugiura 1978) is calculated by ModelTest v3.7 (Posada & Crandall 1998). The AICc is chosen *a priori* over the AIC (Akaike 1974) because they converge as sample size increases, and the AIC in general prevents over-parameterisation by penalizing for free parameters, a distinct advantage over the classic likelihood ratio test, which may demonstrate marginal improvements of likelihood values that correspond with excessive increases in the number of parameters.

Single gene data sets are analysed in GARLI v0.96 to complete four simultaneous independent analyses (from different random start trees) for the default 5,000,000 generations with 1000 non-parametric bootstrap pseudo-replicates each. A 50% majority-rule consensus tree is calculated in PAUP* from the trees produced by GARLI v0.96 during each independent run. Node values are in the form of percent support based on bootstrap pseudo-replicates.

4.3.7 Divergence dating - clock-like behaviour

Two tests are employed to determine if the 16s and COIII data sets adhere to a strict molecular clock model of evolution. The strict molecular clock assumption of a constant nucleotide substitution rate over evolutionary time does not necessarily hold when genes are subject to changes in function, duplication events, environmental variation and mechanisms of DNA damage and repair systems. Intuitively, because of the large geographical range of the *Mytilus edulis* complex taxa, spatial environmental variation and the changes in DNA damage and repair mechanisms one would expect differences in the rate of nucleotide substitution among lineages. In addition, to correctly infer the divergence date of Southern hemisphere *M. galloprovincialis*, the molecular clock assumption must be examined.

Tajima's relative rate test (Tajima 1993) is a simple non-parametric test that requires no prior model choice. Three sequences are examined, one being a known outgroup and the other two are treated as ingroups. If you let n_{ijk} be the number of sites where sequences 1, 2 and 3 have nucleotides *i*, *j* and *k*, respectively, then the assumption of the molecular clock is,

$$E(n_{ijk}) = E(n_{jik}) \text{ (Tajima 1993),} \qquad \text{Eq. 1}$$

regardless of the model of nucleotide substitution or variation of substitution rate among sites. The test statistic approximately follows a chi-square distribution with one degree of freedom and a strict molecular clock can be assumed if the equality in Eq. 1 holds (Tajima 1993). A consensus sequence is generated for each haplogroup/species in DAMBE (Xia & Xie 2000) and haplotypes that are not assayed by markers (i.e. haplogroup assignment is based on placement in the phylogenetic tree) are removed from the data set. Pairwise Tajima's relative rate tests are then conducted in MEGA v4 (Tamura et al. 2007) among haplogroups using *M. californianus* as the outgroup, with the exception of the *M. edulis*/*M.*

galloprovincialis haplogroup divided into two groups based on species. The advantage of using a relative rate test is to determine what branches of the tree are evolving faster or slower than other branches, in essence a local molecular clock test.

The global parametric likelihood ratio test is used to test the hypothesis of a strict molecular clock over the entire tree. The null hypothesis is all branches evolve at a uniform rate and branch lengths are constrained so the tips line up on one axis. The alternative hypothesis allows the branch lengths to vary independently and tips are not constrained to one axis. The test statistic is,

$$-2logL = 2(logL_0 - logL_1) \qquad\qquad \text{Eq. 2}$$

where L_0 and L_1 are the likelihood scores under the null and alternative hypotheses respectively. The test statistic follows a chi-square distribution where the number of sequences minus two gives the degrees of freedom. The likelihood scores are calculated in PAUP* using model specifications as determined by ModelTest under the AICc (Posada & Crandall 1995) for both 16s and COIII data sets.

4.3.8 Divergence dating

The coalescent-based software mDIV (Nielson & Wakely 2001) is used to estimate the date of divergence for the Southern hemisphere *M. galloprovincialis* lineage. An HKY (Hasegawa-Kishino-Yano) (Hasegawa et al. 1985) model of nucleotide substitution, a variation of the GTR model, was the basis for the coalescent estimates of divergence. The program uses MCMC approximation method to estimate the time since divergence for two populations descended from a single panmictic ancestral population with or without subsequent migration (Nielson & Wakeley 2001). Estimates of divergence time were repeated on three individual runs each with 2,000,000 generations and a burn-in of 25% (500,000 generations), the only variable among runs being the random starting number.

This estimation is performed within the coalescent framework, which assumes ancestral sequences are still present in the population.

The maximum migration rate parameter was set to zero for two reasons: (1) the number of free model parameters is greatly reduced with no expected migration, leading to better mixing and convergence of sampled posterior likelihoods and probabilities and (2) the migration rate between hemispheres is so close to zero due to physiological limitations of larval dispersal, that it can be considered negligible. From a biological stand point, migration between hemispheres means larval planktonic survival over a wide range of water temperatures, large geographical distances and presumably long periods of time. Although migration events of this magnitude are possible (as illustrated by the presence of *M. galloprovincialis* in the Southern hemisphere), they are rare enough to allow inference of a migration rate that is effectively zero.

The maximum time since divergence is set to 5 million years as this would pre-date the earliest estimate of *Mytilus* presence in the Atlantic Ocean at 3.5 mya (Vermeij 1991) and the conservative estimate of the Bering Seaway opening at 4.8 mya (Marincovich & Gladenkov 1998), corresponding to the Atlantic migration hypothesis (Hilbish et al. 2000). Branching order on the COIII single gene and combined genes phylogenetic tree obtained in the current study is clear in that the Southern *M. galloprovincialis* lineage and the *M. galloprovincialis*/*M. edulis* haplogroup are sister groups diverging from a common ancestor with the *M. galloprovincialis* Northern haplogroup, therefore the divergence of the Southern lineage cannot predate the divergence of the *M. galloprovincialis*/*M. edulis* haplogroup. In addition, there is consensus among previously published phylogenies, with the exception of Gérard et al.'s (2008) COI phylogeny, that the Southern hemisphere *M. galloprovincialis* lineage did not diverge before *M. edulis* and would therefore be less than 5 million years old.

In contrast to the coalescent-based model of population differentiation used in mDIV, the divergence time of the Southern hemisphere *M. galloprovincialis* lineage is also estimated under a birth-death model of speciation in BEAST v1.5beta2 (Drummond & Rambaut 2007). It is unknown whether a birth-death model of speciation or a coalescent model approach is the most suitable for dating the divergence of the Southern hemisphere *M. galloprovincialis* lineage, therefore, an estimate of absolute divergence time may be considered reliable if there is concordance among estimates obtained from both models. The birth-death model of speciation employed in BEAST is derived from a Yule model of speciation, a linear model describing a simple birth-death process characterised by an average lineage birth rate known as the Yule process (Yule 1925, Drummond & Rambaut 2007).   A data set of combined 16s and COIII genes (for a subset of individuals where both genetic sequences were available) is used for input into BEAST. The models of nucleotide substitution for the partitioned data set are unlinked, the 16s partition defined by an HKY model of nucleotide substitution and the COIII partition defined by a GTR model of nucleotide substitution as defined earlier by ModelTest. A relaxed molecular clock model assuming independent rates of evolution on different branches with an uncorrelated (i.e. parent and daughter lineage rates are not correlated) lognormal distribution is used for the COIII partition and a strict molecular clock model was used for the 16s partition. Haplogroups are defined as monophyletic taxon sets (in congruence with Bayesian phylogeny haplotype distribution) and a calibrated normal distribution is defined for the time to most recent common ancestor (T_{MRCA}) of the *M. galloprovincialis* 'Mediterranean' haplogroup (refer to branching order of combined loci Bayesian tree) at a midpoint 5.95 (SD=1.15) my following the divergence time estimated by the opening of the Bering Strait (Marinkovich & Gladenkov 1998). The starting tree for all runs was user-defined based on the topology obtained from Bayesian analysis (Drummond et al. 2006). An initial BEAST

run is performed to optimize the operator tuning parameters, which are then used as starting points within a uniform distribution for a subsequent 3 independent runs of 10,000,000 MCMC generations each. Assessments of chain mixing and convergence on the posterior distribution of each run are conducted in Tracer v1.4.1 (Rambaut & Drummond 2003) with a burn-in value of 25% and a run is considered successful if the effective sampling size (Drummond & Rambaut 2007) was greater than 300. A majority-rule consensus tree from all runs is constructed in PAUP and then probability values for bipartitions assigned using TreeAnnotator v1.5beta (Drummond & Rambaut 2007).

4.3.9 Hypothesis testing of topological constraints

The topologies derived from this particular data set are tested against topologies previously reported for the COI (Gérard et al, 2008) and 16s (Hilbish et al. 2000; Gérard et al. 2008) genes to determine if discordant topologies derived from different genes and inference methods are significantly different from each other. Three previously reported competing topologies are reduced to their backbone topologies (Figure 4.2), which serve as input for the constraints to be tested against the trees constructed in this study using maximum likelihood and Bayesian inference. The phylogenetic trees published by Gérard et al. (2008) use *M. trossulus* as an outgroup. In the current study, *M. trossulus* sequences for 16s and COIII genes (single gene topologies) did not form a monophyletic basal clade in the tree (with the exception of the ML-16s tree) therefore *M. californianus* is used as an outgroup in the constraining backbone topologies. Although these constraints are applied to the ML-16s tree, the addition of a clade in between *M. californianus* and the earliest *M. edulis/M. galloprovincialis* clade does not affect the outcome of the analysis because only a relative topology is enforced and taxa may be added along the tree without penalty as long as the backbone topology is maintained. The goal of hypothesis testing is to

determine which topology is the most likely to be correct to aid inference of the evolutionary history of the Southern hemisphere *M. galloprovincialis* lineage.
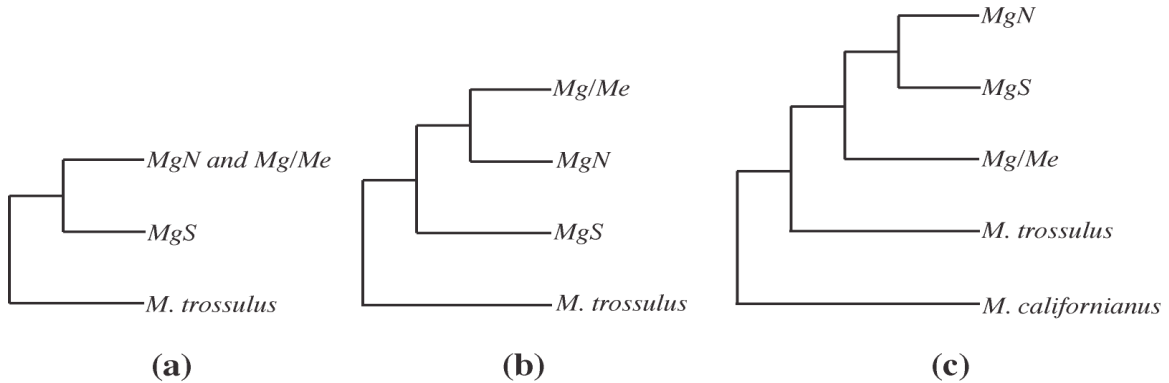
Figure 4.2 – Backbone topologies describing evolutionary relationships of the *Mytilus edulis* species complex. (a) COI phylogeny from Gérard et al (2008), (b) 16s topology from Gérard et al (2008), and (c) 16s topology from Hilbish et al. (2000). MgS = *M. galloprovincialis* Southern haplogroup, MgN = *M. galloprovincialis* Northern haplogroup, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern haplogroup.

Within a maximum likelihood framework of phylogenetic inference, competing topologies have most widely been compared with the non-parametric Kishino-Hasegawa test (KH test) (Kishino et al. 1990), but some authors have contested the use of this test for topologies derived *a posteriori* (i.e. trees produced from a data set) (Goldman et al. 2000, Buckley 2002). The Shimodaira-Hasegawa test (SH test) (Shimodaira & Hasegawa 1999), similar to the KH test, is applicable when *a priori* and *a posteriori* topologies are used (Goldman et al. 2000, Buckley 2002). This investigation uses the Shimodaira-Hasegawa tests in the ML framework.

Bayesian methods can be advantageous over ML methods as there is no re-optimisation of likelihood with multiple bootstrap replicates, rather the posterior distribution is sampled with the MCMC (Buckley 2002). In addition, if different models are used to generate the trees, these models do not have to be nested (Kass & Raftery 1995). The approach uses a Bayes Factor (BF) to evaluate the evidence in favour of a topological theory (Kass & Raftery 1995), calculated as the ratio of marginal likelihoods

derived under separate hypotheses (the BIC gives an approximation of the logarithm of the BF), which lends itself to evolutionary model testing. Since this investigation is interested in testing topologies, the marginal likelihoods can only be approximated through the MCMC sampling of the harmonic mean of the likelihoods derived from trees constructed with and without the topological constraint (termed B_1 and B_0, respectively) (Ronquist et al. 2005), and the ratio of these marginal likelihoods becomes the BF. Kass & Raftery (1995) proposed a table for interpreting the BF in a scientifically meaningful way, based on original work by Jeffreys (1961, in Kass & Raftery 1995) to interpret in half units on the $\log_{10}$ scale. The scale below is a derivative of the original, transformed by doubling the natural logarithm of the BF to mimic the scale of likelihood ratio test statistics (where B_{10} is the BF calculated by the ratio of B_1 to B_0), and should be interpreted as a rough guide based on previous investigations (Kass & Raftery 1995):

$2 \log_e(B_{10})$	Evidence against H_0
0 to 2	Not worth a mention
2 to 6	Positive
6 to 10	Strong
>10	Very Strong

4.3.10 Shimodaira-Hasegawa test – ML trees

To perform topological constraint testing on the ML COIII and 16s trees, first a modified nexus file for each gene (i.e. not combined) is used as input for independent GARLI runs. An initial run uses the modified nexus file to create an unconstrained baseline tree resampled by 1000 bootstrap pseudo-replicates for each gene and two subsequent runs are performed on each gene using the topological constraints derived from Gérard et al. (2008) and Hilbish et al. (2000) (Figure 4.2), each run resampled with 1000 bootstrap replicates. The bootstrap replicates are used to create 50% majority-rule consensus trees for all GARLI runs in PAUP. Each Shimodaira-Hasegawa test was based

on the 50% majority-rule consensus tree for the normal and constrained topologies. The tests are conducted in a pairwise fashion comparing the maximum likelihood topology (T_{ML}) derived from GARLI to each previously published topology (T_0). The test distribution is generated by a fully optimised model version of the resampling estimated log-likelihood (RELL) method described by Kishino et al. (1990) with 1000 bootstrap replicates. The topological constraints are treated as "backbone" (Figure 4.2) to reduce the restrictive properties of a monophyletic constraint. The general hypothesis of the SH test is applied to test how well the published topologies from Gérard et al (2008) and Hilbish et al (2000) fit the current data set.

4.3.11 Bayes Factors – Bayesian Inference

Hypothesis testing using topological constraints is conducted in MrBayes. The form of topological constraints in MrBayes is limited to hypotheses of monophyly, therefore monophyletic groupings are assigned based on topological positions in the published trees (Figure 4.2) and follow the haplogroup distinctions laid out originally by Rawson & Hilbish (1995) and redefined by Westfall et al. (2010). For the purposes of hypothesis testing, haplotypes of putatively identified *M. trossulus* were removed from the COIII data set because they did not form a monophyletic grouping in the BI-COIII tree, and the placement of *M. trossulus* in the phylogenetic tree is not of primary interest for this investigation. With the modified data set, the original analysis is repeated using identical model and MCMC parameters to obtain directly comparable values for posterior probabilities and marginal likelihoods, and the topological constraints are applied to this secondary analysis. The 16s data set retains the *M. trossulus* haplotypes for hypothesis testing because they did not violate the assumption of monophyly. Two hypotheses are tested against the BI-COIII tree: Hypothesis 1 - the monophyletic *M. galloprovincialis*

116

Southern lineage diverged before the Northern lineages *M. galloprovincialis* and *M. galloprovincialis/M. edulis* diverged, therefore the MRCA to all three *M. galloprovincialis* and *M. edulis* lineages is descended from *M. trossulus* (Figure 4.2a) (Gérard et al. 2008). The constraint for this hypothesis is; haplogroups *M. galloprovincialis* Northern and *M. galloprovincialis/M. edulis* Northern form one monophyletic group and *M. galloprovincialis* Southern forms another monophyletic group (Figure 4.2a), consequently this is the same hypothesis that is reflected in the COI topology of Gérard et al. (2008) (Figure 4.2b). Hypothesis 2 - derived from Hilbish et al. (2000) for 16s rRNA gene topology, infers that the *M. galloprovincialis* Southern and Northern lineages are sister groups that descended from their most recent common ancestor with the *M. galloprovincialis/M. edulis* haplogroup (Figure 4.2c) (Hilbish et al. 2000), therefore the constraint for this hypothesis is; *M. galloprovincialis* Northern and Southern lineages form a monophyletic group (Figure 4.2c).

4.4 Results

4.4.1 Sequencing

A total of 308 individuals mussels have sequence data collected, 291 individuals for the COIII gene generating 173 unique female haplotypes (Tables 4.3 and 4.4) and 128 individuals for the 16s gene generating 25 unique female haplotypes Table 4.6. The distribution of unique haplotypes from each country including their taxonomic identification is in Tables 4.3 and 4.4 for the COIII gene and Table 4.6 for the 16s gene. Seven haplotypes are detected in more than one country within the Northern or Southern hemisphere and five haplotypes are detected worldwide; the distributions of these haplotypes can be found in Table 4.5 for the COIII gene and in Table 4.7 for the 16s gene. The taxonomic identity (haplogroup) for each haplotype reported in the above mentioned

117

tables is derived from either the diagnostic 16s RFLP assay, placement within the phylogenetic tree (BI), or a combination of both methods. It's important to iterate the classification of the *M. galloprovincialis* Southern haplogroup is based on both nuclear and mitochondrial genes (Hilbish et al. 2000, Westfall & Gardner in press). Haplotype nomenclature is composed of three parts: the first part is a one-letter species identifier (*M. galloprovincialis* is g, *M. edulis* is e, and *M. trossulus* is t), the second part is a two-letter location identifier based on where the haplotype was identified:

Code	Area
GL	Global
SH	Southern hemisphere
NH	Northern hemisphere
SA	South America
EU	Europe
AC	Atlantic Canada
PC	Pacific Canada

The individual country codes can be interpreted from Tables 4.3 and 4.4. In cases where the haplotype was shared among species (i.e. within the same haplogroup or for putative hybrids), the first single letter code becomes two letters to denote each species. The third part of the haplotype nomenclature identifies what haplotype sequence it is that fulfills the requirements of the first 3 (or 4) letters. For example, gNZ 1 is the first haplotype that belongs to *M. galloprovincialis* and is only found in New Zealand.

Table 4.3 – Haplotype names, taxonomic identities, distributions and Genbank accession numbers for the COIII gene for Southern hemisphere locations.

Country	Region(s)	Haplotype	Taxonomic Identity	Genbank Accession #
New Zealand	RINB (3), CHAI (4), AUCI	gNZ 1	MgS*†	HQ336119
	RINB	gNZ 2	MgS†	HQ336120
	RINB	gNZ 3	MgS†	HQ336121
	RINB, CHAI (3)	gNZ 4	MgS*†	HQ336122
	RINB	gNZ 5	MgS†	HQ336123
	RINB (6), CAMI (3), RIVE, WLGH (4)	gNZ 6	MgS*†	HQ336124
	RINB	gNZ 7	MgS†	HQ336125
	RINB	gNZ 8	MgS†	HQ336126
	RINB	gNZ 9	MgS†	HQ336127
	MUPO, RIVE	gNZ 10	MgS*†	HQ336128
	MUPO	gNZ 12	MgS*†	HQ336129
	MUPO	gNZ 13	Mg/Me*†	HQ336130
	MUPO	gNZ 14	MgS*†	HQ336131
	CHAI	gNZ 15	MgS*†	HQ336132
	CHAI (2)	gNZ 16	MgS*†	HQ336133
	CHAI	gNZ 17	MgS*†	HQ336134
	CHAI	gNZ 18	MgS*†	HQ336135
	AUCI	gNZ 19	MgS*†	HQ336136
	AUCI	gNZ 20	MgS*†	HQ336137
	AUCI	gNZ 22	MgS*†	HQ336139
	AUCI	gNZ 23	MgN†	HQ336140
	AUCI	gNZ 24	MgN†	HQ336141
	AUCI	gNZ 25	MgN†	HQ336142
	CAMI	gNZ 26	MgS*†	HQ336143
	CAMI	gNZ 27	MgS*†	HQ336144
	CAMI	gNZ 28	MgS*†	HQ336145
	RIVE	gNZ 29	MgS†	HQ336146
	RIVE (2)	gNZ 30	MgN†	HQ336147
	RIVE	gNZ 31	MgS†	HQ336148
	RIVE	gNZ 32	MgS†	HQ336149
	MAUI	gNZ 33	MgS*†	HQ336150
	WLGH	gNZ 35	MgS*†	HQ336151
	WLGH	gNZ 36	MgS*†	HQ336152
	WLGH	gNZ 39	MgS*†	HQ336154
	OCPA	gNZ 40	Mg/Me†	HQ336155
	OCPA	gNZ 41	Mg/Me†	HQ336156
	OCPA	gNZ 42	Mg/Me†	HQ336157
	OCPA	gNZ 45	Mg/Me†	HQ336158
	TETI (2), DOVB	gNZ 46	MgS†	HQ336159
	TETI	gNZ 47	MgS*†	HQ336160
	OPUM	gNZ 48	MgS*†	HQ336161
	RUSH	gNZ 49	MgS*†	HQ336162
TOTAL n=42				
Australia	MELB, PORA	gAU 2	MgN*†	HQ336163
	MELB	gAU 3	MgS*†	HQ336164
	MELB	gAU 4	MgS*†	HQ336165
	MELB	gAU 5	MgS*†	HQ336166
	MELB	gAU 6	MgS*†	HQ336167
	MELB	gAU 7	MgS*†	HQ336168
	MELB	gAU 8	MgS*†	HQ336169
	MELB	gAU 10	MgS*†	HQ336170
	MELB	gAU 11	MgS*†	HQ336171
	MELB	gAU 12	MgS*†	HQ336172

Country	Region(s)	Haplotype	Taxonomic Identity	Genbank Accession #
	MELB	gAU 13	MgS*†	HQ336173
	MELB	gAU 14	MgS*†	HQ336174
	MELB	gAU 15	MgS*†	HQ336175
	MELB	gAU 16	MgS*†	HQ336176
	MELB	gAU 17	Mg/Me†	HQ336177
	PORA (6)	gAU 18	MgS*†	HQ336178
	PORA (2)	gAU 19	MgS*†	HQ336179
	PORA	gAU 20	Mg/Me†	HQ336180
	PORA	gAU 21	MgS*†	HQ336181
	PORA (2)	gAU 22	MgS*†	HQ336182
	PORA	gAU 23	MgS*†	HQ336183
	PORA	gAU 24	MgS*†	HQ336184
	PORA	gAU 25	MgS*†	HQ336185
	PORA	gAU 26	MgS*†	HQ336186
	PORA	gAU 27	MgS*†	HQ336187
	PORA	gAU 28	MgS*†	HQ336188
	WONR	gAU 29	MgS†	HQ336189
	WONL	gAU 30	Mg/Me†	HQ336190
TOTAL n=28				
Argentina	ARUS	gAR 1	MgS*†	HQ336191
	ARUS (2)	gAR 2	MgS*†	HQ336192
	ARUS	gAR 3	MgS*†	HQ336193
	ARUS	gAR 4	MgS*†	HQ336194
	ARUS	gAR 5	MgS*†	HQ336196
TOTAL n=5				
Chile	CHCO (2)	gCH 1	MgS*†	HQ336198
	CHCO (2)	gCH 2	MgS*†	HQ336199
	CHCO, CHGT	gCH 3	MgS*†	HQ336200
	CHCO	gCH 4	MgS*†	HQ336201
	CHCO	gCH 5	MgS*†	HQ336202
	CHCO	gCH 6	MgS*†	HQ336203
	CHCO	gCH 7	MgS*†	HQ336204
	CHCO	gCH 8	MgS*†	HQ336205
	CHCO (2), CHCL (2)	gCH 11	Mg/Me*†	HQ336206
	CHCO	gCH 12	Mg/Me*†	HQ336207
	CHCO	gCH 14	MgS*†	HQ336208
	CHGT	gCH 15	MgS*†	HQ336209
	CHGT	gCH 16	MgS*†	HQ336210
	CHGT	gCH 17	MgS*†	HQ336211
	CHGT (2)	gCH 18	MgS*†	HQ336212
	CHCL	gCH 19	MgN†	HQ336213
	CHCL (2)	gCH 20	MgN*†	HQ336214
	CHCL	gCH 22	Mg/Me†	HQ336215
TOTAL n=18				
Falkland Islands	FIST	gFI 1	MgS*†	HQ336216
	FIST	gFI 2	MgS*†	HQ336217
	FIBI (2)	gFI 3	Mg/Me†	HQ336218
TOTAL n=3				

These are haplotypes unique to a single country. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Total number of haplotypes per country are provided. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree.

MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

Table 4.4 – Haplotype names, taxonomic identities, distributions and Genbank accession numbers for the COIII gene for Northern hemisphere locations.

Country	Regions	Haplotype	Taxonomic Identity	Genbank Accession #
Spain	GASP	gSP_1	Mg/Me†	HQ336053
	GASP	gSP_2	Mg/Me†	HQ336054
	GASP	gSP_4	Mg/Me†	HQ336056
	GASP	gSP_5	Mg/Me†	HQ336057
	GASP	gSP_7	Mg/Me†	HQ336059
	GASP	gSP_9	MgN†	HQ336061
	GASP	gSP_10	Mg/Me†	HQ336062
TOTAL n=7				
Russia	BLSE	gRU_1	Mg/Me*†	HQ336063
	BLSE	gRU_2	MgN*†	HQ336064
	BLSE (6)	gRU_3	Mg/Me*†	HQ336065
	BLSE	gRU_4	MgN*†	HQ336066
	BLSE	gRU_5	Mg/Me*†	HQ336067
	BLSE	gRU_6	Mg/Me*†	HQ336068
	BLSE	gRU_7	Mg/Me*†	HQ336069
TOTAL n=7				
Turkey	IZMI	gTK_3	MgN*†	HQ336072
	IZMI	gTK_4	MgN*†	HQ336073
	IZMI	gTK_5	Mg/Me*†	HQ336074
	IZMI	gTK_6	Mg/Me*†	HQ336075
TOTAL n=4				
England	GBLO	eGB_1	Mg/Me*†	HQ336076
	GBLO	eGB_2	Mg/Me*†	HQ336077
	GBLO	eGB_3	Mg/Me*†	HQ336078
	GBLO	eGB_5	Mg/Me*†	HQ336080
	GBLO	eGB_6	Mg/Me*†	HQ336081
	GBLO	eGB_7	Mg/Me*†	HQ336082
	GBLO	eGB_8	Mg/Me*†	HQ336083
	GBMU	eGB_9	Mg/Me*†	HQ336084
	GBMU	eGB_10	Mg/Me*†	HQ336085
	GBMU	eGB_11	Mg/Me*†	HQ336086
	GBMU	eGB_12	Mg/Me*†	HQ336087
	GBMU	eGB_13	Mg/Me*†	HQ336088
TOTAL n=12				
Holland	NLOO	eNL_1	Mg/Me*†	HQ336089
	NLOO	eNL_2	Mg/Me*†	HQ336090
	NLOO (4)	eNL_3	Mg/Me*†	HQ336091
	NLOO	eNL_4	Mg/Me*†	HQ336092
	NLOO (3)	eNL_5	Mg/Me*†	HQ336093
	NLOO (2)	eNL_6	Mg/Me*†	HQ336094
	NLOO	eNL_7	Mg/Me*†	HQ336095
	NLOO	eNL_8	Mg/Me*†	HQ336096
	NLOO	eNL_10	Mg/Me*†	HQ336098
	NLOO	eNL_12	Mg/Me*†	HQ336099
TOTAL n=10				
Canada	PBAY	eCA_1	Mg/Me†	HQ336100

Country	Regions	Haplotype	Taxonomic Identity	Genbank Accession #
	PBAY	eCA_2	Mg/Me†	HQ336101
	PBAY	eCA_4	Mg/Me†	HQ336102
	PBAY	eCA_5	Mg/Me†	HQ336103
	PBAY	eCA_6	Mg/Me†	HQ336104
	PBAY	eCA_7	Mg/Me†	HQ336105
	COMO	tCA_2	Mt*	HQ336107
	CHEM	tCA_4	Mt*	HQ336108
	CHEM	tCA_6	Mt*†	HQ336110
	CHEM	tCA_7	Mt*†	HQ336111
	CHEM	tCA_8	Mt*/Me†	HQ336112
	MYCA	tCA_9	Mt*/Me†	HQ336113
	MYCA	tCA_10	Mt*/Me†	HQ336114
	MYCA	tCA_11	Mt*	HQ336115
	MYCA	tCA_13	Mt†	HQ336116
	MYCA	tCA_14	Mt†	HQ336117
	PBAY	tCA_15	Mt*/Me†	HQ336118
	MYCA, PBAY (2)	etCA_1	Mt*/Me†	HQ336106
TOTAL n=18				

These represent haplotypes unique to a single country. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Total number of haplotypes per country are provided. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

Table 4.5 – Haplotype names, taxonomic identities, distributions and Genbank accession numbers for the COIII gene.

Haplotype	Countries (Regions)	Taxonomic Identity	Genbank Accession #
gSH_2	New Zealand (AUCI) (2) Chile (CHCO)	MgS*†	HQ336138
gSH_3	New Zealand (MAUI) Chile (CHCL, CHCO)	MgS*†	HQ336153
gSA_1	Argentina (ARUS) Chile (CHCO)	MgS*†	HQ336195
gSA_2	Argentina (ARUS) Chile (CHGT)	MgS*†	HQ336197
geEU_1	Spain (GASP) England (GBLO)	Mg/Me*†	HQ336055
eEU_1	England (GBLO) Holland (NLOO)	Mg/Me*†	HQ336079
etNH_1	Holland (NLOO) Canada (COMO, CHEM)	Mg/Me*†	HQ336097
gGL_1	New Zealand (WLGH, OCPA) Australia (MELB) Chile (CHCO, CHCL) Spain (GASP)	MgN*†	HQ336060
gGL_2	New Zealand (MUPO, MAUI, WLGH, OKUB)	Mg/Me*†	HQ336070

Haplotype	Countries (Regions)	Taxonomic Identity	Genbank Accession #
	Australia (MELB, PORA, WONR) Chile (CHCL) Turkey (IZMI)		
gtGL_1	New Zealand (MAUI) Chile (CHCO, CHCL) Turkey (IZMI) Canada (COMO, CHEM, OTTP)	MgN*†	HQ336071
gtGL_2	New Zealand (OCPA) Chile (CHCL) Canada (CHEM)	MgN†	HQ336109
geGL_1	New Zealand (OCPA) Spain (GASP) Holland (NLOO)	Mg/Me†	HQ336058

These represent haplotypes that are identified in more than one country and/or hemisphere. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

Table 4.6 – Haplotype names, taxonomic identities, distributions and Genbank accession numbers for the 16s gene.

Country	Regions	Haplotype	Taxonomic Identity	Genbank Accession #
New Zealand	OKUB	gNZ 1	MgN*	GQ455380
	MAUI	gNZ 3	MgS*†	GQ455382
	CHAI	gNZ 4	MgS*†	GQ455383
	TETI, RUSH, OAKB, OKUB	gNZ 5	MgN*†	GQ455384
	MAUI	gNZ 6	MgS*†	GQ455385
	LYTH	gNZ 7	MgS*†	GQ455386
	RINB	gNZ 8	MgS*†	GQ455387
TOTAL n=7				
Chile	CHCO	gCH 1	MgS*†	GQ455388
	CHCO	gCH 2	MgS*†	GQ455389
	CHCO (3)	gCH 3	Mg/Me*†	GQ455390
	CHCO (7)	gCH 4	MgS*†	GQ455391
TOTAL n=4				
Falkland Islands	FIBI	gFI 1	Mg/Me*†	GQ455392
	FIBI	gFI 2	Mg/Me*†	GQ455393
TOTAL n=2				
Russia	BLSE (6)	gBS 1	Mg/Me*†	GQ455396
TOTAL n=1				
England	GBLO	eGB 2	Mg/Me*†	GQ455405
TOTAL n=1				
Canada	COMO	tPC 1	Mt*†	GQ455400
	COMO	tPC 2	Mt*	GQ455401
	CHEM	tPC 3	Mt*†	GQ455402
	CHEM (3), COMO (2)	tPC 4	Mt*†	GQ455403
	MYCA (2)	tAC 2	Mt*†	GQ455404

Country	Regions	Haplotype	Taxonomic Identity	Genbank Accession #
TOTAL n=5				

These represent haplotypes unique to a single country. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Total number of haplotypes per country are provided. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

Table 4.7 – Haplotype names, taxonomic identities, distributions and Genbank accession numbers for the 16s gene.

Haplotype	Countries (Regions)	Taxonomic Identity	Genbank Accession #
gSH1	New Zealand (MUPO (5), RINB (3), LYTH (2), CAPA (2), AUCI (4), CAMI (6), CHAI (3), JACB, KAIT, KAIM, DOVB, WHAR, OKUB, OPUM, WAIT) Chile (CHGT) (5) Argentina (ARUS) (4) Falkland Islands (FIST) (2)	MgS*†	GQ455395
gSA 1	Chile (CHCO, CHGT (2)) Argentina (ARUS)	MgS*†	GQ455394
etNH_1	England (GBLO (3), GBMU (7)) Canada (MYCA) (3)	Mg/Me*†	GQ455399
gGL_1	New Zealand (MUPO (5), MAUI) Turkey (IZMI) (7) Canada (CHEM)	Mg/Me*†	GQ455397
gGL_2	New Zealand (MAUI) Chile (CHCO) (2) Turkey (IZMI) Russia (BLSE) (3) Canada (CHEM)	MgN*†	GQ455398

These represent haplotypes that are identified in more than one country and/or hemisphere. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

4.4.2 COIII and 16s combination

The incongruence length difference (or partition homogeneity) test for assessing variation of evolutionary properties within and between the single COIII and 16s gene data sets favours the alternative hypothesis that there is more variation between partitions than what would be expected by chance ($p<0.05$) meaning the data sets can not be analysed under a single model of nucleotide substitution. This does not affect the Bayesian inference because MrBayes allows inference based on mixed models and separate estimation of parameters.

A combined data set of the COIII and 16s genes yields 72 unique biallelic haplotypes and represents all haplogroup and species categories. Haplotype identifiers are an amalgamation of the 16s and COIII individual identifiers (in that order). As outlined for the single gene data sets, haplotype names, taxonomic identity and distribution of individual combined haplotypes can be found in Tables 4.8 – 4.10.

Table 4.8 – Haplotype names, taxonomic identities and distributions for the combined data set for Southern hemisphere locations.

Country	Region	Haplotype (16s COIII)	Taxonomic Identity
New Zealand	RINB, CHAI, AUCI	gSH2 gNZ1	MgS*†
	RINB	gNZ8 gNZ24	MgS*†
	RINB (2), CAMI (3)	gSH2 gNZ6	MgS*†
	MUPO	gSH2 gNZ10	MgS*†
	MUPO	gSH2 gNZ12	MgS*†
	MUPO	gGL1 gNZ13	Mg/Me*†
	MUPO	gSH2 gNZ14	MgS*†
	MAUI	gNZ3 gNZ33	MgS*†
	TETI, DOVB, RUSH	gNZ5 gNZ46	MgN*
	OKUB	gNZ1 gGL2	MgN*†
	OPUM	gSH2 gNZ48	MgS*†
	CAMI	gSH2 gNZ26	MgS*†
	CAMI	gSH2 gNZ27	MgS*†
	CAMI	gSH2 gNZ28	MgS*†
	CHAI	gNZ4 gNZ1	MgS*†
	CHAI (2)	gSH2 gNZ4	MgS*†
	AUCI	gSH2 gNZ20	MgS*†
TOTAL n=17			
Chile	CHCO	gCH1 gCH1	MgS*†
	CHCO	gCH4 gCH1	MgS*†
	CHCO	gCH4 gCH3	MgS*†

Country	Region	Haplotype (16s COIII)	Taxonomic Identity
	CHCO	gCH4 gCH4	MgS*†
	CHCO	gCH4 gCH5	MgS*†
	CHCO	gCH4 gSH2	MgS*†
	CHCO (2)	gCH3 gCH11	Mg/Me*†
	CHCO	gCH3 gCH12	Mg/Me*†
	CHCO	gCH2 gSA1	MgS*†
	CHCO	gCH4 gCH14	MgS*†
	CHCO	gCH4 gCH2	MgS*†
	CHGT	gSH2 gCH15	MgS*†
	CHGT	gSH2 gCH16	MgS*†
	CHGT	gSA1 gCH17	MgS*†
	CHGT (2)	gSH2 gCH18	MgS*†
	CHGT	gSH2 gCH3	MgS*†
TOTAL n=17			
Argentina	ARUS	gSH2 gAR1	MgS*†
	ARUS	gSH2 gAR3	MgS*†
	ARUS	gSH2 gAR4	MgS*†
	ARUS	gSH2 gAR5	MgS*†
TOTAL n=4			
Falkland Islands	FIBI	gFI2 gFI3	Mg/Me*†
	FIBI	gFI1 gFI3	Mg/Me*†
	FIST	gSH2 gFI1	MgS*†
	FIST	gSH2 gFI2	MgS*†
TOTAL n=4			

These represent haplotypes unique to a single country. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. Taxonomic identity inferred from * - 16s RFLP and † - placement in BI phylogenetic tree. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis*/*M. edulis* Northern.

Table 4.9 – Haplotype names, taxonomic identities and distributions for the combined data set for Northern hemisphere locations.

Country	Region	Haplotype (16s COIII)	Taxonomic Identity
Russia	BLSE	gBS1 gRU1	Mg/Me*†
	BLSE	gGL2 gRU2	MgN*†
	BLSE (5)	gBS1 gRU3	Mg/Me*†
	BLSE	gGL2 gRU4	MgN*†
TOTAL n=4			
England	GBMU	etNH1 eGB11	Mg/Me*†
	GBMU	etNH1 eGB12	Mg/Me*†
	GBMU	etNH1 eGB13	Mg/Me*†
	GBLO	etNH1 eGB2	Mg/Me*†
	GBLO	etNH1 eGB3	Mg/Me*†
	GBLO	eGB2 geEU1	Mg/Me*†
TOTAL n=6			
Canada	CHEM	tPC4 gtGL2	Mt*†
	CHEM	tPC4 tCA6	Mt*†
	CHEM	tPC4 tCA7	Mt*†
	COMO	tPC4 gtGL1	Mt*†
	COMO	tPC1 tCA2	Mt*†

Country	Region	Haplotype (16s COIII)	Taxonomic Identity
	MYCA	etNH1 tCA9	Mg/Me*†
	MYCA	etNH1 tCA10	Mg/Me*†
	MYCA	etNH1 tCA11	Mg/Me*†
	MYCA	tAC2 tCA13	Mt*†
	MYCA	tAC2 tCA14	Mt*†
TOTAL n=10			

These represent haplotypes unique to a single country. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. * - taxonomic identity inferred from placement in the phylogenetic tree (BI), † - taxonomic identity inferred from 16s RFLP. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis/M. edulis* Northern, Mt = *M. trossulus*.

Table 4.10 – Haplotype names, taxonomic identities and distributions for the combined data set.

Haplotype (16s COIII)	Country (Region)	Taxonomic Identity
gSH2 gSH2	New Zealand (AUCI)	MgS*†
gSA1 gSA1	Argentina (ARUS)	MgS*†
gGL1 etNH1	Canada (CHEM)	Mg/Me*†
etNH1 eEU1	England (GBLO)	Mg/Me*†
gGL1 gGL2	New Zealand (MUPO (3),MAUI) Turkey (IZMI) (6)	Mg/Me*†
gGL2 gtGL1	New Zealand (MAUI) Chile (CHCO) Turkey (IZMI)	MgN*†
gGL2 gGL1	Chile (CHCO)	MgN*†

These represent haplotypes that are identified in more than one country and/or hemisphere. The number in brackets after each region code is the number of individuals in which that haplotype was found, a single individual is inferred if there is no number given. * - taxonomic identity inferred from placement in the phylogenetic tree (BI), † - taxonomic identity inferred from 16s RFLP. MgS = *M. galloprovincialis* Southern, MgN = *M. galloprovincialis* Northern, Mg/Me = *M. galloprovincialis/M. edulis* Northern.

4.4.3 Genetic distance

The average genetic distance within haplogroups is less than the average genetic distance between haplogroups (Tables 4.11 and 4.12) for both the COIII and 16s genes with the exception of *M. trossulus* COIII.

Table 4.11 – The mean genetic distance between haplogroups for 16s (above the diagonal) and COIII (below the diagonal).

	M. galloprovincialis (S) (n=70)	*M. galloprovincialis* (N) (n=8)	*M. galloprovincialis* (Mg/Me) (n=13)	*M. edulis* (Mg/Me) (n=25)	*M. trossulus* (n=10)
M. galloprovincialis (S) (n=9)	-	0.011	0.017	0.016	0.030
M. galloprovincialis (N) (n=3)	0.069	-	0.019	0.017	0.035
M. galloprovincialis (Mg/Me) (n=5)	0.078	0.045	-	0.005	0.037
M. edulis (Mg/Me) (n=3)	0.078	0.043	0.021	-	0.035
M. trossulus (n=5)	0.081	0.043	0.031	0.027	-

Haplogroups are divided into species with S = Southern lineage, N = Northern lineage and Mg/Me = shared *M. galloprovincialis*/*M. edulis* Northern haplogroup. Sample sizes indicated on column header are COIII and row header are 16s.

Table 4.12 – The mean genetic distance within haplogroups for the 16s and COIII genes.

Haplogroup	COIII	16s
M. galloprovincialis (S)	0.026	0.007
M. galloprovincialis (Mg)	0.020	0.007
M. galloprovincialis (Mg/Me)	0.024	0.006
M. edulis (Mg/Me)	0.016	0.003
M. trossulus	0.036	0.015

Haplogroups are divided into species with S = Southern lineage, N = Northern lineage and Mg/Me = shared *M. galloprovincialis*/*M. edulis* Northern haplogroup.

The mean genetic distance within *M. galloprovincialis* haplogroups ranged from 2.0% to 2.6% for the COIII gene and 0.6% to 0.7% for the 16s gene (Table 4.11). The Southern hemisphere *M. galloprovincialis* lineage had the smallest mean COIII genetic distance between the *M. galloprovincialis* Northern lineage at 6.9% and the largest distance between *M. trossulus* at 8.1% (Table 4.12). The mean COIII genetic distance between the Southern hemisphere *M. galloprovincialis* haplogroup and all other species/haplogroups was much greater than the distance between sibling species of the *M.*

edulis complex, a range of 2.1% to 4.5% between all other pairwise comparisons of species/haplogroup combinations. Within the shared *M. galloprovincialis*/*M. edulis* haplogroup, species distinction between *M. edulis* and *M. galloprovincialis* translates to a mean genetic distance of 2.1% for the COIII gene and 0.5% for the 16s gene. In contrast, the mean genetic distance between *M. edulis* and the *M. galloprovincialis* Northern haplogroup (MgN) was 4.3% for COIII and 1.7% for 16s (Table 4.11). *M. trossulus* has a within group genetic distance of 3.6% for COIII and 1.5% for 16s and between all other haplogroups ranging from 2.7% to 8.1% and 3.0% to 3.7%, respectively. The small distance (COIII) between *M. trossulus* and *M. edulis* (2.7%) is a consequence of shared haplotypes among the Atlantic coasts of Canada and Europe and supports previous results of female mitotype gene flow across the Atlantic Ocean (Riginos et al. 2004).

4.4.4 Phylogenetic Analyses

4.4.4.1 COIII single gene data set – Bayesian inference

The model of nucleotide substitution chosen under the a priori specified BIC by ModelTest v3.7 (Posada & Crandall 1998) was the general-time-reversal model with gamma parameter (GTR + Γ) for the COIII data set. Bayesian analysis of the single COIII and 16s gene data sets shows discordant topologies (Figures 4.3 and 4.4, respectively). In the unresolved (represented by polytomies only at the population level and not at the haplogroup/species level) COIII single gene Bayesian tree (Figure 4.3a), individual clades formed by haplotypes from the *M. galloprovincialis* Southern haplogroup, as determined by the 16s RFLP, form a monophyletic grouping that is sister to the polyphyletic *M. galloprovincialis*/*M. edulis* Northern hemisphere haplogroup. The divergence of the two Northern hemisphere haplogroups is shown here to pre-date the divergence between Northern and Southern hemispheres (Figure 4.3a). The *M. galloprovincialis*/*M. edulis*

Northern hemisphere haplogroup contains two main clades, one of which contains *M. galloprovincialis* haplotypes from geographical locations East of the Almeria-Oran oceanographic front and hence, would be considered part of the Northern *M. galloprovincialis* haplogroup. These locations in Russia and Turkey also contain haplotypes that have segregated onto other divergent clades within the tree (the *M. galloprovincialis* Northern clade), suggesting the barrier to gene flow in the Western Mediterranean Sea is not strong enough to cause complete population separation. The distribution of haplotypes from a single geographic origin onto divergent clades of the tree could also represent secondary contact due to either natural or anthropogenic causes.

Within the *M. galloprovincialis* Southern hemisphere haplogroup cluster, there are three clearly separated clades corresponding to South American (Chile and Argentina), New Zealand and Australian samples. The separation is exactly correlated to geography except a single New Zealand haplotype, gNZ 16, appears on the South American clade. This haplotype was detected on the Chatham Islands, a group of offshore islands approximately 600km east of New Zealand, and its appearance on the South American clade hints at some gene flow between these regions. New Zealand and Australian clades are most closely related having diverged after the South American clade diverged. Haplotypes gSH 2 and gSH 3, which are shared between Chile and New Zealand, are present in the South American clade. The first shared haplotype gSH 2 is found in the Auckland Islands (NZ) and Concepcion (Chile) and the second shared haplotype gSH 3 is found in Maud Island (NZ) and Colchogue (Chile).

Hard polytomies on branches indicate population level uncertainty. The polyphyletic sister group containing *M. galloprovincialis*/*M. edulis* Northern haplotypes does not show reliable geographic separation between countries or species. In fact, *M. trossulus* haplotypes are found interspersed throughout this clade grouping as well, which

echoes the small genetic distance between the three species. Hybrid *M. edulis/ M. galloprovincialis* individuals have been identified previously in the Auckland Islands as introduced and Northern hemisphere *M. galloprovincialis* mtDNA haplotypes have been identified at Maud Island (Westfall & Gardner in press) making the inference of natural gene flow impossible to discern from anthropogenic introduction events between Chile and New Zealand. Within the monophyletic *M. galloprovincialis* Southern hemisphere grouping, the regional South American clade is basal to the New Zealand and Australian clades, the latter two groups descending from a common ancestor with South America.

The *M. galloprovincialis* Northern haplogroup is divided into two clades; both appear to diverge before the *M. galloprovincialis*/*M. edulis* haplogroup and one clade shares the most recent common ancestor with the *M. edulis/ M. galloprovincialis* Northern haplogroup and the *M. galloprovincialis* Southern haplogroup, supporting earlier hypotheses of evolutionary relationships between Northern and Southern hemisphere *M. galloprovincialis* haplogroups (Hilbish et al. 2000).

(a)

(b)

Mg N

Mg / Me

Mg S

Mg N

Figure 4.3 – COIII single gene trees.

Bayesian inference (a) and maximum likelihood inference (b), both are rooted with *M. californianus*. Bayesian inference bipartition confidence displayed as posterior probability and maximum likelihood inference bipartition confidence displayed as bootstrap support (%). MgS = *M. galloprovincialis* Southern haplogroup, MgN = *M. galloprovincialis* Northern haplogroup and Mg/Me = *M. galloprovincialis*/*M. edulis* Northern haplogroup.

4.4.4.2 COIII single gene data set – Maximum Likelihood inference

The model of nucleotide substitution chosen under the a priori specified AICc by ModelTest v3.7 (Posada & Crandall 1998) was the general-time-reversal model with gamma parameter (GTR + Γ) for the COIII single gene data set. The 50% majority rule consensus tree generated by PAUP* for the 1000 bootstrap pseudo-replicates repeated in four independent runs in GARLI v0.96 (Zwickl 2006) (Figure 4.3b) showed a similar topology to the Bayesian analysis. The *M. galloprovincialis* Northern haplogroup is basal to both the *M. galloprovincialis*/*M. edulis* Northern and *M. galloprovincialis* Southern haplogroups, the latter two groups being closely related sister groups. *M. trossulus* is paraphyletic and appears in clades dominated by the *M. galloprovincialis* Northern and *M. galloprovincialis*/*M. edulis* Northern haplogroups, indicating incomplete lineage sorting or ongoing gene flow among these sibling species. Within the Southern hemisphere *M. galloprovincialis* cluster, regional differentiation with clades is apparent again and South America is basal to New Zealand and Australian clades (which appear as sister clades).

A close inspection of the distribution of haplotypes in the Northern hemisphere as they are classified into haplogroups is imperfect. This is not surprising by any means because secondary contact over moderate and large geographic scales with introgressive hybridisation allows for cross-over of haplotypes into other haplogroups, especially at the single gene level.

4.4.4.3 16s rDNA single gene data set – Bayesian inference

The model of nucleotide substitution chosen under the a priori specified BIC by ModelTest v3.7 (Posada & Crandall 1998) was the Hasegawa-Kishino-Yano (HKY) (Hasegawa et al. 1985) model for the 16s rDNA data set. The Bayesian 16s tree shows *M. galloprovincialis* Southern lineage haplotypes as unresolved polytomies at the basal portion of the tree and all Northern hemisphere haplotypes, including *M. trossulus*, are derived (Figure 4.4a). Considering these as hard polytomies is possible if we consider that lineage sorting of the Southern hemisphere *M. galloprovincialis* haplogroup has widely progressed and the taxa represented by polytomies belong to the same haplogroup, which is demonstrated in the current study COIII single gene tree and the COI tree from Gérard et al. (2008).

The origin of Kerguelen Island *Mytilus* spp. populations has been a recent topic of debate, and the haplotypes sampled from that area appear in the *M. galloprovincialis*/*M. edulis* clade (sampled by Hilbish et al. 2000) and the unresolved *M. galloprovincialis* Southern polytomy (sampled by Gérard et al. 2008), indicating there are endemic populations in that area which may have been affected by anthropogenic introduction, or conversely, echoes reticulated evolution or shared ancestry. The shared ancestry hypothesis is congruent with the COIII single gene and combined data set phylogenies (Figures 4.3a and 4.5, respectively), which show *M. galloprovincialis* Southern and *M. galloprovincialis*/*M. edulis* Northern haplogroups as closely related sister clades descending from a common ancestor.

4.4.4.4 16s rDNA single gene data set – Maximum Likelihood inference

The model of nucleotide substitution chosen under the a priori specified AICc by ModelTest v3.7 (Posada & Crandall 1998) was the Hasegawa-Kishino-Yano (HKY)

(Hasegawa et al. 1985) model for the 16s rDNA data set. The maximum likelihood phylogenetic reconstruction of the 16s single gene data set (Figure 4.4b) is congruent with the Bayesian Inference reconstruction. We see here the unresolved basal polytomy formed by *M. galloprovincialis* Southern haplotypes, although a lot of the haplotypes form bipartitions with no observable order of divergence. One important difference between the inference methodologies is the maximum likelihood tree shows the divergence of *M. trossulus* to pre-date the divergence of all other species/haplogroups in the maximum likelihood tree, a result that is congruent with our current understanding of *Mytilus edulis* complex evolution (Seed 1992, Riginos et al. 2004). Haplotypes sampled from the Falkland Islands appear in the *M. edulis*/*M. galloprovincialis* clade (sampled in this study) and in the *M. galloprovincialis* Southern polytomy (sampled by Hilbish et al. 2000).
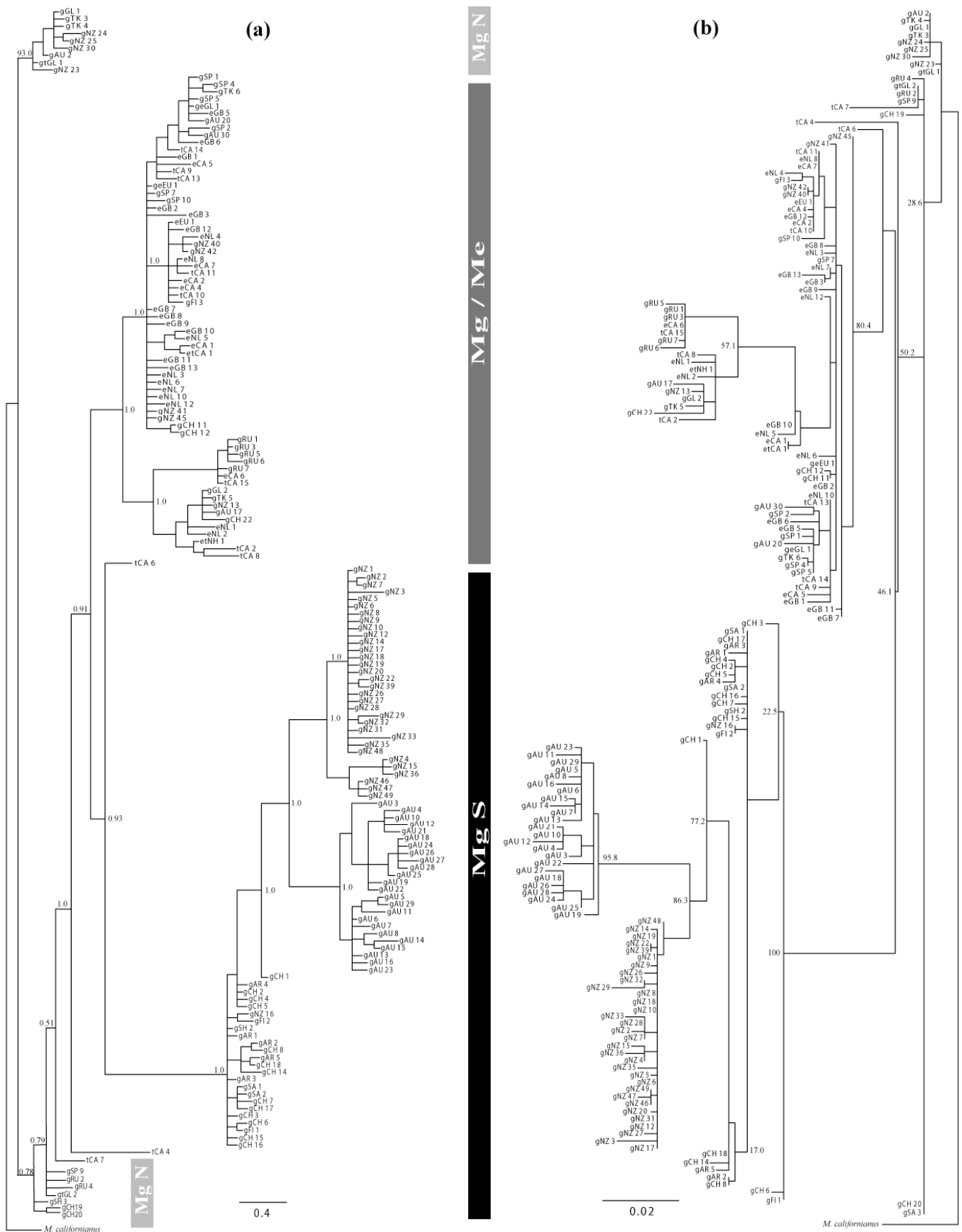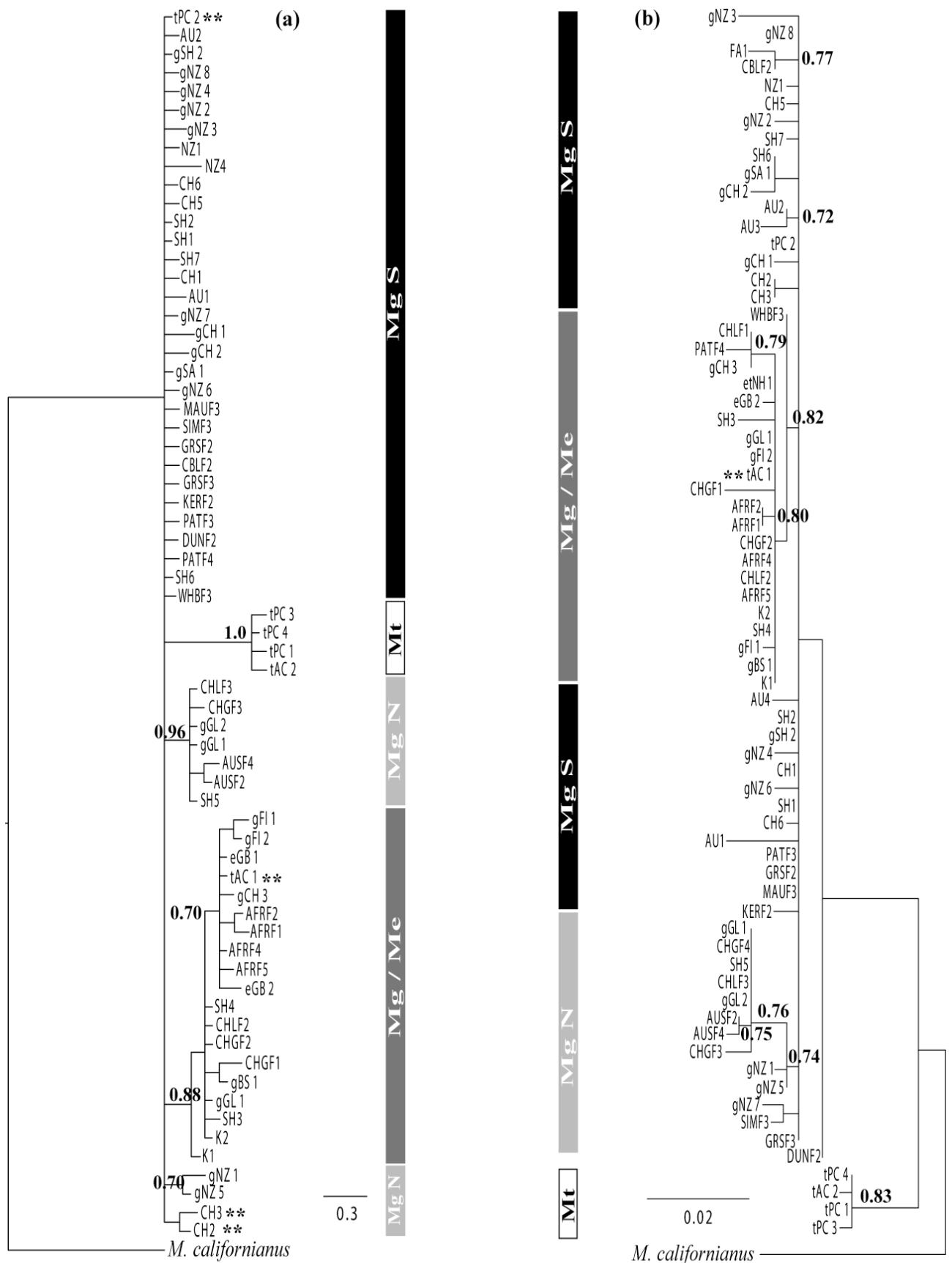
Figure 4.4 – 16s single gene trees.

Bayesian inference (a) and maximum likelihood inference (b), both are rooted with *M. californianus*. Bayesian inference bipartition confidence displayed as posterior probability and maximum likelihood inference bipartition confidence displayed as bootstrap support (%) and is not written if <50%. MgS = *M. galloprovincialis* Southern haplogroup, MgN = *M. galloprovincialis* Northern haplogroup and Mg/Me = *M. galloprovincialis/M. edulis* Northern haplogroup. ** = taxon does not match haplogroup placement. Note scale bars are different.

4.4.4.5 Combined data set – Bayesian Inference

The combined COIII and 16s gene data set under a mixed model of nucleotide substitution hypothetically yields a more accurate phylogenetic reconstruction than either single gene tree. This phylogenetic reconstruction (Figure 4.5) exhibits features from both of the single gene trees including clear clade groupings of haplogroups. *M. trossulus* haplotypes are present as a monophyletic clade, a feature present in the 16s rDNA single gene tree. The *M. galloprovincialis* Southern haplogroup becomes resolved, opposed to the 16s rDNA single gene tree, and is present as a derived haplogroup of which the *M. galloprovincialis/M. edulis* Northern haplogroup is sister to. Like the COIII single gene tree, the *M. galloprovincialis* Northern haplogroup appears basal to all haplogroups except *M. trossulus*. On the single COIII gene tree, the *M. galloprovincialis* Northern haplogroup was broken into two clades, one of them diverging before the other. It is interesting to see the haplotypes gGL 1 and gtGL 1, which clustered on the early diverging *M. galloprovincialis* clade on the COIII single gene tree are now present in the only *M. galloprovincialis* Northern clade on the combined tree. Also like the COIII single gene tree, clade groupings within the *M. galloprovincialis* Southern haplogroup are separated based on regions of South America and New Zealand, with South America appearing basal to New Zealand (note that Australian samples are not included in the combined data set).

There is a cleaner separation of haplotypes into respective haplgroups with the addition of loci although posterior probability support values are lower overall (Figure 4.5). We can easily denote *M. edulis* haplotypes and therefore ascribe the *M.*

galloprovincialis/*M. edulis* haplogroup to this clade. Even with additional loci we still cross-over, for example, Russian haplotypes cluster on a clade that is closely related to the *M. edulis* haplotypes but still considered as part of this haplogroup.
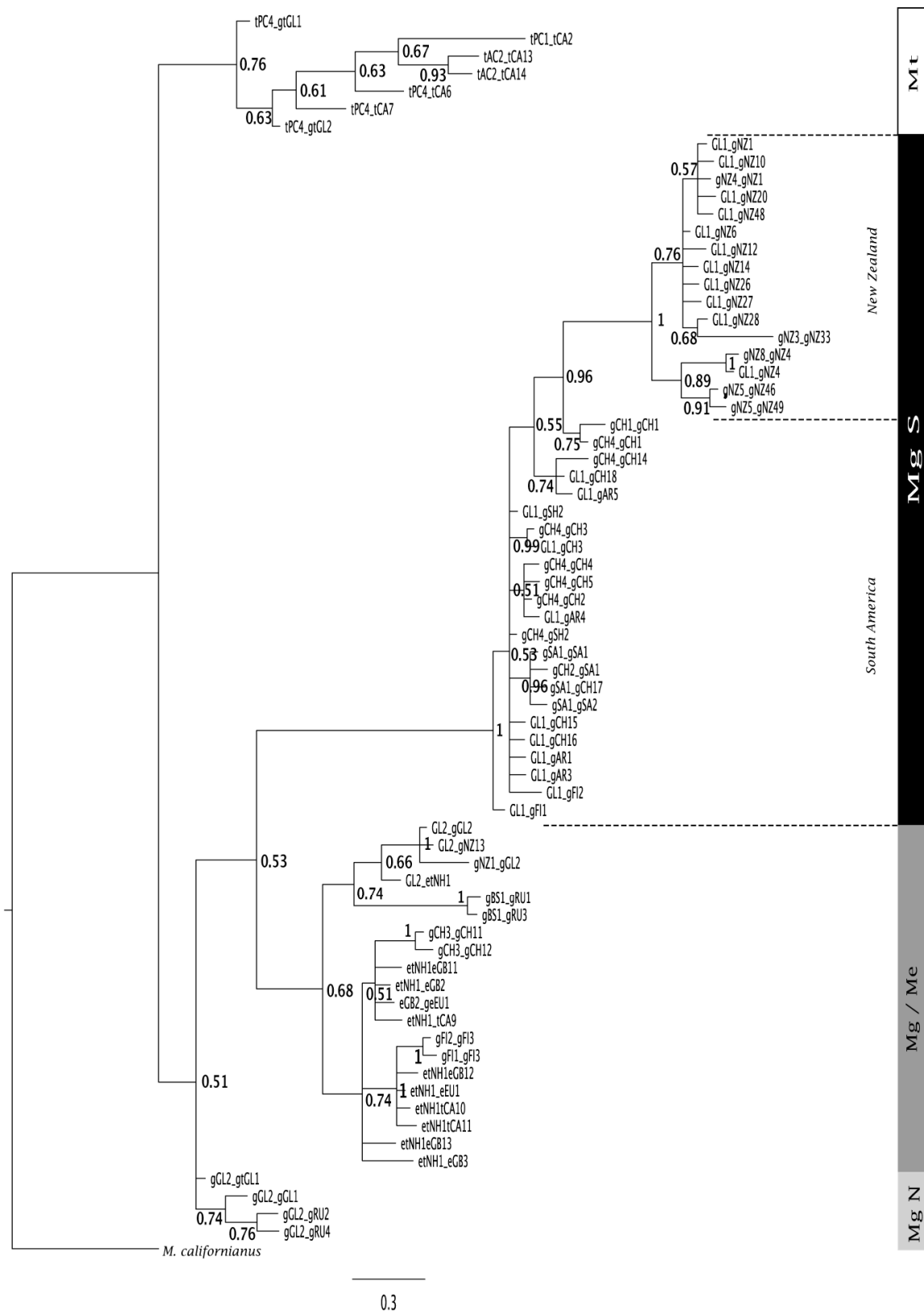
Figure 4.5 – combined COIII and 16s tree.

Bayesian inference, bipartition confidence displayed as posterior probability. MgS = *M. galloprovincialis* Southern haplogroup (separated into regional clade groupings), MgN = *M. galloprovincialis* Northern haplogroup and Mg/Me = *M. galloprovincialis*/*M. edulis* Northern haplogroup

4.4.5 Divergence dating

4.4.5.1 Molecular clock

Tajima's relative rate test implemented in MEGA for single gene COIII and 16s data sets show that the *M. galloprovincialis* Southern lineage is evolving at a significantly different rate than other haplogroups within the *M. edulis* species complex for the COIII gene but not for the 16s gene. Pairwise comparisons indicate the monophyletic *M. galloprovincialis* Southern lineage COIII gene evolves at a different rate than other haplogroups, corresponding to major clades on the tree (Table 4.13). In contrast, the haplogroups which constitute major clades on the 16s tree are evolving at similar rates and can be considered to adhere to a strict molecular clock rate of evolution (Table 4.13).

Table 4.13 – P-values for Tajima's relative rate test (Tajima 1993).

	M. galloprovincialis (S)	*M. galloprovincialis* (N)	*M. galloprovincialis* (Mg/Me)	*M. edulis* (Mg/Me)	*M. trossulus*
M. galloprovincialis (S)	-	0.15730	1.00000	1.00000	0.76320
M. galloprovincialis (N)	0.00000**	-	0.41422	0.41422	0.40538
M. galloprovincialis (Mg/Me)	0.00042**	0.25135	-	1.00000	0.78151
M. edulis (Mg/Me)	0.00468 *	0.05935	0.31731	-	0.78151
M. trossulus	0.00003**	0.52709	0.36571	0.08236	-

16s (above the diagonal) and COIII (below the diagonal) single gene data sets, significant (* = $p<0.05$) and highly significant (** = $p<0.001$) values indicate rejection of the null hypothesis of rate homogeneity.

In contrast, the global parametric likelihood ratio test applied to single COIII and 16s gene data sets testing the adherence to a global molecular clock (throughout all branches of the tree) showed that both data sets do not adhere to a strict molecular clock of nucleotide evolution with highly significant p-values (<0.001 for both data sets) that reject

the null hypothesis of rate homogeneity (Table 4.14). Although there was discrepancy between the global and local molecular clock tests for the 16s rDNA single gene data set, divergence dating in BEAST was based on the results of the relative rate test supporting previously published results for the 16s rDNA gene in *Mytilus* spp. (Gérard et al. 2008, Hilbish et al. 2000).

Table 4.14 – Global parametric likelihood ratio test results

	Likelihood		Test Statistic	Degrees of Freedom	p-value
	Clock	No Clock			
COIII	- 4073.316	- 3945.041	256.550	165	0.00001**
16s	- 1370.605	- 1266.113	208.984	69	5 e-16**

Test for COIII and 16s single gene data sets. ** = p<0.001 - highly significant rejection of null hypothesis of rate homogeneity.

4.4.5.2 Intra- and interlineage estimation

The focus of divergence dating in this paper is for the split between Northern and Southern hemisphere lineages of *M. galloprovincialis*, and additionally, within lineage divergence estimates for clades associated with the Southern hemisphere regions of South America, New Zealand and Australia. The Yule birth-death speciation model applied in the program BEAST for the combined data set partitioned into separate genes utilised a relaxed molecular clock for the COIII partition and a strict molecular clock for the 16s partition, in accordance with the results of relative rate tests. When divergence of *M. trossulus* and *M. edulis* was set to coincide with the current hypothesis of timing for the trans-Arctic interchange at 3.5mya (Vermeij 1991), the divergence estimate of the Southern *M. galloprovincialis* lineage was estimated at 2.22 ± 4.33E-2 mya. When divergence of *M. trossulus* and *M. edulis* was set to coincide with an alternative hypothesis for the first opening of the Bering Strait at 5.95 (SD=1.15) mya (midpoint between 4.8 and

7.4 mya described in Marinkovich & Gladenkov 1998), when the potential for *Mytilus* spp. migration across the Arctic was still possible, the divergence estimate of the Southern *M. galloprovincialis* lineage was estimated at 3.52 (SE 4.07E-2) mya. Because branching order assumed from the combined 16s/COIII Bayesian tree does not match the true evolutionary history of *M. edulis* and M. galloprovinialis due to secondary mixing, this divergence date could be an overestimation.

The coalescent-based mDIV calculations for divergence time (with no calibration points used) based on the COIII single gene data set place divergence between the *M. galloprovincialis*/*M. edulis* Northern haplogroup and the Southern *M. galloprovincialis* haplogroup during the late Pliocene, from 2.04 to 2.66 mya. In addition, divergence dates within the regional clades of the Southern hemisphere *M. galloprovincialis* lineage based on the COIII single gene data set are listed in Table 14.

The divergence dates between South America and the New Zealand/Australia regions are 1.876 to 1.928 mya and between New Zealand and Australia at 0.0513 mya. This estimate concurs with the previously hypothesised route of migration of Northern hemisphere mussels through the Atlantic Ocean (Hilbish et al. 2000) due to the earliest divergence date in South American populations. This is earlier than previous estimates that place the divergence time between 0.84 and 1.0 mya (Hilbish et al. 2000, Gérard et al. 2008).

4.4.6 Topological Hypothesis Testing

4.4.6.1 COIII – Bayes Factors and SH-Test

The topological constraints derived from previously published topologies by Hilbish et al. (2000) and Gérard et al. (2008) (Figure 4.2) applied against the COIII single gene data set were not found to be a significantly worse fit to the data by Bayes Factors

although the likelihood was marginally closer to zero for the unconstrained tree (Table 4.15). Rearranging the COIII single gene topology to change the sister group of *M. galloprovincialis* Southern haplogroup from the *M. galloprovincialis*/*M. edulis* Northern haplogroup to the *M. galloprovincialis* Northern haplogroup (derived from Hilbish et al. 2000, Hypothesis #2) descending from a shared common ancestor with the *M. galloprovincialis*/*M. edulis* haplogroup resulted in a non-significant better fit to the data (likelihood closer to zero) (Table 4.15). Rearranging the COIII topology so *M. galloprovincialis*/*M. edulis* and *M. galloprovincialis* Northern haplogroups are sister clades that sharing a common ancestor with *M. galloprovincialis* Southern haplogroup that descended from the most recent common ancestor with *M. trossulus*, i.e. the Southern *M. galloprovincialis* divergence pre-dates the divergence between Northern *M. galloprovincialis* haplogroups (derived from the COI and 16s trees of Gérard et al. 2008, Hypothesis #1) resulted in a significantly worse fit to the data in the SH-test (Table 4.15). The evidence against the hypothesis of the Southern *M. galloprovincialis* haplogroup indicated from the COI and 16s trees of Gérard et al. (2008) is strong with the COIII single gene data set. The sister status of *M. galloprovincialis* Southern and *M. galloprovincialis*/*M. edulis* haplogroups descending from a common ancestor with the *M. galloprovincialis* Northern haplogroup inferred from the COIII topology holds up when tested against previously published topologies.

4.4.6.2 16s – Bayes Factors and SH-Test

The topological constraints derived from previously published topologies by Hilbish et al. (2000) and Gérard et al. (2008) applied against the 16s single gene data set were found not to be a significantly worse fit to the data (note p=0.055 Table 4.15). Although Bayes Factor evidence for comparison of the current 16s topology to both

hypothetical constraints are not worth a mention, meaning the basal polytomy of the *M. galloprovincialis* Southern haplogroup in the current study holds up in the face of rearrangements to place it in a derived position on the tree, these polytomies tend to indicate irresolvable portions rather than true basal characteristics.

Table 4.15 – Topological hypothesis testing results

		Bayes Factors				Shimodaira-Hasegawa Test			
		Marginal Likelihood Value (Harmonic mean)				Maximum Likelihood Value (Arithmetic mean)			
Data Set	Constraint	normal	constrained	BF	evidence	normal	constrained	test st.	p-value
COIII	**16s** (Hilbish et al. 2000)	-3637.73	-3644.22	0.0	nwm	-3723.48	-3705.00	18.48	0.055
COIII	**16s/COI** (Gérard et al. 2008)	-3637.73	-3666.13	0.0	nwm	-3723.48	-4725.93	1002.45	0.000*
16s	**16s** (Hilbish et al. 2000)	-1341.73	-1339.99	0.0	nwm	-1297.96	-1713.30	415.33	0.012*
16s	**16s/COI** (Gérard et al. 2008)	-1341.73	-1338.22	0.0	nwm	-1297.96	-1725.65	427.68	0.010*

Topological comparisons of Hilbish et al. (2000) and Gérard et al. (2008) with single gene COIII and 16s data sets from the current study. nwm – not worth a mention, qualitative non-significance.

4.5 Discussion

4.5.1 Genetic identity of Southern hemisphere *M. galloprovincialis*

The unique genetic identity of the Southern hemisphere lineage of *M. galloprovincialis* is demonstrated by the identification of COIII haplotypes unique to this lineage and not shared with Northern hemisphere conspecifics, a result shared by two previous mtDNA studies by Hilbish et al. (2000) and Gérard et al. (2008). Any haplotypes identified in this study that are shared between hemispheres are thought to have arisen from anthropogenic-mediated introductions and are not due to secondary natural migration events, leading to a reciprocally monophyletic Southern hemisphere clade. This point is not well considered in the published literature of *Mytilus* taxonomy; the presence of nuclear alleles, allozyme allelomorphs or Northern hemisphere mitochondrial haplotypes

in the Southern hemisphere that are considered diagnostic between species in the Northern hemisphere is invoked to underlie hypotheses of secondary contact with introgression (Daguin & Borsa 2000, Borsa et al. 2007), reticulated evolution (Borsa et al. 2007) and multiple migration events to the Southern hemisphere (Hilbish et al. 2000, Gérard et al. 2008), albeit Gérard et al. (2008) used the ancestral branching order of Southern hemisphere *M. galloprovincialis* (COI) to hypothesize a scenario of shared ancestry and skewed taxonomic preconceptions. The established invasive nature of Northern hemisphere *M. galloprovincialis* (Lowe et al. 2000, Wonham 2004) and it's ability to out-compete native species in places such as South Africa (van Erkom Schurink & Griffiths 1991), the North-east Pacific (Geller 1994) and the Sea of Japan cannot be disregarded when considering the presence of haplotypes shared between hemispheres.

In addition, the elevated presence of haplotypes shared between hemispheres at international port locations (i.e. Wellington Harbour, NZ or Melbourne, AU) or areas of intense aquaculture production (i.e. Marlborough Sounds, NZ) in the Southern hemisphere (Westfall & Gardner in press) provides additional evidence for distribution to be solely due to anthropogenic-mediated introduction. All of the haplotypes identified in this study as shared between hemispheres resolved onto Northern hemisphere clades of the COIII and combined data set phylogenies and cross-validation of a sample subset with the 16s RFLP assay revealed Northern hemisphere identity, further evidence that these haplotypes are of recent Northern hemisphere origin. Additionally, the 16s RFLP identified Northern hemisphere haplotypes sampled from New Zealand, Australia and Chile that weren't shared among hemispheres but clustered onto Northern hemisphere clades.

Genetic distance measures of haplotypes classified into groups based on the diagnostic 16s RFLP and species status (Westfall et al. 2010) along with the criterion of reciprocal monophyly provide strong support for reclassification of the Southern

hemisphere lineage of *M. galloprovincialis* into either a new species within the *M. edulis* complex or a subspecies of *M. galloprovincialis*, the latter a proposal already imparted by Daguin & Borsa (2000). The COIII genetic distances (Table 4.11) in this study between species in the Northern hemisphere ranged from 2.1% (between *M. galloprovincialis* and *M. edulis* from the same haplogroup) to 4.5% (between *M. galloprovincialis* from the Northern and shared haplogroups). When considering the average genetic distance between hemispheres, these values reach more than two-fold the distance among Northern *M. galloprovincialis* from 6.9% to 7.8%, and the distance between *M. trossulus* and Southern hemisphere *M. galloprovincialis* at 8.1%. The situation is similar for the more slowly evolving 16s rDNA gene, although the absolute values of genetic distance are considerably smaller, the genetic distance between hemispheres is roughly two-fold the distance among Northern hemisphere haplogroups and species.

Within the COIII and combined data set phylogenetic trees, Southern hemisphere *M. galloprovincialis* form highly and moderately supported, respectively, reciprocally monophyletic clades that appear to have undergone widespread lineage sorting, a pattern repeated in previously published phylogenies of Gérard et al. (2008), Daguin & Borsa (2000) and Hilbish et al. (2000). If the current levels of divergence aren't sufficient to warrant species or sub-species status, continued reproductive isolation between hemispheres due to geographic distance and inhospitable conditions across tropical regions as a barrier to gene flow will only lead to greater differentiation among the lineages.

4.5.2 Phylogenetic inference

4.5.2.1 Congruence among current and published topologies

The clades associated with predefined haplogroups for *M. galloprovincialis* and *M. edulis* exist on all phylogenetic reconstructions to date (with the exception of COI

phylogeny of Gérard et al. 2008), the only incongruence lies is the branching order of these haplogroups. The 16s based hypothesis of Hilbish et al. (2000) infers the Southern and Northern *M. galloprovincialis* are sister groups and the divergence of the *M. galloprovincialis*/*M. edulis* haplogroup pre-dates the divergence of the former haplogroups. In the current hypothesis derived from COIII and combined data sets, the positions of the Northern *M. galloprovincialis* and shared *M. galloprovincialis*/*M. edulis* are reversed, indicating the Northern *M. galloprovincialis* haplogroup diverged first. The topologies of Gérard et al. (2008) and Daguin & Borsa (2000) are more incongruent with this data set, inferring the Southern *M. galloprovincialis* divergence pre-dates the Northern hemisphere divergence of *M. edulis* and *M. galloprovincialis*, an interesting result because Gérard et al (2008) inferred phylogeny from the 16s and COI genes, the latter a sub-unit of the same gene complex used in this study. The time of Northern and Southern hemisphere divergence estimated at 0.84 mya by Gérard et al. (2008) is discordant with the current hypothesis on the timing of *M. galloprovincialis* divergence in the Mediterranean Sea at 1-2 mya (Seed 1992). If Southern hemisphere *M. galloprovincialis* divergence indeed pre-dated the separation of *M. galloprovincialis* and *M. edulis* in the Northern hemisphere, the estimated time of Northern and Southern hemisphere divergence should pre-date 1-2 mya.

Incongruent topologies constructed from separate mtDNA genes in previously published and current studies highlights the risk of inferring species relationships from single gene trees, therefore, it is important to turn our attention to the combined COIII/16s gene topology (Figure 4.5). The evolutionary relationships inferred by the combined topology are similar to the single gene COIII topology in that Southern lineage *M. galloprovincialis* and the Northern *M. galloprovincialis*/*M. edulis* haplogroup are sister clades that share a common ancestor. The relationship described here is supported by nuclear studies that identify alleles diagnostic of Northern hemisphere *M. edulis* in

Southern hemisphere *M. galloprovincialis*. Since the Southern hemisphere *M. galloprovincialis* haplogroup shares a common ancestor with the *M. galloprovincialis*/*M. edulis* Northern haplogroup, ancestral genes common to both haplogroups could be detected and the distribution of *M. edulis*-type alleles in Southern hemisphere populations is most likely a product of genetic drift and selection.

The order of divergence suggested by COIII and combined phylogenies is not in line with known evolutionary histories; the presence of *M. trossulus* interspersed throughout other Northern hemisphere clades is indicative of introgressive hybridisation through secondary contact with sibling species. Another aspect is the presence of *M. edulis* in a clade derived from the more basal *M. galloprovincialis* Northern clade. In this case, *M. edulis* is known to have diverged from *M. trossulus* well before *M. galloprovincialis* diverged in the Mediterranean Sea. The effects of glaciation events in Northern Europe could be in part responsible for the derived position of *M. edulis* in the tree. If refugia existed during times of glaciation, heavier ice cover in Northern Europe compared to the Mediterranean and Atlantic coast of Spain, the Atlantic *M. galloprovincialis* haplogroup could have repopulated much of the coastline following ice melt and introgressed with the existing *M. edulis* refugial populations.

4.5.2.2 Differentiation within the Southern hemisphere and the Atlantic migration hypothesis

A congruent feature among the COI (Gérard et al. 2008), COIII and combined topologies (current study) is the highly supported separation of clades within the Southern hemisphere that correspond to geographical regions of South America, New Zealand and Australia (Tasmania in Gérard et al. 2008). Gérard et al. (2008) did not find evidence to support an isolation by distance correlation as New Zealand and Tasmanian samples exhibited greater differentiation than South American and New Zealand samples and the

East Australia Current was cited to explain the relative isolation of Tasmanian populations. Monophyletic clades associated with geographical separation is evidence of reduced gene flow and widespread lineage sorting among these regions, and attempts to date the regional divergence times are congruent with the order of divergence inferred from these topologies. For instance, the divergence between South America and New Zealand is estimated at 1.928 mya and between South America and Australia at 1.876 mya, corresponding to the COIII and combined data set topologies that infer New Zealand and Australia as closely related sister groups with New Zealand diverging slightly before Australia (Figures 3.3a and 3.5). This supports the theory of an Atlantic route of *M. galloprovincialis* migration from the Northern hemisphere (Hilbish et al. 2000) with colonisation first appearing in South America and then subsequent migration to New Zealand and Australia followed by complete lineage sorting. Despite these past migration events inferred from topological positioning, there seems to be little to no gene flow occurring with sufficient regularity to maintain intraspecific interbreeding between New Zealand and Australia. Although the phylogenetic evidence to date supports the theory of an Atlantic route of migration to the Southern hemisphere, the absence of endemic *Mytilus* spp. on the Western coast of Africa is still puzzling. The Benguela Upwelling System cools the waters of the west and south coasts of southern Africa enough to support a rich assemblage of temperate flora and fauna. There is no fossil evidence of *Mytilus* spp. along the southern coastline of Africa, and their absence is peculiar. In addition, the presence of an introduced population of *M. galloprovincialis* that has exploded in numbers and displaced many native species shows that this species can successfully inhabit this coastline. If *Mytilus* spp. migrated via the Atlantic Ocean to the southern hemisphere, one would expect evidence of extinct (if not extant) populations.

4.5.2.3 Hypotheses on the origin of Southern hemisphere M. galloprovincialis

Daguin & Borsa (2000) reconciled information from several allozyme loci into a phylogenetic tree that shows Northern and Australasian (New Zealand and Australia) lineages of *M. galloprovincialis* as sister groups and hypothesizes that Australasian *M. galloprovincialis* arose from a proto-*M. galloprovincialis* population that underwent introgression by proto-*M. edulis*. The choice by Daguin & Borsa (2000) to use *M. edulis* as the underlying assumption of an outgroup for this hypothesis is violated by the hypothesis of introgression of *M. edulis* genes and the presence of *M. edulis*-like alleles in Australasian *M. galloprovincialis* (McDonald 1991, Daguin & Borsa 2000, Borsa et al. 2007). The phylogenetic relationships proposed by Hilbish et al. (2000), Daguin & Borsa (2000) and Gérard et al. (2008) necessitate the use of a fourth taxonomic group (free from past introgression with ingroup taxa) as an outgroup because the evolutionary relationships among three *M. galloprovincialis* haplogroups have yet to be unequivocally defined. Using *M. californianus* (Hilbish et al. 2000 and the current study) or monophyletic *M. trossulus* (Gérard et al. 2008) as the outgroup would keep assumptions satisfied while encompassing all possible evolutionary relationships among *M. galloprovincialis* and *M. edulis* haplogroups/lineages.

The difference between the current investigation and all other phylogenetic investigations of *M. edulis* complex taxa lies in the genealogical relationships among the three lineages of *M. galloprovincialis*; the hypotheses from the current study indicate that *M. edulis* and *M. galloprovincialis* from the "North Atlantic" group are the sister taxa to the Southern *M. galloprovincialis*, whilst Hilbish et al. (2000) hypothesize that sister taxon to be *M. galloprovincialis* of the "Mediterranean" group and Daguin & Borsa (2000) and Gérard et al. (2008) hypothesize that the two Northern lineages of *M. galloprovincialis* are

sister taxa and Southern hemisphere *M. galloprovincialis* divergence pre-dates the divergence of Northern hemisphere *M. edulis* and *M. galloprovincialis*.

The divergence of Southern hemisphere *M. galloprovincialis* could have occurred before major genetic mixing events between *M. galloprovincialis* and *M. edulis* in the Atlantic Ocean. The immediate shared ancestry of *M. edulis* and Southern hemisphere *M. galloprovincialis* is substantiated by molecular and biochemical investigations and it is the opinion of this author that a synthesis of investigations to date indicates the Southern *M. galloprovincialis* lineage arose from *M. galloprovincialis* containing *M. edulis* genes. The origin of *M. edulis* alleles could arise from two scenarios: either shared ancestry or introgression through secondary contact. In the shared ancestry scenario, *M. edulis* alleles would be present in *M. galloprovincialis* immediately following divergence in the Mediterranean Sea. Consequently, these alleles would have been lost through selection and/or drift in present-day Northern hemisphere *M. galloprovincialis* but remained in the Southern hemisphere. Introgression through secondary contact indicates *M. galloprovincialis* moving out of the Mediterranean Sea underwent introgressive hybridisation with *M. edulis* and soon after migrated to the Southern hemisphere.

The genetic background of *M. galloprovincialis* belonging to the shared *M. galloprovincialis*/*M. edulis* haplogroup has been subject to introgressive hybridisation for at least 1-2 million years. The coalescent based estimation of the divergence for Southern hemisphere *M. galloprovincialis* at 2.04-2.66 mya is roughly coincident (at the latest estimation) with current ideas of Northern hemisphere evolution. The Strait of Gibraltar first opened in the early Pliocene (close to 5.3 mya) when the Atlantic and Mediterranean met for the first time in 600K years following the Messinian Salinity Crisis (Loget & Van Den Dreissche 2006). In light of this evidence, it is possible the timing of the origin of *M. galloprovincialis* needs to be re-examined; the earliest possible divergence is 5.3 mya. if

blue mussels were distributed along Atlantic Ocean coastlines at that time or around the same time that blue mussels migrated to the Atlantic.

The likelihood-based hypothesis testing of alternate topologies applied to the current single gene COIII and 16s data sets help to reinforce genealogical relationships among the three *M. galloprovincialis* lineages. The maximum-likelihood 16s and COIII single gene topologies withstood most attempts to rearrange the genealogical relationships among the three *M. galloprovincialis* lineages into (1) *M. galloprovincialis* "Mediterranean" and Southern lineages as sister groups (from Hilbish et al. 2000) and (2) *M. galloprovincialis* "Mediterranean" and "North Atlantic" lineages as sister groups with the divergence of Southern hemisphere *M. galloprovincialis* pre-dating the divergence of Northern hemisphere *M. galloprovincialis* and *M. edulis* (from Gérard et al. 2008 and Daguin & Borsa 2000).

The most supported reconstruction in this investigation is from combining loci, indicating the *M. galloprovincialis* "North Atlantic" lineage is sister to the *M. galloprovincialis* Southern lineage. From a purely ecological standpoint, this may seem the most parsimonious relationship. Both Southern and "North Atlantic" *M. galloprovincialis* lineages are distributed in colder waters than the "Mediterranean" lineage, and it would be parsimonious to assume the most recent common ancestor of the Southern *M. galloprovincialis* and *M. galloprovincialis*/*M. edulis* haplogroups was also cold water adapted. This observation also supports the idea of a *M. galloprovincialis* ancestor that contained some *M. edulis* genes either through introgression or shared ancestry.

4.5.3 Future Considerations

To rectify the genealogical relationships among the three lineages of *M. galloprovincialis*, one or more nuclear genes should be added to the substantial amount of mtDNA sequences available and incorporated into future phylogenetic analyses. As stated by Gérard et al. (2008), skewed taxonomic thinking based on nuclear markers considered diagnostic for Northern hemisphere species is also an important consideration. Future investigations simply relying on comprehensive phylogenetic topologies will reduce the bias inherent in the application of single gene markers from Northern hemisphere species. Incomplete or differential lineage sorting from incongruent single gene trees could be overcome by additional loci.

Moving away from the bifurcating nature of phylogenetic trees and into the realm of networks is also promising as lineages arising from the hybridisation of two parent lineages are poorly represented by bifurcations. Networks allow the assumption of lineage independence to be broken and would be suitably applied to *Mytilus edulis* complex taxa to discern phylogenetic and genealogical relationships.

4.6 Acknowledgements

4.7 List of References

Akaike H (1981) Likelihood of a model and information criteria. *Journal of Econometrics* 16: 3-14.

Apte S, Holland BS, Goodwin LS, Gardner JPAG (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Barsotti G, Meluzzi C (1968) Osservazioni su *Mytilus edulis* e *M. galloprovincialis* Lmk. *Conchiglie* 4: 50–58.

Borsa P, Daguin C, Bierne M (2007) Genomic reticulation indicates mixed ancestry in Southern-hemisphere *Mytilus* spp. *Biological Journal of the Linnean Society* 92: 747-754.

Branch GM, Steffani CM (2004) Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology* 300: 189-215.

Buckley TR (2002) Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Systematic Biology* 51: 509-523.

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42: 384–397.

Burzynski A, Zbawicka M, Skibinski DOF, Wenne R (2003) Evidence for recombination of mtDNA in the marine mussel Mytlius *trossulus* from the Baltic. *Molecular Biology and Evolution* 20: 388-392.

Chippindale PT, Wiens JJ (1994) Weighting, partitioning, and combining characters in phylogenetic analysis. *Systematic Biology* 43: 278–287.

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

Cunningham CW (1997) Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 7: 733-740.

Daguin C, Borsa P (2000) Genetic relationships of *Mytilus galloprovincialis* Lmk. populations world-wide: evidence from nuclear-DNA markers. In: Harper EM, Taylor JD, Graeme JA, eds. *The evolutionary biology of bivalvia*. Special publication no. 177. London: Geological Society, London, p. 389-397.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: 699-710.

Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachussets, p. 540.

Filipowicz M, Burzynski A, Smietkana B, Wenne R (2008) Recombination in mitochondrial DNA of European mussels *Mytilus*. *Journal of Molecular Evolution* 67: 377-388.

Fleming CA (1959) Notes on New Zealand recent and tertiary mussels (Mytilidae). *Transactions of the Royal Society of New Zealand* 87: 165-178.

Gardner JPA (2004) A historical perspective of the genus *Mytilus* (Bivalvia: Mollusca) in NZ: multivariate morphometric analysis of fossil, midden and contemporary blue mussels. *Biological Journal of the Linnean Society* 82: 329-344.

Geller JB, Carlton JT, Powers DA (1994) PCR-based detection of mtDNA haplotypes of native and invading mussels on the northeastern Pacific coast: latitudinal pattern of invasion. *Marine Biology* 119: 243-249.

Genta Iturrería SF, Griffin M, Pérez LM (2009) Redescription of the Genus *Modiomytilus* Griffin. From Reunión Anual de Comunicaciones de la Asociación Palontológica Argentina, Museo Argentino de Ciencias Naturales, 2009.

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

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652-670.

Grant WS, Cherry MI (1985) *Mytilus galloprovincialis* Lmk. in southern Africa. *Journal of Experimental Marine Biology and Ecology* 90: 179-191.

Griekspoor A, Groothuis T. 4 Peaks: For Peaks, Four Peaks. Available from: www.mekentosj.com/science/4peaks/

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160-174.

Hilbish TJ, Mullinax A, Dolven SI, Meyer A, Koehn RK, Rawson PD (2000) Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration. *Marine Biology* 136: 69-77.

Hillis DM, Moritz C, Mabel BK, eds (1995) *Molecular Systematics, 2nd edition*. Sinauer Associates, Sunderland, Massachusetts, p. 582.

Hoeh WR, Stewart DT, Sutherland GW, Zouros E (1996) Multiple origins of gender-associated mitochondrial DNA lineages in bivalves (Mollusca: Bivalvia). *Evolution* 50: 2276-2286.

Hoffman RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organisation for the blue mussel, *Mytilus edulis*. *Genetics* 131: 397-412.

Hope JR, Lampert RJ, Edmondson E, Smith MJ, van Tets GE (1977) Late Pleistocene faunal remains from Seton rock shelter, Kangaroo Island, South Australia. *Journal of Biogeography* 4: 363-385

Huelsenbeck JP, Bull JJ, Cunningham CW (1996) Combining data in phylogenetic analysis. *Trends in Ecology and Evolution* 11: 152-158.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

Huelsenbeck JP, Ronquist F (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

Inoue K, Waite JH, Matsouka M, Odo S, Harayama S (1995) Interspecific variations in adhesive protein sequences of *Mytilus edulis, M. galloprovincialis,* and *M. trossulus*. *Biological Bulletin* 189: 370-375.

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

Kijewski T, Zbawicka M, Väinölä R, Wenne R (2006) Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biolpgy* 149: 1371-1385.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications* 13: 235-248.

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution* 31: 151-160.

Krapivka S, Toro JE, Alcapán AC, Astorga M, Presa P, Pérez M, Guiñez R (2007) Shell-shape variation along the latitudinal range of the Chilean blue mussel *Mytilus chilensis* (Hupe 1854). *Aquaculture Research* 38: 1770–1777.

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

Lee SY, Morton BS (1985) The introduction of the Mediterranean mussel *Mytilus galloprovincialis* into Hong Kong. *Malacological Review* 18: 107–109

Loget N, Van Den Driessche J (2006) On the origin of the Strait of Gibraltar. *Sedimentary Geology* 188-189: 341-356.

Lowe S, Browne M, Boudjelais S, De Poorter M (2000) 100 of the world's worst invasive alien species: A selection from the global invasive species database. Published by

The Invasive Species Specialist Group (ISSG), a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), pp. 12.

McDonald JH, Koehn RK (1988) The mussels *Mytilus galloprovincialis* and *M. trossulus* on the Pacific coast of North America. *Marine Biology* 99: 111-118.

McDonald JH, Seed R, Koehn RK (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology* 111: 323-333.

Marincovich L Jr, Gladenkov AU (1999) Evidence for an early opening of the Bering Strait. *Nature* 397: 149-151.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* 46: 523-536.

Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis. Version 2.73 http://mesquiteproject.org

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18: 269-292.

Nielson R, Wakeley J (2001) Distinguishing migration from isolation: a markov chain monte carlo approach. *Genetics* 158: 885-896.

Palumbi SR (1996) Nucleic Acids II: the polymerase chain reaction. In: Hillis DM, Moritz C, Mabel BK, eds. *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts, p. 205-247.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Quesada H, Beynon CM, Skibinski DOF (1995a) A mitochondrial DNA discontinuity in the mussel *Mytlius galloprovincialis* (Lmk.): Pleistocene vicariance biogeography and secondary intergradation. *Biological Journal of the Linnean Society* 12: 521-524.

Quesada H, Zapata C, Alvarez A (1995b) A multilocus allozyme discontinuity in the mussels *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. *Marine Ecology Progress Series* 116: 99-115.

Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer.

Rawson PD, Hilbish TJ (1995) Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution* 12: 893-901.

Rawson PD, Hilbish TJ (1998) Asymmetric introgression of mitochondrial DNA among European populations of blue mussels (*Mytilus* spp). *Evolution* 52: 100-108.

Rawson PD, Joyner K, Hilbish TJ (1996) Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity* 77: 599-607.

Rawson PD, Agrawal V, Hilbish TJ (1999) Hybridisation between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Marine Biology* 134: 201-211.

Riginos C, Sukhdeo K, Cunningham C (2002) Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution* 19: 347-351.

Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* 58: 2438-2451.

Ronquist F, Huelsenbeck JP, van der Mark P (2005) MrBayes v3.1 manual. Available from http://mrbayes.csit.fsu.edu/manual.php.

Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337-1344.

Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6: 461-464.

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

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114-1116.

Sokolov, Eugene P (2000) An improved method for DNA extraction from molluscan tissues. *Journal of Molluscan Studies* 66: 573-575.

Soot-Ryen T (1955) A report on the family Mytilidae (Pelecypoda). Allan Hancock Pacific Expeditions 20: l-175.

Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E (1995) Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. *Molecular Biology and Evolution* 12: 735-747.

Suchanek TH, Geller JB, Kreiser BR, Mitton JB (1997) Zoogeographic distributions of the sibling species *Mytilus galloprovincialis* and *M. trossulus* (Bivalvia: Mytilidae) and their hybrids in the North Pacific. *Biological Bulletin* 193: 187-194.

Swofford DL (1998) PAUP*: Phylogenetic analysis using parsimony (*and other methods) 4.0 Beta. Sinauer Associates, Sunderland, Massachusetts.

Sugiura N (1978) Further analysis of the data by Akaike's information criterion and the finite correction. *Communications in Statistics A-Theories and Methods* 7: 13-26.

Tajima F. (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135: 599-607.

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

Toro JE (1998) PCR-based nuclear and mtDNA markers and shell morphology as an approach to study the Chilean blue mussel, *Mytilus chilensis* (Bivalvia). *Aquatic Living Resources* 11: 347-353.

Toro JE, Ojeda JA, Vergara AM, Castro GC, Alacapan AC (2005) Molecular characterisation of the Chilean blue mussel (*Mytilus chilensis* Hupe 1854) demonstrates evidence for the occurrence of *Mytilus galloprovincialis* in Southern Chile. *Journal of Shellfish Research* 24: 1117-1121.

Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnean Society* 43: 127-148.

Van Erkom Schurink C, Griffiths CL (1991) A comparison of reproductive cycles and reproductive output in four southern African mussel species. *Marine Ecology Progress Series* 76: 123-134.

Vermeij GJ (1991) Anatomy of an invasion: The trans-Arctic interchange. *Paleobiology* 17: 281-307

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

Westfall KM, Gardner JPA (2010) Taxonomic and genetic diversity of Southern hemisphere smooth-shelled blue mussels of the genus *Mytilus* (Mytilidae; Bivalvia) and the identification of Northern hemisphere invaders, in press *Biological Journal of the Linnean Society*.

Westfall KM, Wimberger PH, Gardner JPA (2010) An RFLP to determine if *Mytilus galloprovincialis* Lmk. (Mytilidae; Bivalvia) is of Northern or Southern hemisphere origin. *Molecular Ecology Resources* 10: 573-575.

Wilkins NP, Fujino K, Gosling EM (1983) The Mediterranean mussel *Mytilus galloprovincialis* Lmk. in Japan. *Biological Journal of the Linnean Society* 20: 365–374

Wonham MJ (2004) Mini Review: Distribution of the Mediterranean mussel *Mytilus galloprovincialis* (Bivalvia; Mytilidae) and hybrids in the Northeast Pacific. *Journal of Shellfish Research* 23: 535-543.

Xia X, Xie Z (2001) DAMBE: Data analysis in molecular biology and evolution. *Journal*

of Heredity 92: 371-373.

Yule U (1925) A mathematical theory of evolution, based on the conclusions of Dr. JC Willis… *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 213: 21-87.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas at Austin.

Chapter 5: Blue mussels of the genus *Mytilus* (Mytilidae; Bivalvia) exhibit macro- and micro-phylogeographic structure in the Southern hemisphere.

5.1 Abstract

The biogeographical distribution of many flora and fauna in the Southern hemisphere has resulted from both vicariant and dispersal processes. Long-lived planktonic larvae characteristic of many marine taxa may disperse across ocean basins through strong currents and rafting. To this effect, the West Wind Drift, driven by the Antarctic Circumpolar Current, is believed to mediate gene flow among populations separated by large oceanic distances across the Southern hemisphere. At the microgeographic scale within coastal waters of New Zealand, Chile and Australia, biogeographic boundaries that reduce or prevent gene flow have been identified for some nearshore marine invertebrate species. Blue mussels of the genus *Mytilus* are native to the Southern hemisphere and this distribution represents a unique lineage of *M. galloprovincialis*. The long-lived planktonic larval phase mediates the capacity for long-distance dispersal but previous phylogenetic analyses indicate there is no recent or ongoing gene flow among South America, New Zealand and Australia. Sequence data for two mtDNA cytochrome oxidase complex genes is used to reconstruct the biogeographic history across the Southern hemisphere and test hypotheses of intraregional phylogeographic barriers. Reciprocally monophyletic clades are associated with geographic placement but the order of Southern hemisphere colonization cannot be resolved based on inconsistent placement of South America or New Zealand in the basal position and overlapping confidence intervals of time to most recent common ancestor estimates. Recent and ongoing gene flow exists between the Kerguelen Islands and South America and is most likely mediated by the West Wind Drift whilst barriers to gene flow are detected across the Pacific Ocean and Tasman Sea. Fine-scale phylogeographic disjunction is evident in blue mussels from northwest and southwest regions of New Zealand. Regional population subdivision in this area is most likely a product of past gene flow that is not maintained in the contemporary setting.

5.2 Introduction

5.2.1 Southern hemisphere biogeographic hypotheses

The relative roles of vicariance and dispersal in shaping the distributions of species in the Southern hemisphere has been a topic of debate for over a century. With an increasing amount of molecular data becoming available, vicariance and dispersal theories have both been proven valid depending on the type of organism (animal vs. plant, terrestrial vs. marine) and age of Southern hemisphere occurrence. Occasionally, both theories can reasonably explain the evolutionary relationships among distributions within a single species on different spatiotemporal scales (e.g. Southern Beech tree *Nothofagus*, Linder & Crisp 1995).

Considering the geological time frame, tectonic plate vicariance relates contiguous biotic distribution on the Gondwanan supercontinent and sequential continental break-up to the evolutionary relationships among present day distributions. Vicariance is plainly observed for terrestrial taxa through the comparative phylogenetic approach. In the context of marine taxa with high dispersal capacity, molecular signatures of recent dispersal events or ongoing gene flow usually overshadow signatures of geologically old vicariance processes. Dispersal processes are therefore often invoked to explain the connectivity among taxa and among populations of a single, widely distributed species (Waters & Roy 2004a, de Queiroz 2005, Waters 2008).

Compared to tectonic plate movement, sea level changes alter the connectivity of water masses on much smaller time scales of 10,000 - 100,000 years (Jannson & Dynesius 2002). Reductions in sea levels associated with climatic cycles (i.e. Milankovitch Oscillations) may expose coastal geographic features (i.e. isthmuses, peninsulas, land bridges), which alter the connectivity of water masses and change the dispersal pathways of marine larvae (Jannson & Dynesius 2002, Waters et al. 2005). Conversely, increasing

162

sea levels can break up existing landmasses into series' of archipelagos or island chains (Stephens & Hogg 2004).

Dispersal theories relate present day biotic distributions to sequential unidirectional migration events across water bodies and may be a valid hypothesis to explain blue mussel distribution in the Southern hemisphere. The Atlantic, Indian and Pacific Ocean basins are connected by the Southern Ocean: it's major current system, the Antarctic Circumpolar Current (herein the ACC), is driven eastward by prevailing westerly winds (also called the West Wind Drift). The ACC formed approximately 33.5 mya, well-before the presence of *Mytilus* spp. in the Southern hemisphere, therefore all biogeographic hypotheses of blue mussel distribution and colonization pathways have to consider the relative impact of this major current. The ACC has been identified as a probable cause for dispersal of Asterinid sea stars (Waters and Roy 2004a, Waters 2008) with basal phylogenetic clades distributed in South Africa and derived geographically restricted monophyletic clades in eastern Australia.

Trans-Tasman (Sea) dispersal between New Zealand and Australia is variable among near-shore invertebrate taxa and may be related to larval duration. The genus *Nerita* exhibits ongoing Trans-Tasman gene flow between New Zealand and southeast Australian populations and has a larval duration of 5-6 months (Waters et al. 2005). Conversely, the genus *Austrolittorina* is thought to live in an active larval feeding phase for up to one month and exhibits no ongoing or recent gene flow between New Zealand and southeastern Australian populations (Waters et al. 2006). *Mytilus* spp. have a larval duration time similar to *Austrolittorina* and would therefore be predicted to show no recent or ongoing Trans-Tasman gene flow.

5.2.2 Global patterns of *Mytilus* dispersal and vicariance

Members of the genus *Mytilus* have a high dispersal potential through planktotrophic larvae that inhabit the water column from four to seven weeks and can delay metamorphosis (Bayne 1965) for up to 6 months (Lane et al. 1985) allowing for a great potential dispersal distance. Indeed, gene flow has been detected across the North Atlantic Ocean basin in M. *edulis* (Riginos et al. 2004) and intraspecific population differentiation is apparent on macro-geographic scales in western Europe (Smietkana et al. 2004 and 2009, black sea/med population separation).

Historically, the global antitropical distribution of *Mytilus edulis* species complex members (*M. edulis, M. galloprovincialis, M. trossulus*) is believed to have arisen through dispersal during the larval life history phase with the exception of *M. edulis* and *M. trossulus* divergence mediated by reduced water flow in the Bering Strait during Pleistocene glaciations (Riginos et al. 2004).

Vicariance has been shown to underlie intraspecific blue mussel population differentiation in one documented case (Quesada et al. 1995, Ladoukakis et al. 2002). Pleistocene glaciation caused reductions of sea levels in the North Atlantic Ocean and Mediterranean Sea ~2 mya, which altered the connectivity of these two basins by reducing water flow through the Strait of Gibraltar. *Mytilus* populations contiguously distributed in both regions experienced reduced genetic connectivity through this time period and exhibit population differentiation in present day mtDNA haplotype distributions (Quesada et al. 1995). Blue mussel populations in the Azov and Baltic Seas also display mtDNA differentiation and isolation but these patterns have been attributed more to dispersal and founder effects than vicariance (Smietkana et al. 2004).

5.2.3 Phylogenetics of the '*Mytilus edulis* species complex' global distribution

The order of colonisation inferred from topological branching patterns supports the theory of an Atlantic route of migration to the Southern hemisphere of a Northern hemisphere congener first proposed by Hilbish et al. (2000) and substantiated by this thesis (Chapter 4, Sections 4.4.4.2 to 4.4.4.5). The ancestor of Southern *Mytilus* most likely originated from the Atlantic Ocean or Mediterranean Sea and bypassed the African continent to colonize the coasts of South America. There is no fossil or contemporary evidence of natural blue mussel colonization of the African continent south of the equator and blue mussel presence in South Africa is solely due to human-mediated introduction (Grant & Cherry 1985; Gerard et al. 2008).

Previous attempts to date the divergence of Northern and Southern hemisphere *Mytilus* using fossil calibration have produced varying estimates from different genes: the 16s gene infers a divergence date of 1-2 mya (Hilbish et al. 2000) whilst the COI gene infers a divergence date of 0.84 mya (Gerard et al. 2008). The current thesis estimates the same divergence time in a coalescent framework and from the COIII+16s phylogeny at 2.04-2.66 mya and 2.18-2.26 mya, respectively (Chapter 4, Section 4.4.5.2).

Reciprocal monophyly of clades associated with continental distributions in South America (includes Kerguelen Islands for COI in Gerard et al. 2008), New Zealand and Australia is supported in single COI, COIII and combined COIII+16s (Australia absent) mtDNA phylogenetic analyses, but not in single 16s locus phylogenies (Figure 5.1). Limited gene flow inferred from the monophyletic property of cytochrome oxidase complex clades indicates that connectivity of *Mytilus* across the Southern Pacific Ocean and the Tasman Sea is negligible at present and has been for some time except for a connection between South America and the Kerguelen Islands. Monophyly also indicates complete lineage sorting within each continent and basal placement of the South

American/Kerguelen clade indicates a single colonization event from South America to Australasia or multiple colonization events followed by complete lineage sorting (Figure 5.1c, d).
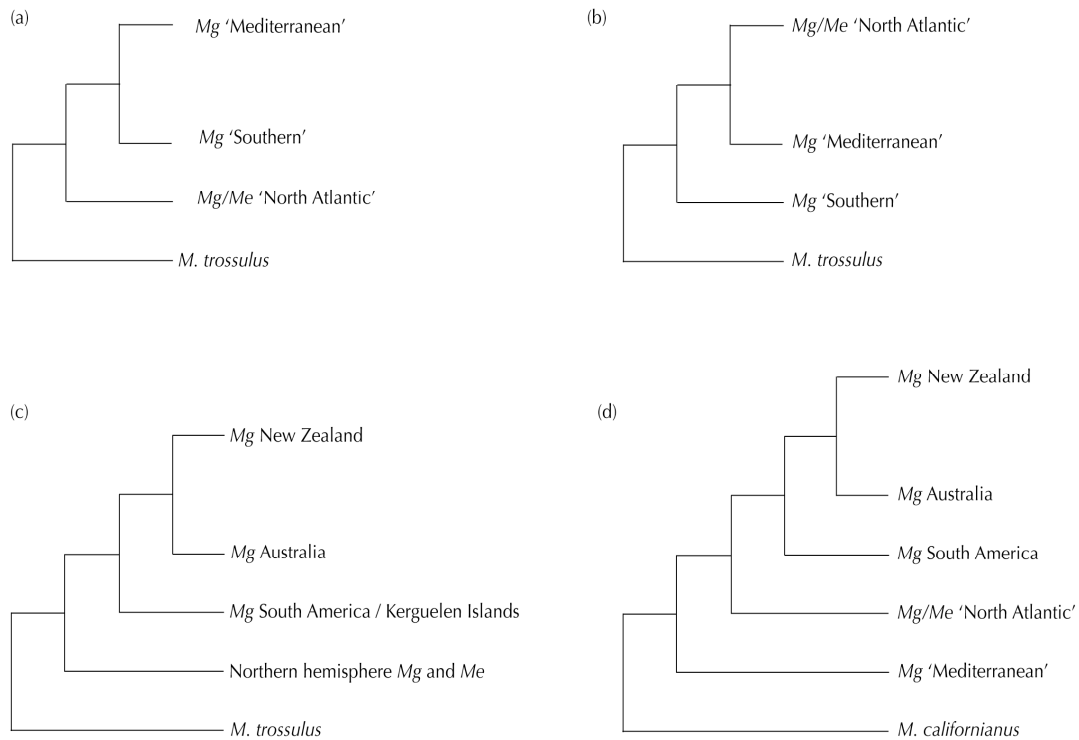
Figure 5.1 – Blue mussel backbone topologies from different genes and data sets.
(a) 16s Hilbish et al. 2000, (b) 16s Gerard et al. 2008, (c) COI Gerard et al. 2008, (d) COIII this thesis. Abbreviations are *Mg = M. galloprovincialis*, *Me = M. edulis*. 'Mediterranean' and 'North Atlantic' haplogroup designations from Rawson & Hilbish 1995 and 'Southern' refers to Southern hemisphere lineage. *M. trossulus* and *M. californianus* are outgroups. (The combined COIII + 16s topology is equal to (d) with Australia removed making *Mg* New Zealand and *Mg* South America sister clades).

5.2.4 The role of vicariance and dispersal in determining intraregional differentiation of *Mytilus*.

5.2.4.1 New Zealand

Some nearshore marine invertebrate species in New Zealand exhibit population level subdivision around 42ºS latitude on the South Island. Northern and Southern population subdivisions are observed in the green-lipped mussel *Perna canaliculus* (Apte & Gardner 2002, Star et al. 2003) and the cushion star *Patiriella regularis* (Waters & Roy 2004b, Ayers & Waters 2005). The most commonly cited reason for this disjunction is upwelling on the eastern coast of the South Island at the same latitude (Ross et al. 2009). An east-west phylogeographic break occurs at the north tip of the North Island that has been shown in the brooding amphipod *Paracorophium lucosi* (Stephens & Hogg 2004). Upwelling at the north tip of the North Island and sea level changes in the Miocene/Pliocene epochs changing the North Island into a series of archipelagos that later lost a marine connection have been cited as possible causes of this genetic disjunction (Ross et al. 2009). Both of these hypotheses will be examined with respect to population subdivision in New Zealand *Mytilus*.

The distribution of *Mytilus* in New Zealand spans 35º – 52º S latitudes and includes the South Island, offshore islands and two pockets on the North Island: the Bay of Islands and the South coast, separated by ~1000 km. coastal distance (Figure 5.2). Although mussels have been found in the Hauraki Gulf, this distribution is mainly composed of non-native individuals (93%, this thesis). The importance of this population as a possible range expansion and hybrid zone for blue mussels in northern New Zealand should be explored in future investigations.

167

5.2.4.2 Australia

An east-west phylogeographic split identified in near-shore marine invertebrate taxa in southern Australia has been linked to Pleistocene low sea level exposure of the Bassian isthmus connecting Tasmania to southern Australia (Waters et al. 2005, Waters et al. 2006, Ayre et al. 2009). Population subdivision for broad-cast spawning marine invertebrates caused by changes in water flow during the time the Bassian Isthmus was exposed is still detected in present day mtDNA sequences and represents ongoing barriers to gene flow between eastern and western populations split around 146oW (Waters et al. 2005, Waters et al. 2006, Ayre et al. 2009).

5.2.4.3 South America

The Chilean coastline inhabited by blue mussels is divided into two biogeoclimatic regions with associated environmental variation important to marine taxa: (1) the Central coast region defined by average sea surface temperature (SST) from ~6º - 11ºC (2) the Fjords coast region defined by SST from ~3º - 6ºC (Figure 5.2b) and steep salinity clines within the fjords (Fernandez et al. 2000). Nearshore marine invertebrates exhibit differentiation between two biogeographical regions that roughly align to the biogeoclimatic region boundaries: (1) Peru-Chile and (2) Magellanic (Brattström & Johannsen 1983, Lancelloti & Vasquez 1999) (Figure 5.2b). Despite the importance of biogeographical regions to enforce population differentiation in marine invertebrates, allozymic variation of *M. chilensis* revealed no population differentiation along the Chilean distribution of this species along an 1800-kilometer sample distribution (Toro 2006). A moderate isolation by distance relationship was identified between Punta Arenas (55ºS) and the rest of the sample distribution (to 36ºS) (Toro 2006).

In and around South America, *Mytilus* are distributed throughout the entire Falkland Islands and much of Argentina and Chile reaching a northern limit at ~30ºS. The taxonomic status of the Chilean blue mussel, *M. chilensis*, has been reviewed in this thesis and confirmed as the Southern hemisphere lineage of *Mytilus*, therefore, all mussels considered within this study are ascribed to a single species.

5.2.5 Current investigation

The aim of this investigation is to determine the evolutionary relationships of reciprocally monophyletic clades associated with continental distributions (New Zealand, Australia and South America) (Gerard et al. 2008) and infer a likely scenario for blue mussel colonization and biogeographic distribution in the Southern hemisphere by comparison to other marine taxa with similar life history characteristics. At a smaller spatial scale, phylogeographical relationships within South America, New Zealand and Australia will be explored through nested clade phylogeographical analyses and sets of AMOVA analyses. Intraregional genetic differentiation of *Mytilus* will be compared to genetic disjunction formed by phylogeographical breaks previously identified for marine taxa in South America, New Zealand and Australia.

5.3 Materials & Methods

5.3.1 Sampling regime

Blue mussels were sampled from New Zealand, Chile, Argentina, the Falkland Islands and Australia for the COIII gene (Figure 5.2) and from New Zealand, Chile, Australia and the Kerguelen Islands (Figure 5.2) by Gerard et al. (2008) for the COI gene (Table 5.1). Between two loci, this sample distribution encompasses almost the entire Southern hemisphere distribution of native mussels except for a few oceanic islands. Note

that blue mussels in South Africa constitute a completely non-native population (Gerard et

al. 2008) and will not be considered further here.

Table 5.1 – Southern hemisphere locations, geographical coordinates and codes for COIII and COI (Gerard et al. 2008) sample sites.

Gene	Country	Location	Sample Size	Latitude	Longitude	Code
COIII	New Zealand	TeTii Bay	3	35º 08.768' S	174º 00.258' E	TETI
		Dove's Bay	2	35º 11.789' S	174º 01.860' E	DOVB
		Russell Harbour	1	35º 15.571' S	174º 07.195' E	RUSH
		Opua Marina	1	35º 18.668' S	174º 07.208' E	OPUM
		Maud Island	6	41º 01.505' S	173º 53.764' E	MAUI
		Mussel Point	10	41º 43.989' S	174º 16.052' E	MUPO
		Wellington Harbour	10	41º 17.000' S	174º 50.000' E	WLGH
		Chatham Islands	13	43º 35.956' S	176º 39.939' W	CHAI
		Riverton	8	46º 21.084' S	168º 0.875' E	RIVE
		Ringitingi Beach	16	46º 53.259' S	167º 59.907' E	RINB
		Auckland Island	10	50º 29.391' S	166º 16.767' E	AUCI
		Campbell Island	6	52º 31.600' S	169º 06.835' E	CAMI
	Chile	Concepcion	20	36º 44.003' S	73º 07.830' W	CHCO
		Golfo Trinidad	7	49º 58.557' S	75º 12.085' W	CHGT
	Argentina	Ushuaia	9	54º 48.136' S	68º 15.483' W	ARUS
	Falkland Is.	Stanley Wharf	2	51º 41.573' S	57º 49.283' W	FIST
	Australia	Wonboyn River	2	37º 15.033' S	149º 57.556' E	WONR
		Melbourne	20	37º 53.532' S	144º 54.910' E	MELB
		Port Arthur	20	43º 08.489' S	147º 51.397' E	PORA
COI	New Zealand	Wellington Harbour	10	41º 37.000' S	174º 48.000' E	WHB
		Dunedin	10	45º 54.000' S	170º 30.000' E	DUN
		George Sound	10	44º 54.000' S	167º 20.000' E	GRS
	Chile	Maullin	10	41º 27.000' S	73º 37.000' W	MAU
		Canal de las Montañas	39	51º 51.000'S	73º 20.000' W	PAT
	Kerguelen Is.	Kerguelen Islands	99	49º 20.000' S	70º 10.000' E	KER
	Australia	Hobart	9	42º 50.000' S	147º 20.000' E	HOB
		Simpson's Bay	6	43º 17.000' S	147º 18.000' E	SIM
		Cloudy Bay Lagoon	9	43º 26.000' S	147º 13.000' E	CBL
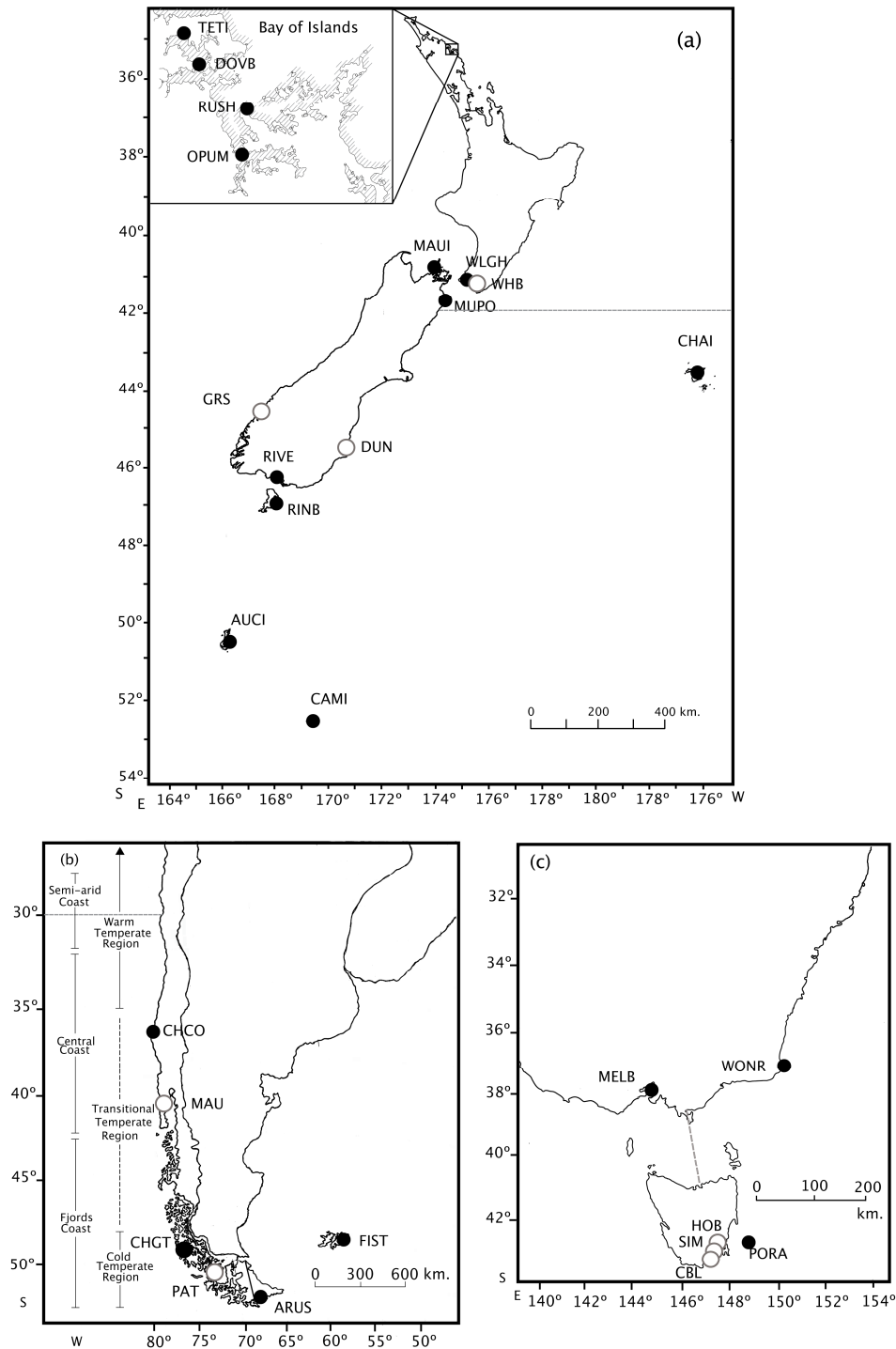
Figure 5.2 – Southern hemisphere blue mussel sample sites for COIII (filled circle) and COI (empty circle) haplotypes.
Kerguelen Islands are not pictured but one sample site occurs there. Biogeographic boundaries identified in other marine taxa are represented by dashed lines. An additional dashed line at ~30ºS in Chile represents the northern limit of *Mytilus* distribution. Codes for locations are given in Table 5.1.

171

5.3.2 Data set construction

Methods for sequence acquisition and alignment for the cytochrome oxidase subunit III gene is described in Chapter 4, Section 4.3.3. Southern hemisphere origin haplotypes from South America (Chile, Argentina and Falkland Islands), New Zealand and Australia identified as Southern lineage *Mytilus* through Bayesian and Maximum Likelihood phylogenetic reconstruction methods (refer to Chapter 4, Figures 4.3 to 4.5) were only used in the following analyses to prevent specious inference due to contemporary introduction of non-native *Mytilus* (Table 5.2). When taxonomic status disagreed among methods or data sets the haplotype in question was not used. Because some sample locations were removed based on the taxonomic identity of haplotypes, updated maps with the locations of individuals used in the current study are in Figure 5.2.

The COIII data set consisted of 79 unique haplotypes plus the outgroup haplotype *M. californianus* (collected form Otter Point, British Columbia, Canada, Genbank accession HQ336052) (Table 5.2). Haplotype nomenclature is composed of three parts consistent with previous haplotype nomenclature from Chapter 4: the first part is a one-letter species identifier (although only *M. galloprovincialis* is g used here), the second part is a two-letter location identifier based on where the haplotype was identified (Table 5.2).

Table 5.2 – COIII haplotype codes, locations and Genbank accession numbers used for phylogenetic reconstruction.

Country	Haplotype	Location(s)	Genbank Accession no.
New Zealand	gNZ 1	RINB (3), CHAI (4), AUCI	HQ336119
	gNZ 2	RINB	HQ336120
	gNZ 3	RINB	HQ336121
	gNZ 4	RINB, CHAI (3)	HQ336122
	gNZ 5	RINB	HQ336123
	gNZ 6	RINB (6), CAMI (3), RIVE, WLGH (4)	HQ336124
	gNZ 7	RINB	HQ336125
	gNZ 8	RINB	HQ336126
	gNZ 9	RINB	HQ336127
	gNZ 10	MUPO, RIVE	HQ336128

Country	Haplotype	Location(s)	Genbank Accession no.
	gNZ 12	MUPO	HQ336129
	gNZ 14	MUPO	HQ336131
	gNZ 15	CHAI	HQ336132
	gNZ 16	CHAI (2)	HQ336133
	gNZ 17	CHAI	HQ336134
	gNZ 18	CHAI	HQ336135
	gNZ 19	AUCI	HQ336136
	gNZ 20	AUCI	HQ336137
	gNZ 22	AUCI	HQ336139
	gNZ 26	CAMI	HQ336143
	gNZ 27	CAMI	HQ336144
	gNZ 28	CAMI	HQ336145
	gNZ 29	RIVE	HQ336146
	gNZ 31	RIVE	HQ336148
	gNZ 32	RIVE	HQ336149
	gNZ 33	MAUI	HQ336150
	gNZ 35	WLGH	HQ336151
	gNZ 36	WLGH	HQ336152
	gNZ 39	WLGH	HQ336154
	gNZ 46	TETI (2), DOVB	HQ336159
	gNZ 47	TETI	HQ336160
	gNZ48	OPUM	HQ336161
	gNZ 49	RUSH	HQ336162
Australia	gAU 3	MELB	HQ336164
	gAU 4	MELB	HQ336165
	gAU 5	MELB	HQ336166
	gAU 6	MELB	HQ336167
	gAU 7	MELB	HQ336168
	gAU 8	MELB	HQ336169
	gAU 10	MELB	HQ336170
	gAU 11	MELB	HQ336171
	gAU 12	MELB	HQ336172
	gAU 13	MELB	HQ336173
	gAU 14	MELB	HQ336174
	gAU 15	MELB	HQ336175
	gAU 16	MELB	HQ336176
	gAU 18	PORA (6)	HQ336178
	gAU 19	PORA (2)	HQ336179
	gAU 21	PORA	HQ336181
	gAU 22	PORA (2)	HQ336182
	gAU 23	PORA	HQ336183
	gAU 24	PORA	HQ336184
	gAU 25	PORA	HQ336185
	gAU 26	PORA	HQ336186
	gAU 27	PORA	HQ336187
	gAU 28	PORA	HQ336188
	gAU 29	WONR	HQ336189
Argentina	gAR 1	ARUS	HQ336191
	gAR 2	ARUS (2)	HQ336192
	gAR 3	ARUS	HQ336193
	gAR 4	ARUS	HQ336194
	gAR 5	ARUS	HQ336196
Chile	gCH 1	CHCO (2)	HQ336198

Country	Haplotype	Location(s)	Genbank Accession no.
	gCH 2	CHCO (2)	HQ336199
	gCH 3	CHCO, CHGT	HQ336200
	gCH 4	CHCO	HQ336201
	gCH 5	CHCO	HQ336202
	gCH 6	CHCO	HQ336203
	gCH 7	CHCO	HQ336204
	gCH 8	CHCO	HQ336205
	gCH 14	CHCO	HQ336208
	gCH 15	CHGT	HQ336209
	gCH 16	CHGT	HQ336210
	gCH 17	CHGT	HQ336211
	gCH 18	CHGT (2)	HQ336212
Falkland Islands	gFI 1	FIST	HQ336216
	gFI 2	FIST	HQ336217
Argentina & Chile	gSA 1	ARUS (2), CHCO	HQ336195
Argentina & Chile	gSA 2	ARUS, CHGT	HQ336197
New Zealand & Chile	gSH 2	AUCI (3), CHCO	HQ336138

Location codes can be interpreted from Table 5.2. The number in brackets following a location code is the absolute number of individuals with that haplotype.

A second data set was constructed from COI sequences derived from Gerard et al. (2008) (Table 5.3) consisting of 49 unique haplotypes. Sequences sampled from Australia (Tasmania), New Zealand and Chile together form a monophyletic Southern hemisphere clade, which is further separated into monophyletic clades associated with the country of origin (Gerard et al. 2008). The outgroup used with this data set is *M. trossulus* (Genbank Accession U68744.1 GI:2501772) (Gerard et al. 2008). Gerard et al. (2008) did not ascribe species status to Southern hemisphere *Mytilus* spp. The individuals from whom these sequences were derived were not available for taxonomic identification therefore the taxonomic status is derived from geographical context and the congruent reciprocally monophyletic property of Southern hemisphere clades with COIII data. It is assumed these individuals represent *Mytilus* of the Southern hemisphere lineage. Mussels from Dichato, Chile and Nedlands, Australia resolved onto the Northern hemisphere portion of the phylogenetic tree (Gerard et al. 2008) and were not included in this study.

Table 5.3 – COI haplotype codes, locations and Genbank accession numbers used for phylogenetic reconstruction from Gerard et al. (2008).

Country	Haplotype	Location (s)	Genbank accession no.
New Zealand	WHB3	Wellington Harbour	AM905151
	WHB4	Wellington Harbour	AM905152
	WHB5	Wellington Harbour	AM905153
	WHB6	Wellington Harbour	AM905154
	DUN1	Dunedin	AM905146
	DUN3	Dunedin	AM905148
	GRS2	George Sound	AM905156
	GRS3	George Sound	AM905157
	GRS4	George Sound	AM905158
	GRS5	George Sound	AM905159
	GRS6	George Sound	AM905160
	NZ1	Wellington Harbour, Dunedin, George Sound	AM905149, AM905147, AM905155
Australia (Tasmania)	CBL1	Cloudy Bay Lagoon	AM905161
	CBL3	Cloudy Bay Lagoon	AM905163
	CBL4	Cloudy Bay Lagoon	AM905164
	SIM2	Simpson's Bay	AM905166
	SIM3	Simpson's Bay	AM905167
	HOB1	Hobart	AM905168
	HOB2	Hobart	AM905169
	HOB3	Hobart	AM905170
	HOB4	Hobart	AM905171
	HOB5	Hobart	AM905172
	HOB6	Hobart	AM905173
	AU1	Cloudy Bay Lagoon, Simpson's Bay	AM905162, AM905165
Chile	MAU2	Maullin	AM905181
	MAU4	Maullin	AM905183
	MAU5	Maullin	AM905184
	MAU6	Maullin	AM905185
	MAU7	Maullin	AM905186
	PAT3	Canal de las Montañas (Patagonia)	AM905189
	PAT7	Canal de las Montañas (Patagonia)	AM905193
	PAT8	Canal de las Montañas (Patagonia)	AM905194
	PAT9	Canal de las Montañas (Patagonia)	AM905195
Kerguelen Islands	KER4	n/a	AM905199
	KER5	n/a	AM905200
	KER8	n/a	AM905203
	KER9	n/a	AM905204
	KER10	n/a	AM905205
	KER11	n/a	AM905206
	KER12	n/a	AM905207
	KER13	n/a	AM905208
	KER14	n/a	AM905209
	KER15	n/a	AM905210
	KER16	n/a	AM905211
South America	CHK1	Canal de las Montañas (Patagonia), Kerguelen Islands	AM905191, AM905202

Country	Haplotype	Location (s)	Genbank accession no.
	CHK2	Canal de las Montañas (Patagonia), Kerguelen Islands	AM905190, AM905201
	CHK3	Canal de las Montañas (Patagonia), Kerguelen Islands	AM905192, AM905196
	CHK4	Maullin, Canal de las Montañas (Patagonia), Kerguelen Islands	AM905180, AM905188, AM905198
	CHK5	Maullin, Canal de las Montañas (Patagonia), Kerguelen Islands	AM905182, AM905187, AM905197

Note – some haplotypes were shared within New Zealand, within Australia and between Chile and the Kerguelen Islands. Multiple Genbank accession numbers for one unique haplotype and notation CHK reflects Chile/Kerguelen origin. The three-letter haplotype codes (WHB, DUN, GRS, CBL, SIM, HOB, MAU, PAT, KER) reflect sample location map codes (Figure 5.2).

5.3.3 Bayesian phylogenetic reconstruction

The program MrBayes v3.1.2 (Huelsenbeck & Ronquist 2001) was used to analyse the single COI and COIII data sets each with two simultaneous independent runs after optimisation of MC^3 operators. The model of nucleotide substitution was determined for the COI and COIII data sets (ModelTest v3.7, Posada & Crandall 1998) under the Bayesian Information Criterion (BIC) (Schwarz 1978) assumed *a priori* as the model with the largest posterior probability. Default values of uninformative (flat) prior distributions are specified *a priori*.

Convergence was assessed in two ways: (1) within each run by the effective sample size (ESS) of each parameter of interest using Tracer v1.4.1 (Rambaut & Drummond 2003) (minimum ESS=300) and, (2) between runs with the average standard deviation of clade probabilities (referred to as split frequencies) tending towards zero and the potential scale reduction factor of continuous variables tending towards 1 after convergence. After several initial runs to achieve convergence on the posterior distribution, the following MC^3 operators were applied to subsequent analyses: 10,000,000 generations, 8 chains, 2 swaps per generation and a chain temperature of 1.00. Independent run results were summarised

into a maximum clade credibility tree (total tree probability > 0.95) with nodal support (posterior probability) estimated for each clade.

The program BEAST v1.5beta2 (Drummond & Rambaut 2007) was used to estimate node times in COI and COIII data sets. Two divergence estimates of Northern and Southern hemisphere *Mytilus* were obtained from previous analyses based on (1) a coalescent model = 2.04-2.66 mya and (2) a fossil calibration node = 2.18-2.26 mya (Chapter 4, Section 4.4.5.2). Because estimates are so close, the estimate obtained from the coalescent model was used for both data sets in this instance under a normal prior distribution (2.35 ± 0.31SE). A monophyletic constraint was placed on ingroup haplotypes and the calibration point set at the T_{MRCA} of the ingroup. Two independent runs were performed with the following parameters: 10,000,000 generations sampled every 1,000, 25% burn-in, coalescent (constant size) tree prior, the BIC best fit model of nucleotide substitution as identified by ModelTest and a relaxed uncorrelated lognormal molecular clock where the mean substitution rate was not fixed.

5.3.4 Nested clade phylogeographic analysis (NCPA)

Testing phylogeographic patterning of COI and COIII haplotypes within the Southern hemisphere and within each reciprocally monophyletic region was carried out using NCPA. The program TCS v1.0 (Clement et al. 2000) was used to create single 95% probability parsimony networks of COI and COIII haplotypes (Templeton et al. 1992). Nesting of haplotypes into hierarchical clades follows the rules of Templeton et al. (1987) and ambiguity rules of Templeton and Sing (1993). Testing the null hypothesis of random geographic distribution in all clades within the nested hierarchy was performed using Geodis V2.0 (Posada et al. 2000) and significance assessed using the Dunn-Sidak correction for multiple comparisons. Where the null hypothesis of random association is

rejected, the updated inference key (http://darwin.uvigo.es/download/geodisKey_28Apr09.pdf) originally provided by Templeton et al. (1995) identified potential evolutionary processes occurring for the individuals or regions (represented by individuals) of interest. Congruence between COI and COIII networks with respect to inference of evolutionary process lends support to hypotheses and minimizes the chance of false positives (Templeton 2004). Minimum oceanic distance is used instead of great circle distance to mimic connectivity of a marine dispersing organism (Google Earth v1.2) (Tables 5.4 and 5.5).

Table 5.4 – Pair-wise minimum oceanic distances (km.) between COIII Australian sample locations and COI Southern hemisphere locations.

Data Set	Location	Location	Distance (km.)
COIII	WONR	MELB	584.6
	WONR	PORA	646.0
	MELB	PORA	741.5
COI	MAU	KER	5963.0
	MAU	PAT	770.1
	KER	PAT	5192.1
	CBL	SIM	26.8
	CBL	HOB	53.9
	SIM	HOB	29.5
	WHB	GRS	584.4
	WHB	DUN	375.4
	GRS	DUN	309.0

Codes for COIII and COI sample locations can be found in Tables 5.2 and 5.3, respectively. COI sample location distances are given for sample sites within countries because the haplotype network 95% parsimony limit is reached within each country.

Table 5.5 – Pairwise minimum oceanic distance (km.) above the diagonal between COIII study sites (New Zealand and South America).

	TETI	DOVB	RUSH	OPUM	MAUI	WLGH	MUPO	CHAI	RIVE	RINB	AUCI	CAMI	CHCO	CHGT	ARUS	FIST
TETI	-	5.5	12.8	21.7	933.8	997.9	1147.5	1316.2	1892.8	1933.8	2102.2	2304.2	6166.3	5195.7	5259.5	5706.3
DOVB		-	11.2	16.2	927.4	992.4	1142.0	1310.7	1885.7	1928.3	2096.7	2298.6	6159.4	5188.7	5256.6	5699.4
RUSH			-	6.5	925.3	1002.0	1138.1	1304.9	1885.9	1924.4	2093.5	2295.3	6162.0	5191.3	5255.2	5701.9
OPUM				-	930.7	1010.9	1147.1	1313.8	1892.0	1933.3	2102.4	2304.2	6163.9	5193.2	5257.1	5703.9
MAUI					-	129.4	131.3	896.2	986.9	921.4	1303.6	1394.8	5691.3	4943.1	4980.7	5393.3
WLGH						-	72.9	763.9	865.9	859.2	1235.9	1324.4	5561.9	4813.7	4851.3	5263.9
MUPO							-	790.8	800.1	786.3	1178.7	1272.5	5560.3	4810.9	4842.7	5270.8
CHAI								-	1262.9	1258.6	1498.5	1436.0	5100.0	4385.5	4448.0	4895.5
RIVE									-	68.2	510.8	689.3	5566.7	4753.0	4733.9	5127.0
RINB										-	438.6	627.6	5540.4	4714.7	4703.7	5084.4
AUCI											-	297.2	5412.4	4564.6	4525.9	4898.7
CAMI												-	5231.8	4385.2	4351.3	4726.7
CHCO													-	1641.9	2310.9	2271.6
CHGT														-	668.9	1629.6
ARUS															-	807.9
FIST																-

Codes for COIII sample locations can be found in Table 5.2.

5.3.5 Intraregional phylogeographic hypothesis testing

Hypothesis testing of previously identified phylogeographic breaks within the regions of New Zealand, South America (Chile) and Australia was performed for both COI and COIII data sets. COI haplotypes unique to the Kerguelen Islands were not used because this location is not included the spatial scale of phylogeographic breaks observed in other nearshore marine invertebrate species. Analyses of molecular variance (AMOVA) (Excoffier et al. 1992) were carried out in Arlequin v2.001 (Schneider et al. 2000) under the infinite alleles model. Statistical significance was assessed with 10,000 non-parametric permutations. The term 'region' refers to New Zealand, South America or Australia. The term 'location' refers to the sample locations contained within each region. For haplotypes shared between locations and/or regions one copy was placed within each hypothetical group to represent the number of individuals that haplotype was identified from. For haplotypes that were identified in more than one individual at a single location, the number of haplotypes specified in the AMOVA analysis represents the absolute number of individuals with that haplotype.

5.3.5.1 New Zealand distribution

Two phylogeographic breaks previously identified for nearshore marine invertebrate species (Ross et al. 2009) were tested as *a priori* haplotype groupings with AMOVA analyses in Arlequin. The groups corresponding to phylogeographic breaks form the following hypotheses that predict more genetic variation is contained between groups than within groups: (1) a division at 42ºS to give Northern and Southern groups (2) a division around the east cape of the North Island to give northwest and southeast groups (including all South Island and offshore locations) (Table 5.6). Refer to Figure 5.2a for hypothetical phylogeographic boundaries and locations from previous studies. Hypothesis

1 was repeated for COI data with the structure: Group I (North) WHB (5) and Group II (South) GRS (6) and DUN (3). Hypothesis 2 could not be repeated for COI because haplotypes were not sampled from the northwest of New Zealand (Gerard et al. 2008).

Table 5.6 –AMOVA structures for population subdivision hypotheses within New Zealand (COIII data set).

Hypothesis	Group number	Group name	Regions (number of individuals)	Sample size
1	I	North	OPUM (1), TETI (3), DOVB (1), RUSH (1), WLGH (7), MAUI (1), MUPO (3)	17
	II	South	CHAI (12), RIVE (5), RINB (16), AUCI (7), CAMI (6)	46
2	I	Northwest	OPUM (1), TETI (3), DOVB (1), RUSH (1)	6
	II	Southeast	WLGH (7), MAUI (1), MUPO (3), CHAI (12), RIVE (5), RINB (16), AUCI (7), CAMI (6)	57

Phylogeographic boundaries in hypotheses 1 and 2 are visualized in Figure 5.2a.

5.3.5.2 South American distribution

The main phylogeographic disjunction observed for marine invertebrate species along the Chilean coastline follows three regions (Figure 5.2b). One Chilean sample site from this study is located in northern portion of the transitional temperate region and another in the cold temperate region (Lancelloti & Vasquez 1999). Additionally, although the presence of biogeographic barriers for nearshore invertebrate marine taxa on the Atlantic coast of South America has not been investigated in published studies, the Argentine and Falkland Islands locations are located south of the hypothesized break at 48ºS and will be included in the cold temperate region in a second analysis (Table 5.7).

This hypothesis predicts that more genetic variation will be contained between biogeographic regions than within biogeographic regions. An AMOVA was performed in Arlequin to test the relative contribution of the *a priori* grouping in hypothesis 2 to the observed genetic variance. Hypothesis 1 was evaluated with a pairwise F_{ST} calculation in Arlequin. Hypothesis 1 was repeated for COI data with the structure: Group I (North) MAU (7) and Group II (South) PAT (9) (Figure 5.2b), and was evaluated with pair-wise F_{ST} calculation in Arlequin.

Table 5.7 – AMOVA structures for population subdivision hypotheses within South America (COIII data set).

Hypothesis	Group number	Group name	Regions (number of individuals)	Sample size
1	I	North	CHCO (13)	13
	II	South (Chile)	CHGT (7)	7
2	I	North	CHCO (13)	13
	II	South (Chile, Argentina and Falkland Islands)	CHGT (7), ARUS (8), FIST (2)	17

Phylogeographic boundary in hypotheses 1 and 2 is visualized in Figure 5.2b.

5.3.5.3 Australian distribution

Testing the east-west phylogeographic break observed in *Nerita* (Waters et al. 2005) in southern Australia was performed with AMOVA analyses in Arlequin (Table 5.8).

Table 5.8 – AMOVA structures for population subdivision hypothesis within Australia (COIII data set).

Hypothesis	Group number	Group name	Regions (number of individuals)	Sample size
1	I	East	PORA (17), WONR (1)	18
	II	West	MELB (13)	13

Phylogeographic boundary in hypothesis 1 is visualized in Figure 5.2c.

The phylogeographic boundary identified in *Nerita* is situated at Wilsons Promontory, Southern Australia (Waters et al. 2005), at ~ 146ºW longitude. Melbourne lies to the west of this boundary and Port Arthur and the Wonboyn River lie to the east. Refer to Figure 5.2c for the hypothetical phylogeographic boundary from Waters et al. (2005) (hypothesis 1). Hypothesis 1 could not be repeated for COI because sample all locations lie to the east of 146ºW.

5.4 Results

5.4.1 DNA sequence data

The best choice nucleotide substitution models as determined by ModelTest under the BIC were TrN+G (gamma=0.1949, COI) and TIM+G (gamma=0.1681, COIII). The COIII and COI data sets were simply aligned with outgroups, the COI M. *trossulus* sequence had 32 missing characters at the 3' end and two missing characters at nucleotide sites 606 and 607.

5.4.2 Phylogenetic analysis

Bayesian phylogenetic reconstructions are presented for COIII (a) and COI (b) data sets in Figure 5.3. Both reconstructions exhibit general congruence and high nodal support with respect to the reciprocally monophyletic property of clades associated with the geographical regions South America (includes Kerguelen Islands in COI), New Zealand and Australia. COI Kerguelen Island haplotypes resolve onto the South American clade in all phylogenetic reconstructions, herein reference to the COI South American clade includes Kerguelen Island. Both data sets show basal polytomies for South American haplotypes. Intraregional divergence in New Zealand and Australia is more prominent in the COIII than the COI phylogeny; for example, the COIII New Zealand clade infers three coalescent events between subclades A and B (Figure 5.3) and the New Zealand MRCA, whereas the COI New Zealand clade only contains one coalescent event. The same pattern is evident in the Australian regional clade with subclades C and D (Figure 5.3). COIII haplotype NZ16 resolves onto the South American clade but is exclusively found at the Chatham Islands location ~ 850 km east of mainland New Zealand. This affects the reciprocal monophyly status of the South American clade.
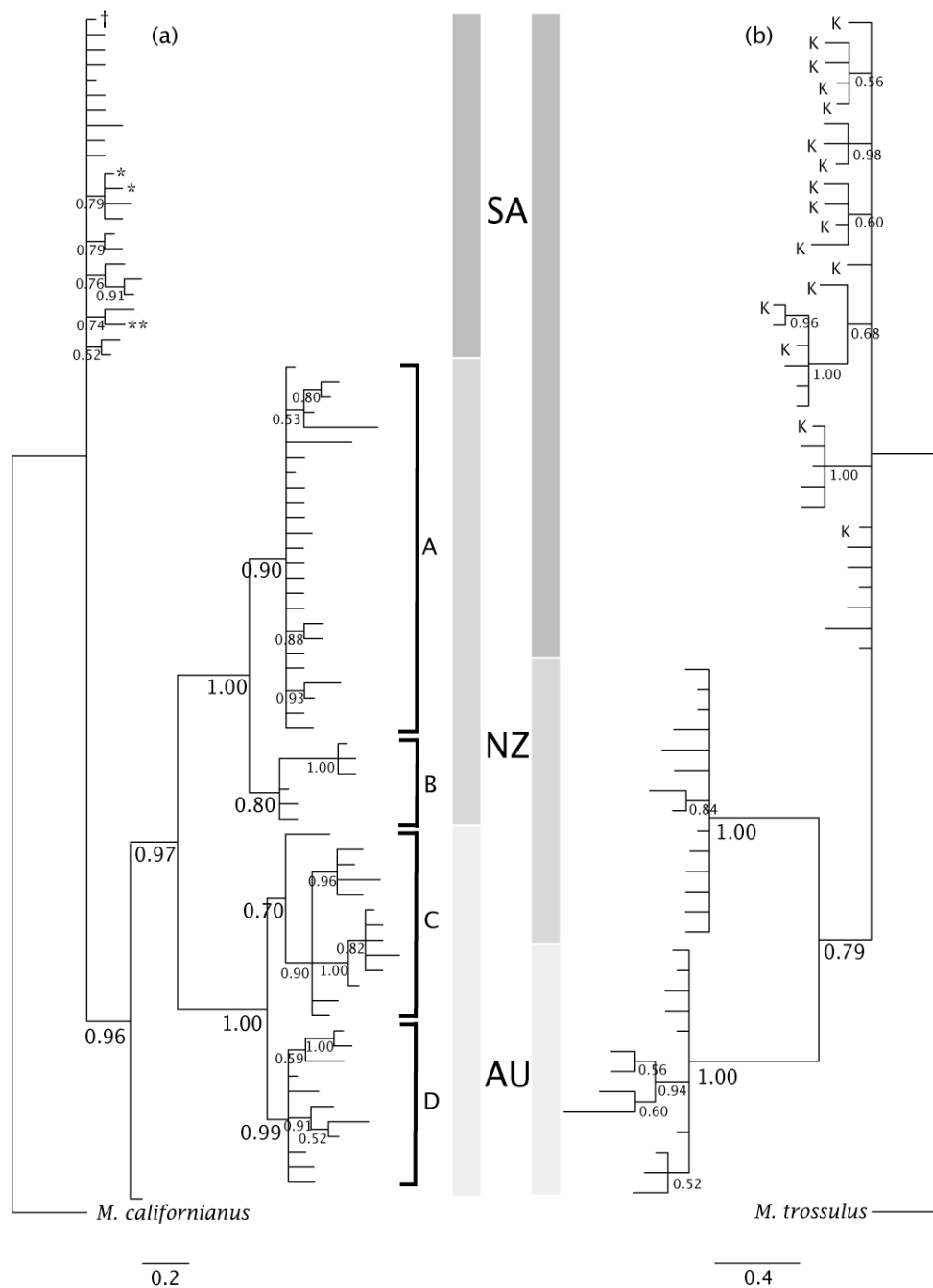
Figure 5.3 – COIII (a) and COI (b) phylogenies of Southern hemisphere *Mytilus galloprovincialis*.
Grey bars represent the geographic distribution of reciprocally monophyletic clades. Posterior probabilities are given as numbers at each node. K denotes Kerguelen Island haplotypes. Subclades A and B represent New Zealand divergence and subclades C and D represent Australian divergence. Note that distance to the outgroup is not representative of true genetic distance. * denotes haplotypes common to more than one South American country. ** denotes haplotypes common to South America and New Zealand. denotes the only New Zealand haplotype found in South America.

5.4.3 Dated phylogenies

The dated phylogenetic reconstructions (BEAST) of the COIII (a) and COI (b) data sets are presented in Figure 5.4. The two data sets are congruent with respect to regional topology, enforcing the reciprocally monophyletic property of regional clades also observed in MrBayes reconstructions (Figure 5.3). Intraregional topology cannot be directly compared because of asymmetric sampling schemes. The evolutionary relationships among regional clades inferred by BEAST are not congruent with relationships inferred by MrBayes. Placement of the South American polytomy remains unresolved in MrBayes reconstructions whilst the same regional clade is placed in a derived position sister to Australia (NZ(AU, SA/K)) in BEAST reconstructions.

Nodes A through D (Figure 5.4) represent the same geographical associations of coalescent events across data sets. The two oldest nodes, B and D, represent MRCAs for Australian/South American populations and the entire Southern hemisphere, respectively. The dates of these nodes are roughly concordant across data sets, node D represents the ingroup calibration point set at 2.35 (SD 0.31) that and node B represents the last coalescent event before Southern hemisphere colonization; estimated at 2.39 mya and 2.04 mya in COIII and COI data sets, respectively. The confidence intervals of nodal dates overlap for the deep coalescent events (nodes B and D, Figure 5.4).
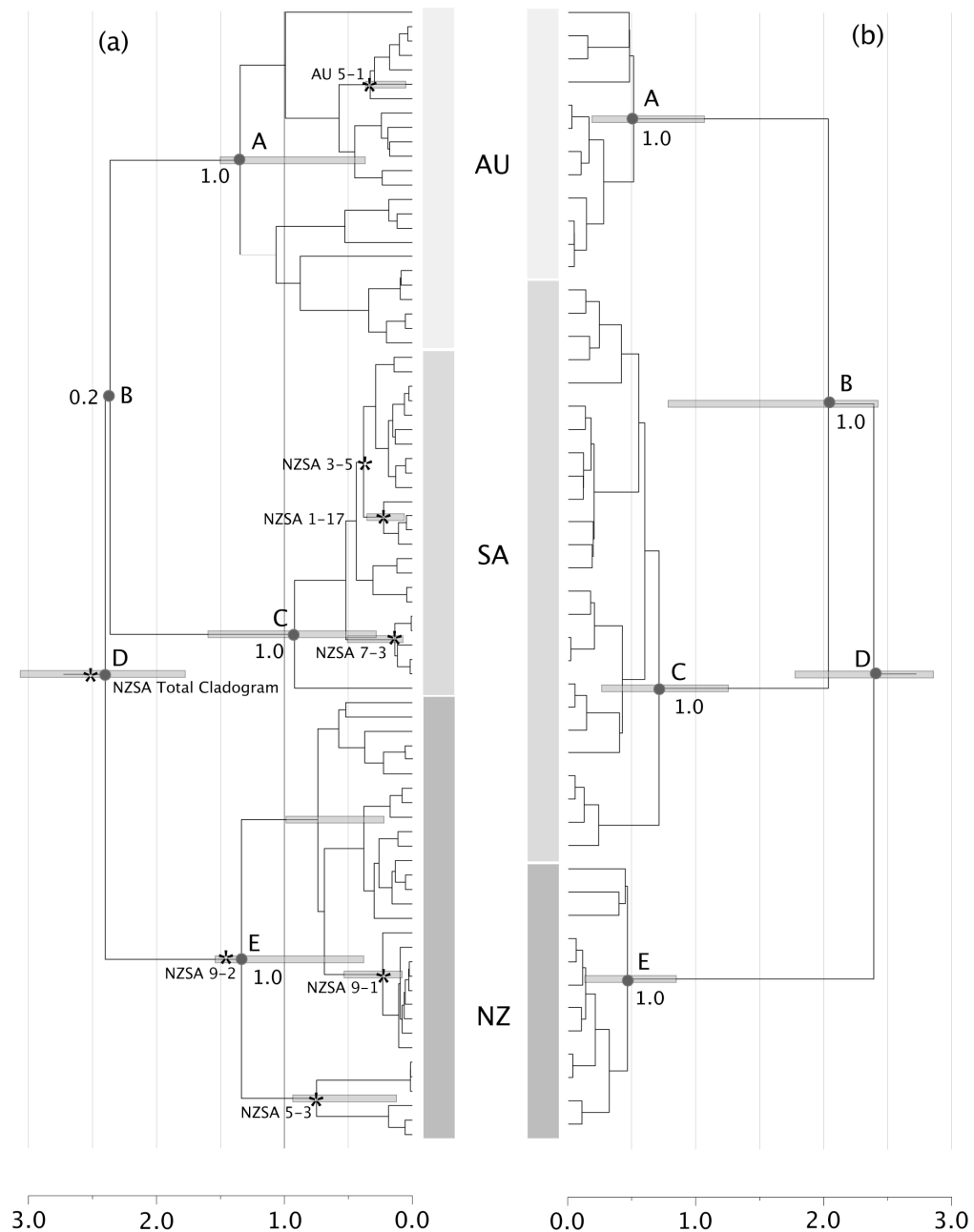
Figure 5.4 – dated phylogenies constructed from (a) COIII and (b) COI data sets in BEAST.
Vertical grey bars represent the geographic distribution of reciprocally monophyletic clades. Horizontal grey bars represent the confidence interval for specific node dates (not all nodes are presented). The scale bar is millions of years. Nodes indicated with (A to E) are congruent between data sets and referenced in the text. Nodes indicated with correspond to statistically significant NCPA clades. The haplotype network and clade number is specified: NZSA denotes the New Zealand/South America haplotype network and AU denotes the Australia haplotype network. Outgroup root not shown. Posterior probability values are provided for labelled nodes.

187

Nodes A, C and E represent the T_{MRCA} of Australian, South American and New Zealand regional clades, respectively (Figure 5.4), but the estimated dates for nodes differ between data sets. The difference of estimated dates between data sets is more pronounced for coalescent events A, C and E than for the deeper coalescent events B and D. All confidence intervals of nodes A, C and E overlap with each other but not with the confidence intervals for nodes B and D (except COI node B). Even though confidence intervals overlap, nodes A, C and E of the COI topology infer earlier coalescent events than corresponding nodes of the COIII topology.

5.4.4 Nested clade phylogeographic analysis

5.4.4.1 COI

The COI data set represents three unconnected haplotype networks at the 95% parsimony criterion associated with geographical distributions of haplotypes and regional clades of phylogenetic trees (data not shown). There were no statistically significant nested clades contained within New Zealand or Australian networks. One nested clade is statistically significant with respect to nonrandom geographical association of genetic differentiation in the South American network. This clade (2-1, Table 5.9) includes haplotypes from Maullin, Canal de las Montañas (both Chile) and the Kerguelen Islands and restricted gene flow with isolation by distance is inferred.

Table 5.9 – COI and COIII haplotype network hierachical clades with significant geographic and genetic associations.

Data Set	Network	Clade	Nested Clades	D_c	D_n	Steps in Key	Inference
COI	SAK	2-1	1-1	ns	≥	1, 2, 11, 17, 4.	Restricted gene flow with isolation by distance
			1-2	ns	ns		
			1-3	ns	ns		
			I–T	ns	≥		
COIII	AU	5–1	4–1	≤	≤	1, 2, 11, 12, 13, 14.	Sample design cannot discriminate among contiguous range expansion, long distance colonization or past fragmentation
			4–2	ns	≥		
			I–T	ns	ns		
COIII	NZSA	1–17	ARUS/CHCO	ns	≥	1, 2, 11, 17, 4.	Restricted gene flow with isolation by distance
			CHGT	ns	ns		
			ARUS/CHGT	ns	ns		
			I–T	ns	≥		
COIII	NZSA	3–5	2–7	ns	ns	1, 2, 3, 5, 6.	Insufficient resolution to distinguish between range expansion/colonization and restricted dispersal/gene flow
			2–8	≤	≤		
			2–9	ns	≥		
			I–T	ns	ns		
COIII	NZSA	4–2	2–2	≥	≥	1, 2, 3, 4.	Restricted gene flow with isolation by distance
			3–2	ns	ns		
			I–T	≥	ns		
COIII	NZSA	5–3	4–3	ns	≥	1, 19.	Allopatric fragmentation
			2–4	≤	≤		
			I–T	≤	≤		
COIII	NZSA	7–3	3–5	≥	≥	1, 2, 3, 4.	Restricted gene flow with isolation by distance
			6–4	≤	≤		
			I–T	≥	≥		
COIII	NZSA	9–1	AUCI	ns	ns	1, 2.	Inconclusive outcome – tip interior status cannot be determined
			OPUM	ns	≥		
			RIVE	ns	ns		
			1–2	ns	ns		
			6–1	ns	ns		
COIII	NZSA	9–2	1–5	ns	ns	1, 2.	Inconclusive outcome – tip interior status cannot be determined
			4–2	ns	ns		
			7–1	ns	ns		
			8–1	ns	≥		
COIII	NZSA	Total Cladogram	12–1	≤	≤	1, 19.	Allopatric fragmentation
			12–2	≤	≥		

SAK = South American / Kerguelen Island, AU = Australian and NZSA = South American / New Zealand. D_c = within clade distance D_n = nested clade distance. ≤ and ≥ = significantly small and large (distance), respectively. I–T = interior/tip clade association. 'Steps in Key' refers to Geodis inference key

(http://darwin.uvigo.es/download/geodisKey_28Apr09.pdf). All referenced clades are visualized in Figure 5.4.

5.4.4.2 COIII

The COIII data set represents two unconnected haplotype networks at the 95% parsimony criterion: (1) includes South American and New Zealand haplotypes (data shown) and (2) includes Australian haplotypes (data not shown). Australian nested clade 5-1 (phylogenetic tree Figure 5.4a, Table 5.9) is statistically significant with respect to nonrandom geographical association between Port Arthur and Melbourne sample locations (Figure 5.2c). The sampling design proved to be inadequate to discriminate among competing hypotheses of contiguous range expansion, long distance colonization and past fragmentation. The South American/New Zealand haplotype network is presented in Figures 5.5 and 5.6 with statistically significant nested clades (Table 5.9). Six nested clades had significant geographic/genetic associations and the total cladogram was bisected into South American and New Zealand clades with allopatric fragmentation inferred from the Geodis reference key. Another allopatric fragmentation inference was made for clade 5-3 which includes a division between Bay of Islands and a subset of Chatham Islands and Ringitingi Beach haplotypes (Figures 5.4a and 5.5, Table 5.9).
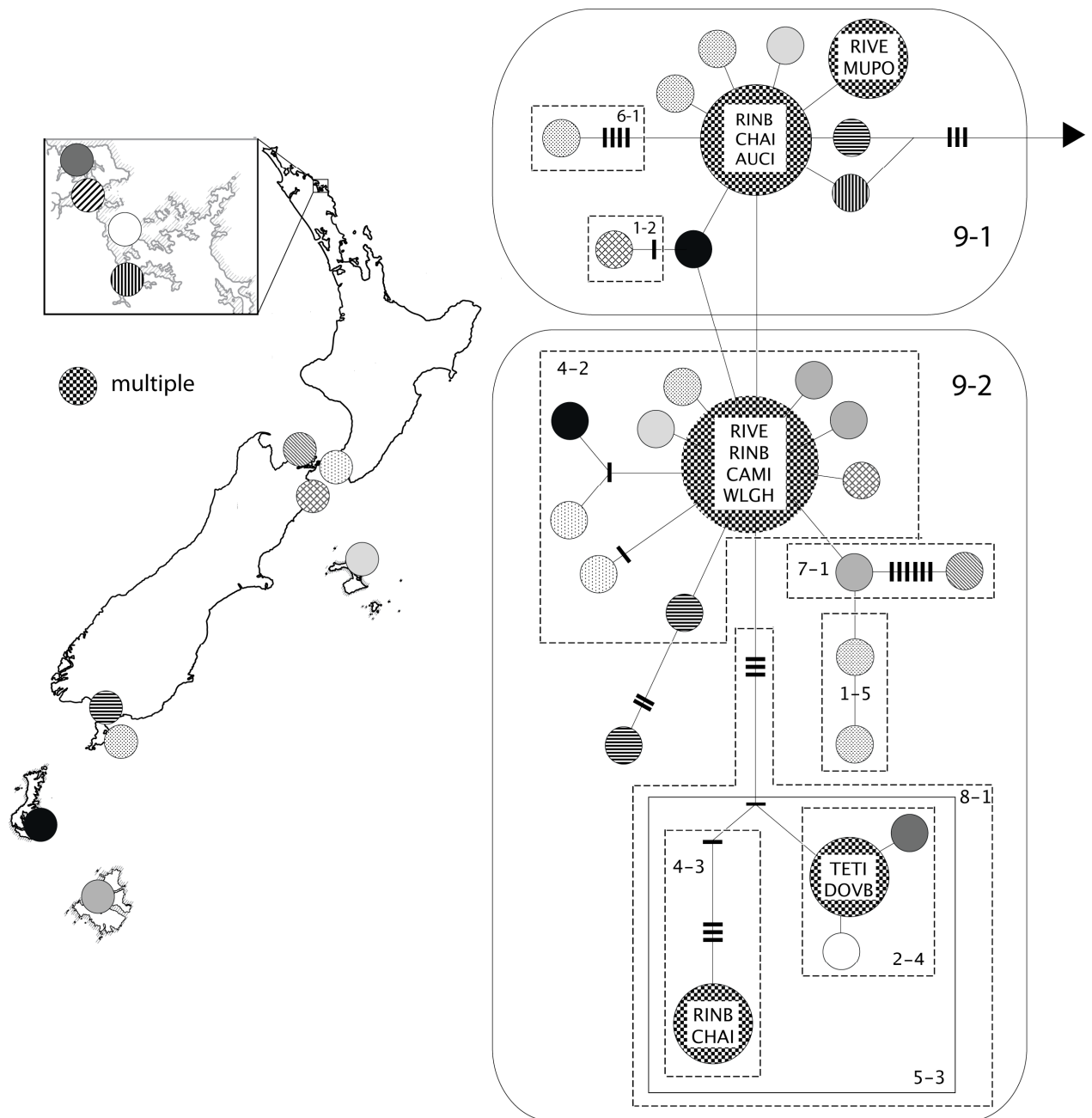
Figure 5.5 – COIII NCPA New Zealand.
Circles represent haplotypes and are scaled to the number of individuals with each haplotype. Each line represents one nucleotide substitution. The arrow indicates a 95% parsimony connection to South America (Figure 5.6). Statistically significant clades are outlined in solid line and associated nested clades are outlined in dashed line. Bars represent missing haplotypes. Haplotypes shared among multiple locations have location codes (Table 5.1) listed.
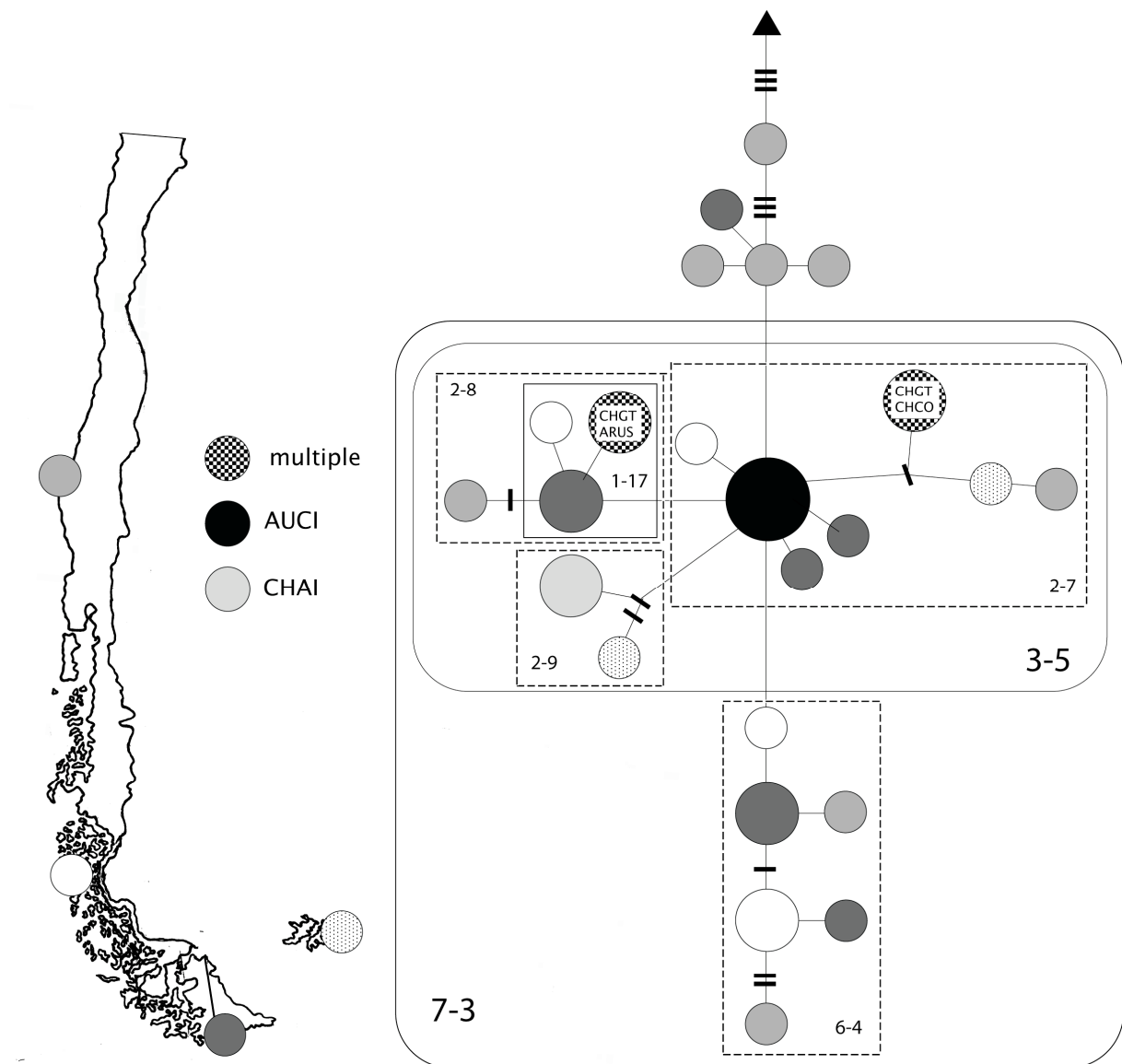
Figure 5.6 – COIII NCPA South America.
Circles represent haplotypes and are scaled to the number of individuals with each haplotype. Each line represents one nucleotide substitution. The arrow indicates a 95% parsimony connection to New Zealand (Figure 5.5). Statistically significant clades are outlined in solid line and associated nested clades are outlined in dashed line. Bars represent missing haplotypes. Haplotypes shared among multiple locations have location codes (Table 5.1) listed.

Restricted gene flow with isolation by distance was inferred for three clades

containing: (1) mixed Ushuaia (Argentina) and Golfo Trinidad haplotypes (Chile) (1-17),

(2) all New Zealand haplotypes except Bay of Islands and Maud Island (4-2), and (3)

mixed South American and Auckland Islands haplotypes (7-3) (Table 5.9). Inconclusive outcomes for NZSA clades 9-1 and 9-2 are derived from ambiguous mutational connections (loops) and most likely represent a false positive. The presence of an Auckland Island haplotype in the middle of the South American network infers some recent or ongoing gene flow between the two regions.

Nested in the Australian network clade 5-1, clade 4-1 contains haplotypes exclusively from Port Arthur and 4-2 contains almost exclusively Melbourne haplotypes (and a single Port Arthur haplotype) (data not shown).

5.4.5 Intraregional biogeographic hypothesis testing

AMOVA results for intraregional hypotheses are listed in Table 5.10. For hypotheses evaluated with pair-wise location F-estimates (South America hypothesis 1): COI pair-wise F_{ST} value is -0.008 (p=0.40) for Maullin and Canal de las Montañas and COIII pair-wise F_{ST} value is 0.012 (p=0.33) for Concepcion and Golfo Trinidad.

Table 5.10 – AMOVA results for intraregional biogeographic hypothesis testing.

Gene/ Region	Hypothesis	Source of variation	Percent variation	F-estimate	F-estimate p-value
COIII/ NZ	1	Among groups	4.84	$F_{CT} = 0.48$	0.06
		Among sample locations within groups	18.78	$F_{SC} = 0.20$	0.02
		Within sample locations	76.39	$F_{ST} = 0.24$	<0.001
COIII/ NZ	2	Among groups	29.19	$F_{CT} = 0.29$	0.02
		Among sample locations within groups	9.65	$F_{SC} = 0.14$	0.02
		Within sample locations	61.6	$F_{ST} = 0.39$	0.002
COI/ NZ	1	Among groups	1.10	$F_{CT} = 0.01$	0.33
		Among sample locations within groups	-0.75	$F_{SC} = -0.01$	0.32
		Within sample locations	99.65	$F_{ST} = 0.003$	0.49
COIII/ SA	2	Among groups	2.81	$F_{CT} = 0.03$	0.50
		Among sample locations within groups	-2.84	$F_{SC} = -0.02$	0.50
		Within sample locations	99.53	$F_{ST} = 0.004$	0.46
COIII/ AU	1	Among groups	-16.97	$F_{CT} = -0.17$	1.00
		Among sample locations within groups	53.58	$F_{SC} = 0.46$	0.05
		Within sample locations	63.39	$F_{ST} = 0.37$	<0.001

NZ = New Zealand, SA = South America, AU = Australia.

Out of all proposed intraregional hypotheses of biogeographic breaks across the Southern hemisphere locations and within two data sets, only the proposed northeast-southwest population disjunction in New Zealand is detected with the COIII gene. In this case 29.19% of the genetic variation within the whole data set is contained between northeast and southwest groups whilst only 9.65% variation is contained among locations between groups (Table 5.10). This is accompanied by significant F-estimates. The null hypothesis of no geographic association in accordance with biogeographical boundaries observed in other near-shore marine invertebrate species cannot be rejected in all other cases.

5.5 Discussion

5.5.1 Southern hemisphere phylogeography: vicariance vs. dispersal

The tectonic-plate mediated vicariance model predicts evolutionary relationships within a species should mirror the sequence of Gondwanaland continental break-up: (Africa(New Zealand(Southern South America, Australia))) (Sanmartin & Ronquist 2004). It is known that Northern and Southern hemisphere *Mytilus* divergence is relatively recent and the Southern hemisphere distribution is extremely unlikely to have been influenced by Gondwanan-related tectonic vicariant processes. Until data is presented that could date the presence of extant members of the genus *Mytilus* in the Southern hemisphere earlier than what has been proposed here, tectonic plate mediated vicariance processes are not considered further for hypotheses on the phylogeographic distribution of *Mytilus*.

The pattern of blue mussel genetic differentiation across South America, New Zealand and Australia congruent between single gene COIII and COI phylogenies lends support to the inferred evolutionary relationships. Basal polytomies in the phylogenies reconstructed in MrBayes formed by South American (and Kerguelen Island in COI) haplotypes most likely represent irresolvable nodes and not the true evolutionary relationships among geographically structured clades. The irresolvable nature of these polytomies is evident when we compare topologies reconstructed in MrBayes to topologies reconstructed in BEAST. The topologies between phylogenetic programs are incongruent with respect to the inferred evolutionary relationships of reciprocally monophyletic regional clades even though both programs use the Bayesian statistical platform. BEAST phylogenies place New Zealand at the basal position in both data sets. This is not a direct conflict of alternate evolutionary relationships, rather, the incongruence between programs indicates that evolutionary relationships between South American, Australian and New Zealand reciprocally monophyletic clades are irresolvable with these genes. In both

programs, the Australian regional clade is resolved in a derived position on the tree, therefore the COI and COIII genes can confidently infer the general relationships: (1) (south South America, (New Zealand, Australia)) or (2) (New Zealand, (south South America, Australia)).   These general relationships have been observed across many Southern hemisphere taxa and are termed the plant southern pattern (PSP) and southern Gondwana pattern (SGP), respectively (Sanmartin & Ronquist 2004).

The PSP and SGP patterns are usually underlain by competing hypotheses of dispersal and vicariance, respectively (Sanmartin & Ronquist 2004).   Since the SGP pattern has been linked to tectonic-mediated vicariance associated with the break up of Gondwana (Cranston & Edward 1992), the most likely phylogeographical scenario for blue mussel Southern hemisphere distribution is linked to dispersal processes underlying the PSP pattern.   For marine taxa, dispersal hypotheses include observations of directional eastward dispersal underlain by the West Wind Drift (associated with the ACC) and Trans-Tasman dispersal between Australia and New Zealand.

5.5.2 Eastward dispersal and the West Wind Drift

COI haplotypes from South America and the Kerguelen Islands resolve onto a paraphyletic clade and no geographic structure other than isolation by distance is detected between these locations.   Ongoing gene flow between South America and the Kerguelen Islands is inferred from the close evolutionary relationships among haplotypes unique to and shared between these regions.   The shortest dispersal route between these regions infers a present day eastward dispersal direction for *Mytilus* from South America to the Kerguelen Islands (Figure 5.7) via the West Wind Drift.
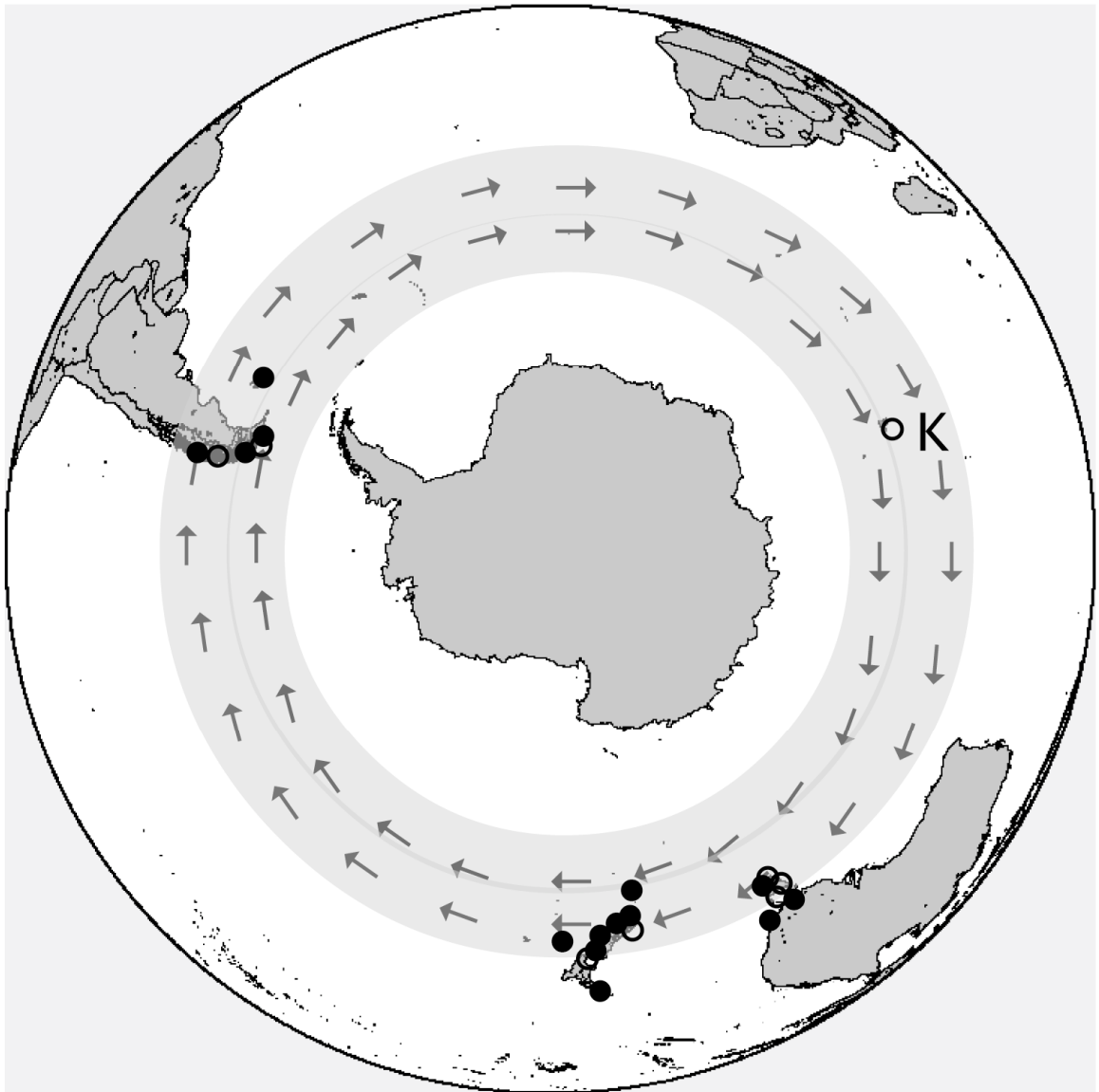
Figure 5.7 – The West Wind Drift (grey circle with directional arrows).
Dark circles indicate COIII sample locations and hollow circles indicate COI sample locations (Gerard et al. 2008). K = Kerguelen Islands.

Phylogenetic topologies do not directly indicate dispersal processes outside of the South American/Kerguelen Island connection, but additive information from haplotype networks and nested clade phylogeographical results provide qualitative observations on relative amounts of gene flow that underlie a general dispersal pathway between New Zealand and South America. COIII haplotypes from New Zealand and South America (but not Australia) were connected under a 95% parsimony criterion and a single haplotype from the Chatham Islands resolved onto the South American phylogenetic clade. Both of

these results indicate higher (historic) gene flow between New Zealand and South America than between either of these countries and Australia. Gene flow was most likely greater between South America and New Zealand in the past and has subsequently either ceased or is significantly reduced at present, supported by an allopatric fragmentation NCPA inference and the reciprocally monophyletic property of both regional clades.

Trans-Tasman gene flow has been observed in highly dispersing marine invertebrates (Waters et al. 2005) but the Tasman Sea is the only body of water included in this investigation that *Mytilus* do not disperse over. Reciprocal monophyly of regional clades and haplotype networks that do not include Australian haplotypes at the 95% parsimony criterion (both COI and COIII) indicate there is no gene flow between these regions either at present or in the recent past. The fairly high mutation rate of the cytochrome oxidase gene complex in *Mytilus* spp. begets a specific temporal component to this observation and any signature of gene flow (in mtDNA) in the distant past (but less than ~2.5 mya) would have been erased by present day.

The distance between South America and the Kerguelen Islands is much greater than the distance between New Zealand and Australia (Table 5.4, Figure 5.7), which leads to the question of why gene flow is ongoing between the former regions but not the latter. Gene flow between South America and the Kerguelen Islands could possibly have ceased in the recent past and lineage sorting has not progressed far enough to generate monophyly of Kerguelen Island haplotypes, or, oceanic islands could mediate ongoing gene flow in a stepping stone model. Conversely, continental currents around Australia at latitudes of *Mytilus* distribution could prevent gene flow on the temporal scale of *Mytilus* larval duration (4-7 weeks).

Trans-Pacific gene flow is inferred by the phylogenetic placement of a Chatham Islands haplotype on the South American clade and the (mutationally) close connections of

an Auckland Islands haplotype to South American haplotypes within the COIII/NZSA haplotype network (Figure 5.6).

5.5.3 Dated phylogenetics

The overlapping confidence intervals for dates associated with nodal groupings B/D and A/C/E (Figure 5.4) indicates the true coalescent date cannot be inferred with confidence. The confidence interval for node D does not overlap with confidence intervals for the remaining younger nodes (A, C, E both data sets) inferring that genetic differentiation among regional clades occurred after blue mussels initially colonized the Southern hemisphere. The confidence interval for COIII node B does not overlap with confidence intervals for the remaining younger nodes inferring that any genetic divergence contained within Australian and South American regional clades did not predate the divergence of South American and Australian populations. This infers very minimal gene flow between these regions in the past and at present day. Both observations on the timing of differentiation support the single trans-equatorial migration hypothesis of Hilbish et al. (2000).

5.5.4 Intraregional phylogeographic patterning

5.5.4.1 New Zealand

A greater number of coalescent events between tips and the COIII New Zealand MRCA compared to the same COI clade most likely represents the asymmetrical sampling between data sets. AMOVA and NCPA (COIII) analyses strongly support a population subdivision in the northwest portion of New Zealand that has also been observed in the brooding amphipod *Paracorophium lucosi* (Stephens & Hogg 2004). This population is disjunct from the remaining New Zealand distribution including offshore islands. The

genetic differentiation identified from the COIII sampling scheme is not present in the COI phylogeny because the northwest portion of the country was not sampled. Conversely, a north-south phylogeographic disjunction in New Zealand at 42ºS is not supported by NCPA or AMOVA analyses in COI and COIII genes, a surprising result if compared to phylogeographic distribution of the green-shell mussel *Perna canaliculus*, which has a larval duration time similar to *Mytilus* and exhibits a North-South genetic split at 42ºS (Apte & Gardner 2002; Star et al. 2003). The absence of blue mussels from the North Island between the Hauraki Gulf and the south coast may contribute to the observed genetic disjunction by preventing gene flow between these areas through geographically intermediate populations.

NCPA results also indicated allopatric fragmentation within the New Zealand subclade B (COIII/NZSA clade 5-3) (Figure 5.5, Table 5.9) corresponding to differentiation between the Bay of Islands and a subset of haplotypes from the Chatham Islands and Ringitingi Beach. Genetic differentiation of subclade B from the remaining Chatham Islands and Ringitingi Beach haplotypes indicates past gene flow between the offshore islands and northern New Zealand that does not occur at present day. From the dated phylogeny, gene flow ceased between these areas ~ 750,000 years before present (T_{MRCA} of subclade B).

5.5.4.2 Australia

The same relationship of the number of coalescent events between tips and the MRCA observed for the New Zealand COI and COIII clades is also present the Australian clade. COI haplotypes are restricted to Tasmania whereas COIII haplotypes span Tasmania and the provinces of South Australia and New South Wales. Although a significant association between haplotype and geographic distribution doesn't exist in

Australia, more coalescent events in the COIII topology infer a greater amount of genetic differentiation between subclades C and D than between COI haplotypes. Since asymmetric sampling of COI haplotypes prevented the observation of population disjunction in New Zealand (for that gene), the same phenomenon might be present in Australia.

AMOVA analyses do not support a distinct east west phylogeographic split centered around Wilsons Promontory in South Australia as observed in *Nerita* (Waters et al. 2005), *Scutus* and *Austrolittorina* gastropods (Waters et al. 2006), but NCPA results indicate a significant relationship between only a subset of Port Arthur and Melbourne haplotypes. Because mussels were not sampled between these locations, NCPA could not distinguish between inferences of contiguous range expansion, long distance colonization or past fragmentation. Long distance colonization can be ruled out in light of the dispersal capacity of mussel larvae, therefore, geographically intermediate locations should be sampled to distinguish contiguous range expansion from past fragmentation. A significant geographical relationship did not extend beyond the subset of haplotypes in the COIII/AU clade 5-1 to include all haplotypes sampled from Port Arthur and Melbourne, therefore some level of gene flow at present day is also inferred.

5.5.4.3 South America

AMOVA analyses did not support a North-South phylogeographic disjunction around 48ºS within Chile or South America for either COI or COIII genes. The COI/SAK nested clade 2-1 containing haplotypes from Maullin, Canal de las Montañas and the Kerguelen Islands exhibited restricted gene flow with isolation by distance. This relationship was not detected beyond the haplotypes in clade 2-1 to include the remaining haplotypes sampled from the same locations, so the degree of gene flow restriction is

examined with other evidence.   Five haplotypes are shared between Chile and the Kerguelen Islands, indicating a substantial amount of gene flow either ongoing or in the recent past.

5.6 References

Apte S, Holland BS, Godwin LS, Gardner JPA (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Ayers KL, Waters JM (2005) Marine biogeographic disjunction in central New Zealand. *Marine Biology* 147: 1045-1052.

Ayre DJ, Minchinton TE, Perrin C (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology* 18: 1887-1903.

Bayne BL (1965) Growth and delay of metamorphosis of the larvae of *Mytilus edulis* (L). *Ophelia* 2: 1-47.

Brattström H, Johanssen A (1983) Ecological and regional zoogeography of the marine benthic fauna of Chile. *Sarcia* 68:289-339.

Clement M, Posada D, Crandall KA. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657-1659.

Cranston PS, Edward DHD (1992) A systematic reappraisal of the Australian Aphroteniinae (Diptera: Chironomidae) with dating from vicariance biogeography. *Systematic Entomology* 17: 41-54.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

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

Fernandez M, Jaramillo E, Marquet PA, Moreno CA, et al. (2000) Diversity, dynamics and biogeography of Chilean benthic nearshore ecosystems: an overview and guidelines for conservation. *Revista Chilena de Historia Natural* 73: 797-830.

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

Grant WS, Cherry MI (1985) *Mytilus galloprovincialis* Lmk. in southern Africa. *Journal of Experimental Marine Biology and Ecology* 90: 179-191.

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

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

Jansson R, Dynesius M (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics* 33: 741–777.

Ladoukakis ED, Saavedra C, Magoulas A, Zouros E (2002) Mitochondrial DNA variationin a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Molecular Ecology* 11: 755-769.

Lancellotti D, Vasquez JA (1999) Biogeographical patterns of benthic bacroinvertebrates in the southeastern Pacific littoral. *Journal of Biogeography* 26: 1001-1006.

Linder HP, Crisp MD (1995) *Nothofagus* and Pacific biogeography. *Cladistics* 11: 5-32.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

De Queiroz A (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology and Evolution* 20: 68-73.

Quesada H, Beynon CM, Skibinski DOF (1995) A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* (Lmk.): Pleistocene vicariance biogeography and secondary intergradation. *Molecular Biology and Evolution* 12: 521-524.

Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* 58: 2438-2451.

Ross PM, Hogg ID, Pilditch CA, Lundquist CJ (2009) Phylogeography of New Zealand's coastal benthos. *New Zealand Journal of Marine and Freshwater Research* 43: 1009-1027.

Sanmartin I, Ronquist F (2004) Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology* 53: 216-243.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: a software for population genetics data analysis. Version 2.000. University of Geneva, Geneva, Switzerland.

Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6: 461-464.

Smietanka B, Zbawicka M, Wolowicz M, Wenne R (2004) Mitochondrial DNA lineages in the European populations of mussels (*Mytilus* spp.). *Marine Biology* 146: 79-92.

Star B, Apte S, Gardner JPA (2003) Genetic structuring among populations of the greenshell mussel *Perna canaliculus* revealed by analysis of randomly amplified polymorphic DNA. *Marine Ecology Progress Series* 249: 171–182.

Stevens MI, Hogg ID (2004) Population genetic structure of New Zealand's endemic corophiid amphipods: evidence for allopatric speciation. *Biological Journal of the Linnean Society* 81: 119–133.

Templeton AR (2004) Using haplotype trees for phylogeographic and species inference in fish populations. *Environmental Biology of Fishes* 69: 7-20.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134: 659-669.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117: 343-351.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics*, 132, 619-633.

Templeton AR, Routman E, Phillips C (1995) Separating population structure from population history: A cladistic anal- ysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, Ambystoma tigrinum. *Genetics* 140: 767–782.

Toro JE, Castro GC, Ojeda JA, Vergara AM (2006) Allozymic variation and differentiation in the Chilean blue mussel, *Mytilus* chilensis, along it's natural distribution. *Genetics and Molecular Biology* 29: 174-179.

Waters JM (2008) Marine biogeographical disjunction in temperate Australia: historical landbridge, contemporary currents, or both? *Diversity and Distributions* 14: 692–700.

Waters JM, Roy MS (2004a) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology* 13: 2797–2806.

Waters JM, Roy MS (2004b) Out of Africa: the slow train to Australasia. *Systematic Biology* 53: 18–24.

Waters JM, King TM, O'Loughlin PM, Spencer HG (2005) Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology* 14: 2789–2802.

Waters JM, McCulloch G, Eason J (2006) Marine biogeographical structure in two highly dispersive gastropods: implications for trans-Tasman dispersal. *Journal of Biogeography* 34: 678-687.

Chapter 6: Population and seascape genetics of *Mytilus galloprovincialis* Lmk. (Mytilidae; Bivalvia) in New Zealand.

6.1 Abstract

Consistently realising the potential for pelagic marine larvae to disperse over large spatial scales has been recently redefined by data suggesting dispersal distances are much smaller and open to stochastic variables. Members of the *Mytilus edulis* species complex produce planktotrophic larvae that live the water column for 3-7 weeks, allowing for potentially large dispersal distances. In New Zealand, the sibling species *M. galloprovincialis* is common on rocky shorelines of the South Island and offshore islands, and is also present in two pockets in the North Island. Many nearshore marine invertebrates in New Zealand share a population differentiation pattern of a North/South division along a biogeographic break at roughly 42ºS latitude on the eastern side of the South Island. This study investigates the geographic pattern of population subdivision, degree of connectivity among subdivisions, and the influence of environmental variation on allelic frequency profiles using microsatellite loci developed for *M. galloprovincialis* and it's sibling species *M. trossulus* from the Northern hemisphere. Bayesian clustering and distance methods suggest a population structure which does not adhere to that of other New Zealand marine invertebrates; two greatly differentiated populations (F_{ST} = 0.16) indicate biogeographic barriers on the Western side of the South Island and down the middle of a large North Island sheltered bay. The Northwestern population consists of the northernmost sample location and the northwest tip of the South Island and a second, almost panmictic population consists of everything else from the second northernmost sample location to the southernmost offshore island. Isolation by distance is observed within the Southern population over moderate spatial scales. Intermediate gene flow among sample locations clustering around the southern boundary of the Northwest population suggests this biogeographic barrier is semi-permeable to gene flow. Two sample locations in the North Island ~15km apart are assigned to separate populations (pairwise F_{ST} = 0.22); artificial population structure created by non-native alleles or temporal variation of gamete release are considered as possible contributing sources. Annual mean sea surface temperature and winter solar radiation are significantly linked to allelic frequency profiles in the Southern population. High winter solar radiation relative to the annual mean at the southernmost sample location could drive an intermittent gamete release response, contributing to larval

dispersal and isolation by distance. Further characterization of the genome by identifying introgressed genes in hybrid individuals and allelic frequency profiles of reference species from the Northern hemisphere would distinguish naturally occurring from artificially created (non-native alleles) population subdivision.

6.2 Introduction

6.2.1 Marine population connectivity

Marine organisms have long been considered to consistently disperse across large, ocean-wide scales allowing for spatially greater genetic exchange and continuous, panmictic populations (Palumbi 2003). Scenarios depicting either panmixia or 'stepping stone' isolation by distance have classically modeled marine population differentiation (Selkoe et al. 2008). Many benthic marine invertebrates release long-lived larvae capable of large dispersal distances and it is assumed these capabilities are fulfilled by each generation. Larval duration greater than 10 days has been hypothesized as a predictor variable for dispersal potential (Ross et al. 2009) but recent studies describing genetic differentiation at small spatial scales belie the panmixia assumption (Bay et al. 2006).

Connectivity of populations in the oceanic environment is a product of many variables although larval behaviour, oceanographic circulation, climatic events and food availability are the main driving factors; stochasticity of these variables affects larval transport on small temporal scales (i.e. between single generations) leading to unpredictable dispersal and recruitment (Siegel et al. 2008). Nearshore marine invertebrates generally release a large number of larvae with high mortality rates before and just after settlement with up to 98.6% mortality before reproductive maturity (Thorsen 1966). 'Realized connectivity' relates the success rate of larvae in reaching reproductive maturity in contrast to the actual number of gametes released. This similar to effective

population size (N_m) (Hamilton et al. 2008) and depends on: (1) dispersal rates and scales and (2) pre and post-settlement mortality.

New spatiotemporal dispersal estimates reveal distances traveled by larvae may be orders of magnitude less than originally thought and self-recruitment is an underestimated contributor to larval settlement rates and patterns (Cowen et al. 2006). Hydrodynamics plays a role in larval dispersal and recruitment. At the time of gamete release, larvae may stay in the water column directly above the parental habitat or be transported to another layer that is advected to the open ocean, only to return in pulses of water envelopes likely to contain many larvae through upwelling or tidal bores (Pineda 1994, Metaxas 2001, Siegel et al. 2003).

Most benthic invertebrates have a planktonic larval stage lasting from hours to months and the mortality of larvae can be very high during the time before settlement. Coastal current circulation patterns differ greatly throughout the world but can carry larvae away from the coast and suitable settlement habitat. In addition, predation in the plankton and physical stresses of ultraviolet radiation, increased temperature and decreased oxygen levels affect the fitness and survivorship of pre-settlement larvae.

Recruitment rates are dependent on early post-settlement mortality rates and have been documented for sessile invertebrates from 13% to 28% (Bergeron & Bourget 1986, Young 1991, Gosselin & Qian 1996). Early post-settlement mortality generally follows a Type III survivorship curve (survival rate rapidly decreases then levels off) with the highest mortality within the first few days to weeks depending on the species (Gosselin & Qian 1997). Main causes of high mortality (Hunt & Scheibling 1997) are (1) delayed metamorphosis, reducing the fitness of larvae or juveniles (Pechenik 1990), (2) biological disturbance (predation) (Gosselin & Chia 1995), (3) physical disturbance including temperature and hydrodynamics, and (4) physiological stress. Chemico-physical cues also

play a role in settlement rates and scales (Pawlik 1992, Rodriguez et al. 1993). Although rare events of extreme long range dispersal where larvae are exported great geographical distances shape global colonization patterns, it is becoming more apparent that the difference between potential and realized dispersal can shape population level differentiation within smaller spatiotemporal scales (Swearer et al. 2002, Cowen et al. 2006, Becker et al. 2007).

The primary aim of seascape genetics is to correlate environmental and oceanographic variation with genetic dissimilarities among populations and individuals without the necessity of defining populations *a priori*. Correlating these variables may identify potential causes of selection or antagonizing forces affecting observed spatial genetic discontinuities, clines and random patterns. Uniform distribution of a species across a marine landscape is the most suitable type of data for seascape analysis where cryptic boundaries in gene flow may be identified with no obvious cause (i.e. major oceanographic feature or current system) (Manel et al. 2003).

6.2.2 The *Mytilus edulis* species complex

Three sister species comprise the *Mytilus edulis* species complex (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) and all three taxa naturally hybridise in areas where their distribution is sympatric. Population genetic surveys of sister species in the Northern hemisphere centered around allozyme variation from the 1970's to the 1990's (Beaumont et al. 2007) and has since moved towards mitochondrial DNA variation within male and female mitolineages. Despite the development of several microsatellite suites (Presa et al. 2002, Gardestrom & Pereya 2007, Yu & Li 2007) the application of these highly polymorphic and neutral markers in population genetic studies has not occurred in the literature.

Broadcast spawning *Mytilus* species release gametes into the water column where they are randomly fertilized and develop into planktotrophic larvae that live for approximately three to seven weeks (Bayne 1976). Metamorphosis may be delayed for up to six months (Lane et al. 1985) with the adaptation of a long monofilament thread to increase drag (Sigurdsson et al. 1976). This early life history larval phase represents the only opportunity for blue mussel dispersal but coastal oceanographic features may enable larval dispersal on the scale of hundreds of kilometers.

Southern European *M. galloprovincialis* populations exhibit a biogeographic split between the Mediterranean Sea and Atlantic Ocean that occurs at the Almeria-Oran oceanographic front (Quesada et al. 1995, Ladoukakis et al. 2002) and is best described by the male mitolineage (Smietanka et al. 2009). Within the female mitolineage, *M. edulis* and *M. galloprovincialis* are significantly differentiated in the Atlantic Ocean (Smietanka et al. 2009) and there is a slightly separated population of *M. galloprovincialis* in Iceland (Riginos et al. 2004, Smietanka et al. 2004). The Azov Sea is the last population differentiated by female mitolineage dissimilarities in Europe (Smietanka et al. 2009). Many hybrid zones are also restricted to certain areas due to population demographics and pure populations show clinal variation in allelic frequencies mostly attributed to temperature and salinity gradients.

Oceanographic physicochemical variability plays a selective role in population demography for many marine invertebrates and especially *Mytilus* species. Salinity gradients (oceanic versus estuarine) affect allele frequencies at the (allozyme) *Lap* locus in *M. edulis* (Koehn et al. 1976, 1980) and *M. trossulus* (McDonald & Siebenaller 1989). This correlation in *M. galloprovincialis* depends on the specific lineage, Gardner (1994) described *M. galloprovincialis* North Atlantic lineage intolerance to low and fluctuating salinity in estuaries but Skalamera et al. (1999) found no selection between low or high

salinity in Mediterranean lineage *M. galloprovincialis*. The latter study could not correlate genetic dissimilarity and ecological variation or geographical distance (medium scale under 150 km.) (Skalamera et al. 1999). More generally, size structure and growth rates of *M. galloprovincialis* are correlated with salinity in the Baltic Sea (Westerborn et al. 2002). Other selective and adaptive studies of allozyme variation show temperature, wave exposure and intertidal height (from low low tide) affect the structure of populations (including hybrid zones).

6.2.3 New Zealand phylogeography of marine invertebrates

Phylogeographic studies of New Zealand coastal benthic invertebrates show general patterns of connectivity and dispersal common across taxa. Species such as *Jasus* spp. (rock lobster), with larval periods exceeding one year, are genetically homogeneous across their nationwide distribution (Brasher et al. 1992, Ovendon et al. 1992). A review of New Zealand taxa phylogeographic studies by Ross et al. (2009) identifies a general genetic disjunction between Northern and Southern populations that occurs around the North part of the South Island. This pattern has emerged for two species that share life history characteristic of pelagic larvae with a duration of greater than 10 days (Ross et al. 2009) namely the green-lipped mussel (*Perna canaliculus*) (Apte & Gardner 2002, Star et al. 2003) and the cushion star (*Patiriella regularis*) (Waters & Roy 2004, Ayers & Waters, 2005). Direct comparisons of multiple genetic methods for the determination of population differentiation across taxa are impossible, but general patterns including the North-South phylogeographic split may be investigated with fine scale spatial sampling.

An East-West phylogeographic disjunct has been identified within the North Island. The amphipod *Paracorophium lucosi*, a brooding crustacean, showed genetic discontinuity between east and west on the North Coast of the North Island (Stephens & Hogg 2004).

Upwelling at the North tip of the North Island was cited as the probable cause for this disjunct although oceanographic sampling or modeling was not investigated (Stephens & Hogg 2004).

Disparate patterns of phylogeography across taxa suggest species-specific life history characteristics, larval behaviour and factors other than oceanic and nearshore currents affect the connectivity of marine invertebrates (Ross et al. 2009).

6.2.4 New Zealand oceanographic features

Major current systems of the South Pacific are disrupted and separate on the continental shelf of New Zealand (Ross et al. 2009). Two main surface water masses, the Subtropical and Subantarctic Waters, meet in the subtropical convergence zone (Heath, 1982). Subtropical Water derives from a split and subsequent partial eastward flow of the East Australian Current (EAC) at 31°S, 158°E in Australia (Stanton 1981). The eastward departure of the EAC is the Tasman Front, which meanders through the North Tasman Sea and reattaches to New Zealand's northeastern continental shelf as the East Auckland Current (EAUC) (Heath 1982, Roemmich & Sutton 1998, Sharples & Grieg 1998).

The northeastern continental shelf is an area of unique bathymetric features and predominant easterly winds promote seasonal upwelling (Booth 1974, Roemmich & Sutton 1998, Sharples & Grieg, 1998) to deliver nutrients to surface layers. The EAUC comes very close to shore around Cape Brett (Figure 6.1) (Bay of Islands) and summer downwelling events (four events over two months) associated with winds that change the net 'direction of near-bottom water movements to a northwest along-shelf flow with significantly reduced southeastern surface water flow' (Sharples & Grieg 1998) create a northwest moving current (Figure 6.1) (Booth 1974). During summer months, these downwelling events coupled with strong internal tides around the Bay of Islands could

direct the movement of larvae deeper in the water column from a general southeast direction to a northwest direction.
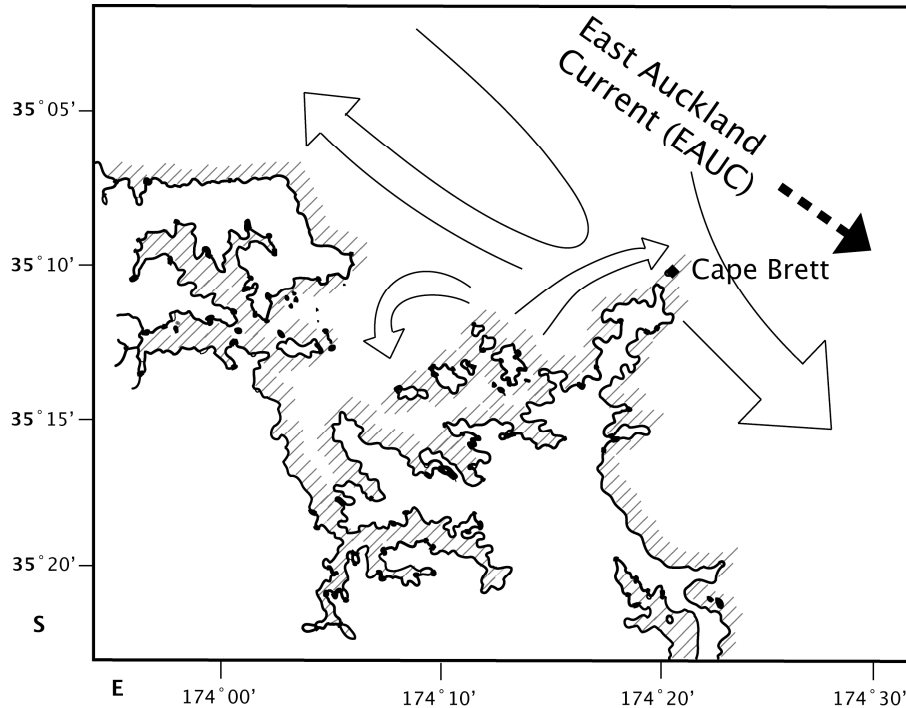
Figure 6.1 – Oceanic currents in the Bay of Islands. Image adapted from Booth, 1974.

Easterly flow of subtropical water across the Tasman Sea north of the Subtropical Convergence zone splits along the west coast of the South Island, forming the southward moving Southland Current and the northward moving intermittent (on long time scales, maximum of 4-6 weeks) Westland Current (Shirtcliffe 1990). The Westland Current forms depending on a prevailing southwesterly wind enduring for at least several days (Stanton 1976, Shirtcliffe 1990). At Cape Farewell, the Westland Current mixes with the D'Urville Current and flows eastwards into Cook Straight (Heath 1982). Periods of strong upwelling between Kahurangi Point (or a little further upstream) and Cape Farewell are dependent on the strength of the Westland Current (Shirtcliffe 1990).

6.2.5 Southern hemisphere *Mytilus* spp.

In the Southern hemisphere, *M. galloprovincialis* is naturally distributed on rocky to soft sediment intertidal zones throughout subtropical to subpolar waters in South America, New Zealand, Australia, and the Kerguelen Islands with an invasive population in South Africa. The native lineage of *M. galloprovincialis* in the Southern hemisphere is genetically distinct from the Northern hemisphere conspecific and has colonized South America, New Zealand and Australia for at least one million years (Hilbish et al. 2000, Gardner 2004, Gérard et al. 2008).

Mytilus chilensis in Chile is a homogeneous panmictic interbreeding population across an 1800km study distribution, average F_{ST} values 0.011-0.055 and Nei's genetic distance of 0.004-0.048 for 7 allozyme loci (Toro 2006).

6.2.6 Current investigation

Extensive intraspecific differentiation is possible (Quesada et al. 1995) among *Mytilus* populations even over relatively short distances. Population subdivision based on microsatellite allele frequencies from blue mussel samples located throughout mainland New Zealand and offshore islands will offer another piece of the picture for phylogeography of benthic marine invertebrates. The relative strength of identified barriers to gene flow will be assessed with estimates of contemporary and historical migration rates among populations.

The current investigation aims to incorporate classical population genetic measures based on microsatellite allele frequencies with contemporary Bayesian methods for the identification of population subdivision. Additionally, migration rate estimates and the relative influence of environmental variation on the distribution of allelic frequency profiles within and among New Zealand blue mussel populations is also presented.

6.3 Materials and Methods

6.3.1 Sample collection and taxonomic identification

Blue mussels were collected from 13 locations throughout the New Zealand mainland and offshore islands, and approx. 30 individuals were collected from each location for analysis with microsatellite loci. Collection dates, locations and sample sizes are listed in Table 6.1 and maps of sample locations on North and South Islands in Figures 6.2(a,b), respectively. In New Zealand, the presence of non-indigenous *M. galloprovincialis* from the Northern hemisphere was documented in Chapter 3 using two RFLP assays (one targeting the mtDNA 16s gene and the other targeting the nDNA Me15/16 gene). The Me15/16 RFLP assay was also applied to all 388 individuals genotyped for microsatellite loci (see Chapter 3, Section 3.3.2 for methods). Because the presence of non-indigenous individuals may bias the population genetic parameters estimated here, analyses were repeated without the Northern conspecifics or hybrids where indicated (the alternative data set). Northern conspecifics were classified based on nuclear Me15/16 RFLP genotypes as either a pure or hybrid individual. Nuclear genotypic designation for hemisphere of origin is in Table 6.1, the number of individuals genotyped as 'MgS' is the sample size for the alternative data set.
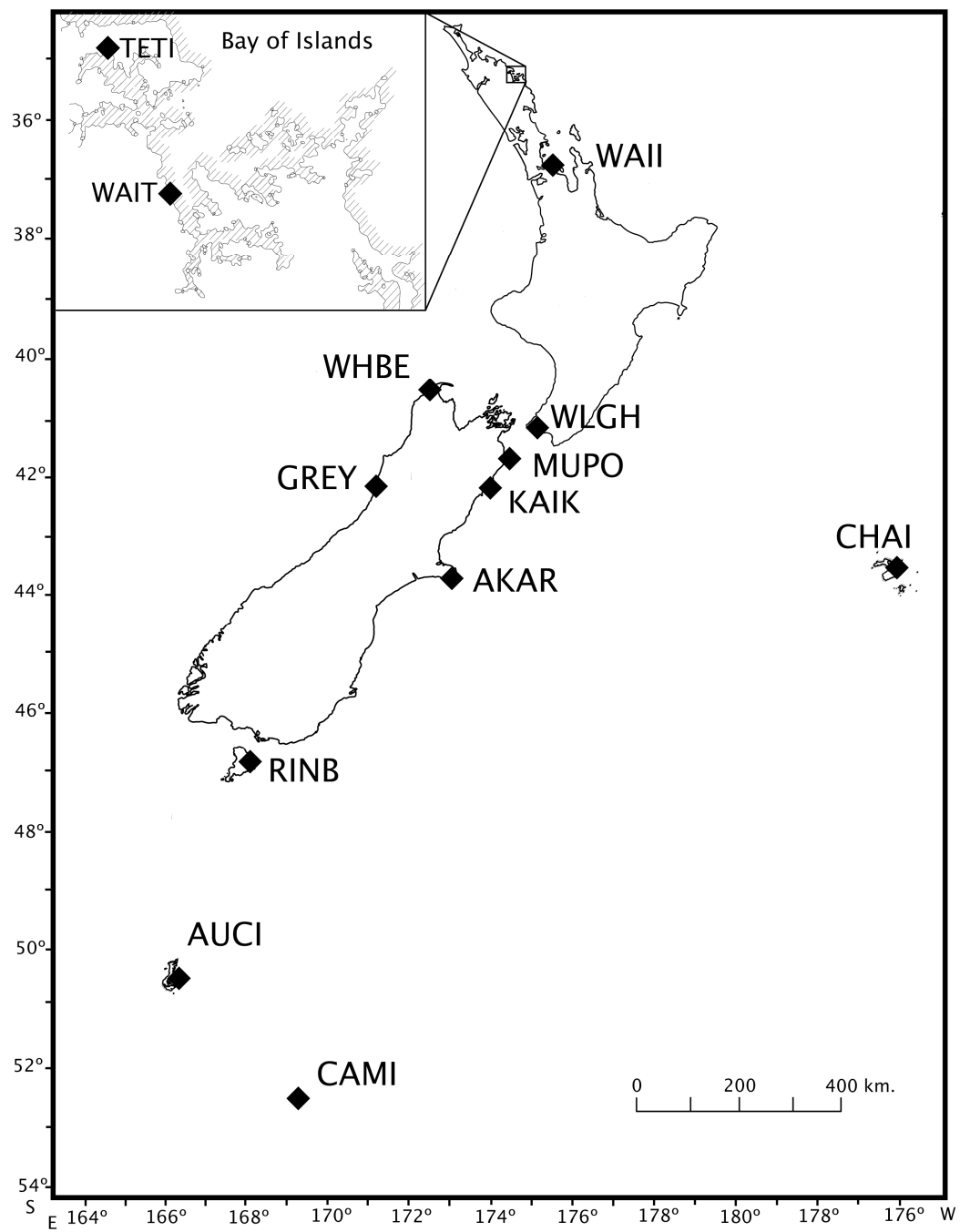
Figure 6.2 – Map of sample locations (codes in Table 6.1) in New Zealand. Inset is Bay of Islands.

Table 6.1 – Sample locations, codes, geographical coordinates, collection dates, sample sizes and genotype assignment*

Sample Location	Code	Latitude	Longitude	Date Collected (dd-mm-yy)	Sample Size (n)	MgS/ MgN/ Hybrid
TeTii Bay	TETI	35º 08.768' S	174º 00.258' E	04-10-05	30	30/0/0
Waitangi	WAIT	35º 16.718' S	174º 05.383' E	06-10-05	29	29/0/0
Waiheke Island	WAII	36º 47.272' S	175º 05.264' E	21-04-09	30	15/3/12
Whakariki Beach	WHBE	40º 30.223' S	172º 39.592' E	26-07-08	30	25/3/2
Wellington Harbour	WLGH	41º 17.000' S	174º 50.000' E	03-18-07	30	17/4/9
Mussel Point	MUPO	41º43.989' S	174º 16.052' E	16-01-08	30	23/1/6
Kaikoura	KAIK	42º 24.260' S	173º 41.103' E	19-02-08	29	28/0/1
Greymouth	GREY	42º 27.457' S	171º 12.357' E	06-03-08	30	30/0/0
Chatham Islands	CHAI	43º 35.956' S	176º 39.939' W	01-12-02	30	30/0/0
Akaroa	AKAR	43º 40.323' S	172º 57.914' E	20-02-08	30	30/0/0
Ringitingi Beach	RINB	46º 53.259' S	167º 59.907' E	27-01-07	30	30/0/0
Auckland Island	AUCI	50º 29.391' S	166º 16.767' E	16-02-04	29	29/0/0
Campbell Island	CAMI	52º 31.600' S	169º 06.835' E	20-03-04	30	30/0/0

* from Me15/16 RFLP (MgS – *M. galloprovincialis* Southern hemisphere, MgN – *M. galloprovincialis* Northern hemisphere, and hybrid MgS/MgN) for all sampled locations in New Zealand.

6.3.2 Molecular methods

Nine microsatellite loci were chosen from several previously published papers based on amplification success and presence of polymorphism in Southern hemisphere *M. galloprovincialis*. Source, repeat motif and expected size range are described in Table 6.2.

Table 6.2 – Microsatellite characteristics from previously published papers.

Locus	Repeat Motif	Size Range (bp)	Species (development)	Source
Mgµ1	$(TG)_n$	168 - 208	*M.galloprovincialis* N	Presa et al. (2002)
Mgµ3	$(TG)_n$	143 - 151	*M.galloprovincialis* N	Presa et al. (2002)
Mgµ5	$(TTTG)_n$	140 - 158	*M.galloprovincialis* N	Presa et al. (2002)
MGE003	$(GTG)_5$	450 - 456	*M.galloprovincialis* N	Yu & Li (2007)
MGE004	$(CTA)_6$	168 - 180	*M.galloprovincialis* N	Yu & Li (2007)
MGE005	$(TGA)_6$	220 - 244	*M.galloprovincialis* N	Yu & Li (2007)
MGE007	$(CAA)_5$	242 - 260	*M.galloprovincialis* N	Yu & Li (2007)
MT203	$(CA)_8$	161 - 197	*M. trossulus*	Gardeström et al. (2007)
MT282	$(GT)_9$	336 - 354	*M. trossulus*	Gardeström et al. (2007)

Repeat motif is listed in 5'-3' direction. Species (development) is the species the locus was developed from and *M. galloprovincialis* N is the Northern hemisphere lineage of *M. galloprovincialis*. Source is the study from which the locus was identified.

See Chapter 2B (Section 2.3.2B) for genomic DNA extraction methods. PCR conditions are listed in Table 6.3 along with realized size ranges for all loci found in *M. galloprovincialis* collected from New Zealand sample locations. A random (evenly spaced across loci) 10% subset of individuals was sequenced to confirm correct microsatellite repeat motif. Fluorescent-labeled primers were used to amplify PCR products for genotype fragment analysis on an Applied Biosystems (ABI) 3730 Genetic Analyzer by technical staff at the Allan Wilson Centre Genome Service (http://www.allanwilsoncentre.ac.nz). Allelic fragment sizes were identified with Peak Scanner V1.0 (Applied Biosystems) using the internal size standard GeneScan™ - 500 LIZ™(Applied Biosystems).

Table 6.3 – PCR MgCl$_2$ concentrations, cycling conditions and observed size range for nine loci.

Locus	Temp. (°C)	MgCl$_2$ (mM)	PCR Cycling Conditions	Size Range (bp)
Mgµ1	TD	1.8	94°C, 2'; [94°C, 30"; 60°/55°C**, 30"; 72°C, 30"] X 50; 72°C, 15'; 4°C, ∞	128-224
Mgµ3	TD	1.5	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	n/a
Mgµ5	TD	1.5	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	131-151
MGE003	TD	1.5	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	364-474
MGE004	48.5	2.1	94°C, 2'; [94°C, 30"; 48.5°C, 30"; 72°C, 30"] X 40; 72°C, 15'; 4°C, ∞	144-228
MGE005	TD	1.5	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	217-253
MGE007	TD	1.6	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	239-254
MT203	53.0	1.5	94°C, 2'; [94°C, 30"; 53°C, 30"; 72°C, 30"] X 35; 72°C, 15'; 4°C, ∞	117-245
MT282	TD	1.5	94°C, 2'; [94°C, 30"; 60°/50°C*, 30"; 72°C, 30"] X 30; 72°C, 15'; 4°C, ∞	324-388

n/a size range means the locus has an imperfect allele array. TD in Temp means a 'touchdown' PCR. * means annealing temperature of each cycle dropped 1°C to 50°C (a total of 20 cycles) and stayed at 50°C for the remaining cycles. ** means annealing temperature of each cycle dropped 0.1°C to 50°C (a total of 50 cycles).

6.3.3 Determination of discrete populations (K)

Identifying discrete populations from an almost continuous distribution (not continuous between Wellington Harbour and the Bay of Islands sample locations) of blue mussels was attempted by two methods. Bayesian clustering algorithms and distance based methods were employed in several ways to determine the correct number of discrete cryptic populations distributed throughout the New Zealand range.

6.3.3.1 Bayesian clustering methods

Considering the locations sampled throughout mainland and offshore New Zealand as a single sample, the program Structure (Pritchard et al. 2000) uses individual-based clustering to detect population structuring rather than defining groups of individuals into *a priori* populations. Structure employs Bayesian model-based clustering algorithms to (1) detect genetic structuring within multilocus genotypic data, (2) assign individuals to differentiated populations and, (3) assess percentage of each sample location's membership

in each cluster. Structure runs were completed under the assumptions of (1) correlated allele frequencies and (2) admixture within populations, meaning a proportion of alleles in the sampled generation arose from previous generations of other populations. *A priori* information on sampling location is used by Structure to detect clustering where weak or statistically insignificant clusters exist. Allowing individuals to be classified into sampling locations infers that ancestral characteristics may be shared within each location based on a likelihood that individuals from a single sample location are from the same population that is higher than the likelihood that individuals from separate sample locations are from the same population. There is no information included in the location prior about the actual or realized distance between sample locations. Where there is no correlation of population structure with location (i.e. the assumption of shared ancestry based on geographical proximity is not met), the model prior does not affect the inferred cluster information. The *a priori* assumption of location was used for all Structure runs. Each value of *K* was tested eight times and the Mann-Whitney U-test was used as an alternative to Bayes' Theorem (Pritchard et al. 2000) to assess the probabilities of different *K* values. Structure clustering algorithms have two main assumptions: (1) all populations are in Hardy Weinberg equilibrium (HWE), and (2) linkage equilibrium between loci within populations (Pritchard et al. 2000). Because population structure was unknown before Bayesian analysis, the assumption of HWE in populations was extended to an assumption of HWE in sample locations. Initial global HWE analyses (Genepopv4.3.1) indicated only four loci were in equilibrium for >50% of the sampling locations and none of the loci were linked. To best meet the assumptions of Structure's clustering algorithms, the analysis was repeated for loci in HWE for >50% of sample locations under the same configuration as listed above. The program BAPS v5.3 (Corander et al. 2008; Tang et al. 2009) was also

used to infer the genetic substructure of New Zealand blue mussels as an additive source of support for inferences of *K* from Structure.

To accommodate the potential bias associated by including non-indigenous Northern hemisphere *M. galloprovincialis*, Structure and BAPS computations were repeated without these individuals with the same configurations listed above. This amounts to three separate data sets analysed by Structure and BAPS; (1) all loci with all taxa, (2) HWE loci with all taxa and (3) HWE loci with Southern *M. galloprovincialis* lineage.

6.3.3.2 Distance Based Methods

Under the step-wise mutation model (SMM), Cavalli-Sforza & Edwards (1967) Chord (D_C) and Nei et al. (1983) D_A distances perform better than other distance measures to reconstruct NJ trees (Takezaki & Nei 1996). Both of these distance measures were calculated for pairwise data sets; (1) all loci, (2) loci in HWE, (3) all taxa, and (4) Southern lineage *M. galloprovincialis*, with the program POPULATIONS 1.2.30 (© 1999, Olivier Langella, available at http://bioinformatics.org/~tryphon/populations/) and neighbour-joining (NJ) trees calculated in 'neighbor' of the PHYLIP package (Felsenstein 1989). NJ methods do not assume equal rates of evolution throughout all parts of the tree and may be more applicable for data sets where mutation rate constancy has not been estimated.

6.3.4 Hardy-Weinberg Equilibrium, F-statistics, linkage disequilibrium and Analysis of Molecular Variance (AMOVA)

Observed heterozygosity (H_O), expected heterozygosity (H_E) and conformation to Hardy-Weinberg Equilibrium were tested in Genepop v4.0.7 (Raymond and Rousset 1995, Rousset 2008) by two global tests with the alternate hypotheses of heterozygote excess and deficiency. The powerful global U-test (Rousset & Raymond 1995) is concerned with two

separate alternative hypotheses of heterozygotes excess and deficiency and p-values were estimated by a Markov Chain (MC) algorithm with 1000 dememorization steps and 1000 iterations for each of 20 batches (Guo & Thompson 1992). The MC parameters were considered sufficient if the p-value SE (standard error) was less than 0.01.

Tests for genotypic association between loci pairs were performed by Genepop. A global Fisher's exact test was performed on the log-likelihood ratio statistic for each pair of loci across all sample sites.

Population differentiation was also tested with Genepop, with populations corresponding to the sample sites. P-values are estimated by an exact test of log-likelihood (G) values. The null hypothesis is that genotypes are drawn from the same distribution in all sample sites. The test was repeated for all pairs of sample sites for all loci.

F-estimates were carried out by the program FSTAT (Goudet 1999) on two data sets comprising (1) sample locations as populations and (2) populations based on genetically distinct clusters as identified by Bayesian and distance methods. Indices of F_{IS} and F_{ST} are popular measures of inbreeding within and differentiation among populations, respectively. F_{IS} can be conveniently thought of as the probability that two individuals are identical by descent.

Under the infinite alleles model (IAM) (Kimura & Crow 1964), an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was designed to partition variance into four hierarchical categories: variance (1) among populations, (2) among sample locations within populations, (3) among individuals within sample locations and (4) within individuals using the program Arlequin V2.001 (Schneider et al. 2000). This analysis was repeated with three hierarchical partitions when sample locations were not considered and individuals were grouped into populations based on Bayesian/distance analyses: (1) variance among populations, (2) among individuals within populations and (3) within

individuals. Statistical significance of nested genetic variance was assessed with 10,000 non-parametric permutations within the hierarchies listed above for both analyses (Excoffier et al. 1992).

6.3.5 Temporal scales of migration rate and gene flow estimates

Classical population genetic measures such as F_{ST} uncover contemporary population structure as a function of historical connectivity and gene flow patterns. Inferring gene flow from indirect measures such as F_{ST} is only pertinent if equilibrium among opposing forces of selection, drift and mutation exists within a population that conforms to an island model. Even if these assumptions are met, this measure becomes one of historical importance only as marine populations may take hundreds to thousands of generations to reach equilibrium. Therefore this method of estimating gene flow becomes one of estimating historical patterns maintained by contemporary populations (Hedgecock et al. 2007). Historical estimates of migration rate were performed with sample locations pooled into populations using Genepop.

Measuring contemporary gene flow is possible through direct measures such as detection of first generation migrants and assignment of individuals to population of origin. The program GeneClass2 (Piry et al. 2004) was used to detect first generation migrants within the original framework of 13 sample locations. The configuration of GeneClass2 detection of first generation migrants analysis used the criterion 'likelihood of the individual genotype within the population where the individual has been sampled' (L_home) (Piry et al. 2004) to take into account unknown (not sampled) source populations because of potential long-distance dispersal from areas like South America and Australia. Individuals identified as first generation migrants were removed from the data set and the remaining individuals were used as reference populations to assign first

223

generation migrants to their correct population of origin. Assignment of first generation migrants is based on 10,000 iterations of a Monte Carlo resampling method of the simulation algorithm proposed by Paetkau et al. (2004) that takes reference population sample size into account (Piry et al. 2004) for all eight sampled loci for individuals with alpha threshold of $p < 0.01$, as this threshold has shown to be representative in simulated data sets (Paetkau et al. 2004). The possibility that the source population wasn't sampled has to be taken into account when interpreting migration estimates derived from the frequency and distribution of first generation migrants because GeneClass2 does not create virtual populations to account for this possibility. Assignment of first generation migrants to sampled populations was based on a Bayesian criterion (Rannala & Mountain 1994) used with the resampling method described above (Paetkau et al. 2004) with the threshold probability of $P > 0.1$. The difference between the exclusion threshold of $p < 0.01$ and the assignment threshold of $p > 0.1$ reduces the chance that individuals excluded from their population of collection are not assigned if the population of origin was not sampled (Underwood et al. 2007).

6.3.6 Spatial genetic structure

6.3.6.1 Isolation by distance

Testing for genotypic differentiation that follows the 'isolation by distance' model (Wright 1943) is important as a first point of analysis before attempting to define cryptic discreet populations. Isolation by distance was first described by Wright (1943) and is characterized by increasing genetic differentiation with increasing geographic distance. Populations within close proximity to each other exchange more migrants than populations further from each other, a measure closely related to the realized dispersal of the organism.

Two methods were used to assess the relationship between a pairwise genetic distance matrix and a pair-wise geographic distance matrix. Both approaches used sample locations as putative populations. Since both approaches infer geographic distance from the great circle distance between two pairs of latitudinal and longitudinal points on a map, realized distances in the oceanic environment do not conform to great circle distances and can be arbitrarily defined based on minimum coastal distance or oceanographic patterns. There is not enough information regarding the relationship between the timing of gamete release and oceanographic circulation patterns to accurately estimate the direction of larval dispersal for *M. galloprovincialis* in New Zealand, therefore the minimum coastal distance was determined for each population pair (Google Earth v1.2) (Table 6.4). Genetic distance was calculated as Slatkin's index of similarity (M) (Slatkin 1993) and a Mantel test employed tô determine the relationship between log transformed genetic similarity and geographic distance using the program Isolation By Distance Web Service (Bohonak 2002) (http://ibdws.sdsu.edu/~ibdws/). Log transformation of the genetic dissimilarity matrix and geographic distance is necessary for two-dimensional habitats (Slatkin 1993).

Table 6.4 – Pairwise minimum coastal distances (kilometers) among sample locations.

	AKAR	AUCI	CAMI	CHAI	GREY	MUPO	KAIK	RINB	TETI	WAII	WAIT	WHBE	WLGH
AKAR	-	913.2	1005.5	852.2	771.4	262.2	173.4	524.2	1409.7	1243.2	1403.2	516.0	335.1
AUCI		-	297.2	1498.5	1086.4	1178.7	1084.4	438.6	2102.2	2107.0	2096.5	1334.9	1235.9
CAMI			-	1436.0	1314.3	1272.5	1172.7	627.6	2304.1	2157.7	2298.3	1514.2	1324.4
CHAI				-	1296.4	790.8	813.9	1258.6	1316.2	1153.5	1307.9	1044.6	763.9
GREY					-	505.6	594.4	735.4	1104.3	1316.1	1096.7	251.8	490.8
MUPO						-	88.8	786.3	1147.5	981.0	1141.1	253.9	72.9
KAIK							-	697.6	1236.3	1069.8	1229.8	342.6	161.6
RINB								-	1933.8	1767.4	1927.4	1040.2	859.2
TETI									-	225.1	15.8	873.2	997.9
WAII										-	219.4	1098.3	943.1
WAIT											-	867.9	1005.0
WHBE												-	230.1
WLGH													-

6.3.6.2 Seascape genetics

Genetic differentiation among many marine species can be extremely low and interpretation may be increased by the addition of non-genetic information (Selkoe et al. 2008). Two general types of non-genetic variables are considered here are to be treated with different statistical approaches: (1) geographic information, and (2) environmental information. Spatial autocorrelation analysis, genetic landscape shape interpolation and using Monmonier's Algorithm to identify biogeographic boundaries are three statistical approaches integrating genetic and geographic data performed by the program Alleles in Space (AIS) (Miller 2005). For the following analyses, geographic latitude and longitude coordinates obtained from iTouch Map (http://itouchmap.com/latlong.html) were transformed into Universal Transverse Mercator (UTM) coordinates using the Research Coordination Network (University of Montana, United States of America) webservice (www.rcn.montana.edu/resources/tools/coordinates.aspx). Spatial autocorrelation analysis identifies non-random patterns of genetic relatedness to geographical placement and can determine at which specific spatial scales an identified pattern is significant (Sokal & Oden 1978, Manel et al. 2003, Miller 2005). Although a statistical relationship between geographical and genetic variation can be identified, Monmonier's maximum distance algorithm can be used to define exactly where discontinuities exist (Monmonier 1973,

Manel et al. 2003). AIS (Miller 2005) builds a Delaunay triangulation network overlayed with a genetic dissimilarity matrix derived from residual genetic distances. The algorithm follows the following three steps; (1) identify the largest genetic distance between two points on the triangulation and create the first barrier segment here, (2) follow the initial barrier until an external edge or previously defined segment is encountered, and (3) follow the barrier identified in (1) in the opposite direction until an external edge or previously defined segment is encountered (Monmonier 1973). The number of barriers to be identified must be defined beforehand. The algorithm was repeated with one and two barriers, the latter corresponding to the clustering of sample locations as identified by Structure, BAPS and neighbour-joining trees.

Landscape shape interpolation is designed to map genetic distances across a landscape and allow inference of genetic distance in geographical areas where no data was recorded. More appropriate to the current study is using this graph to visualize diversity patterns across the sampled landscape in relation to previously described genetic clustering of populations and biogeographical boundaries. The 3-dimensional plot obtained by this method has X and Y-axes corresponding to geographical locations and a Z-axis that corresponds to genetic distance.

Relationships among allelic frequencies and environmental variation across sample locations and within populations are explored using 'bioenv' and 'vegdist' routines in the package Vegan v1.17-2 (Oksanen et al. 2010). The routine 'vegdist' calculates a community dissimilarity matrix using the 'Bray-Curtis' index for allelic frequencies and Euclidean distances for scaled environmental variables. 'Bioenv' then finds the best subset of environmental variables to explain allelic frequency variation according to Spearman rank correlations among distances and dissimilarities. These routines are performed for all sample locations and within each population. Northwestern and Southwestern populations

contain too few sample locations to create the community dissimilarity matrix in 'vegdist'. Partial mantel tests ('mantel' routine in Vegan v1.17-2) (10,000 permutations) were performed among the highest-ranking variables to determine significance (with the correct scaling used by 'vegdist'). Note the maximum number of dissimilarity matrices for a partial mantel test is three (one dedicated to allelic frequency variation and two for environmental variation) and pair-wise tests are performed in cases where 'bioenv' identified more than two environmental variables to best explain allelic frequency variation. Nine environmental variables are derived from the New Zealand Marine Environment Classification scheme (Ministry for the Environment 2005) for each sample location as follows:

Variable	Identifier	Measurement
Winter sea surface temperature	sst_wint	^{O}C
Summer sea surface temperature anomaly	sst_anom	^{O}C
Annual amplitude of sea surface temperature	sst_anamp	^{O}C
Spatial gradient annual mean sea surface temperature	sst_grad	$^{O}C\,km^{-1}$
Tidal current (depth averaged)	tidal	$m\,s^{-1}$
Sediment type	sediment	categorical value
Freshwater fraction	freshwater	$0.01m^{-1}$
Winter solar radiation	rad_wint	Wm^{-2}
Annual mean solar radiation	rad_mean	Wm^{-2}

6.4 Results

6.4.1 Determination of K

Structure results for data sets containing all eight loci and (1) all taxa and (2) Southern hemisphere *M. galloprovincialis* revealed the presence of two separate clusters (Table 6.5). One cluster includes the sampling locations of Te Tii Bay, Whakariki Beach and a moderate representation of individuals from Greymouth whilst the other cluster includes the remainder of the sample sites (Figure 6.2a). The same pattern is evident when only loci in HWE (for >50% of the sample locations) for Southern lineage were tested (Table 6.5). Conversely, using only loci in HWE altered the results of the data set with all

taxa to show three clusters (K=3) (Table 6.5, Fig X). This clustering arrangement shows the previous structure of one cluster containing TeTii Bay and Whakariki Beach and part of Greymouth (obtained from K=2), a second cluster containing the majority of individuals at Wellington Harbour and Waiheke Island with just under 50% of individuals from Mussel Point, and a third cluster with the remaining sample locations (Figure 3a). The second cluster described above disappears with the removal of non-native individuals, it's important to note that Wellington Harbour, Waiheke Island, and to a lesser extent, Mussel Point, had a large proportion of non-native individuals that were removed and the entire Waiheke Island location was removed from the data set because 28 of 30 individuals had non-native alleles. It is impossible to determine if the population structure inferred by including all taxa and hybrids (K=3) is artificially created from the presence of non-native alleles with the knowledge that introgressive hybridization in New Zealand is common among these lineages (Chapter 3). Classifying taxonomic identity of each allele under study is not possible without reference microsatellite allele sizes from Northern hemisphere pure species. This investigation takes the cautionary approach to remove all non-native individuals and hybrids as identified from one nuclear gene, but has the added effect of not knowing if this selective removal affects inference of the true population structure. In analyses restricted to Southern lineage mussels, sample locations with non-native individuals removed indicate a higher proportion of individuals belonging to a single cluster (i.e. no admixture) (Figures 6.3b,c). The same general pattern is evident when the number of loci are restricted to HWE in >50% of the sample locations; the proportion of individuals belonging to their assigned cluster is larger with exceptions at Mussel Point, Ringitingi Beach and Waitangi (Table 6.5).

The data set that will be used to infer population structure is restricted to the Southern lineage and loci in HWE (although using all loci doesn't affect clustering

results). Although this data set is used in this case, it's important to note differences in clustering arrangements from including all taxa and more loci. This relates to two clusters (*K*=2) (herein referred to as populations): (1) the Northwest population consists of Whakariki Beach and TeTii Bay and (2) the Southern population includes the remaining sample locations (Figure 6.2). The proportion of individuals in either population from each sample location is presented in Table 6.6. Greymouth will be grouped with the Southern population but note that this location experiences intermediate gene flow with the Northwestern population. The data set using just Southern hemisphere lineage *M. galloprovincialis* is chosen only for analysis of population structure because this is important to investigate larval dispersal and biogeographic boundaries that naturally occur in New Zealand. However, most analyses from herein will include all taxa in order to describe the current state of population genetics in New Zealand with some analyses repeated for both data sets for comparative purposes.

Table 6.5 – Clustering information for four data sets and two programs (Structure and BAPS)

	DATA SET			
	ALL Loci		HWE Loci	
	Structure	BAPS	Structure	BAPS
All Taxa	*K*=2 [TETI+WHBE] [remaining]	*K*=3 [TETI+WHBE] [CAMI] [remaining]	*K*=3 [TETI+WHBE] [WLGH+MUPO+ WAII] [remaining]	*K*=3 [TETI+WHBE] [GREY] [remaining]
Southern hemisphere **M. galloprovincialis**	*K*=2 [TETI+WHBE] [remaining]	*K*=3 [TETI+WHBE] [CAMI] [remaining]	*K*=2 [TETI+WHBE] [remaining]	*K*=5 [TETI] [GREY] [WHBE] [WLGH+AUCI+CAMI] [remaining]

K is the number of inferred clusters and the codes for sample locations making up each cluster are listed in [ ] brackets. 'remaining' is used to denote sample locations not included in other clusters.

Population structure inferences from BAPS analyses are meant as a checkpoint for the Structure analyses. In this case, the two programs are only slightly incongruous when considering just Southern lineage mussels and all loci, with the addition of Campbell Island as a separate population in BAPS.

Table 6.6 – Proportion of individuals from each sample location belonging to *K* populations. (NW – Northwestern and S – Southern)

| | | All Loci | | HWE Loci | |
Location	n	NW	S	NW	S
TeTii Bay	22	0.958	0.042	0.998	0.002
Wellington Harbour	16	0.194	0.806	0.051	0.949
Greymouth	30	0.292	0.708	0.268	0.732
Whakariki Beach	25	0.817	0.183	0.751	0.249
Mussel Point	18	0.000	1.000	0.010	0.990
Akaroa	29	0.005	0.995	0.003	0.997
Waitangi	19	0.011	0.989	0.052	0.948
Auckland Islands	28	0.098	0.902	0.084	0.916
Ringitingi Beach	30	0.002	0.998	0.004	0.996
Kaikoura	28	0.060	0.940	0.002	0.998
Campbell Island	30	0.060	0.940	0.002	0.998
Chatham Islands	30	0.012	0.988	0.010	0.990

Results from Structure (Pritchard et al. 2000). Data sets comprise only Southern lineage *M. galloprovincialis* and results are presented for data sets containing all loci and loci in HWE for >50% of the sample locations.
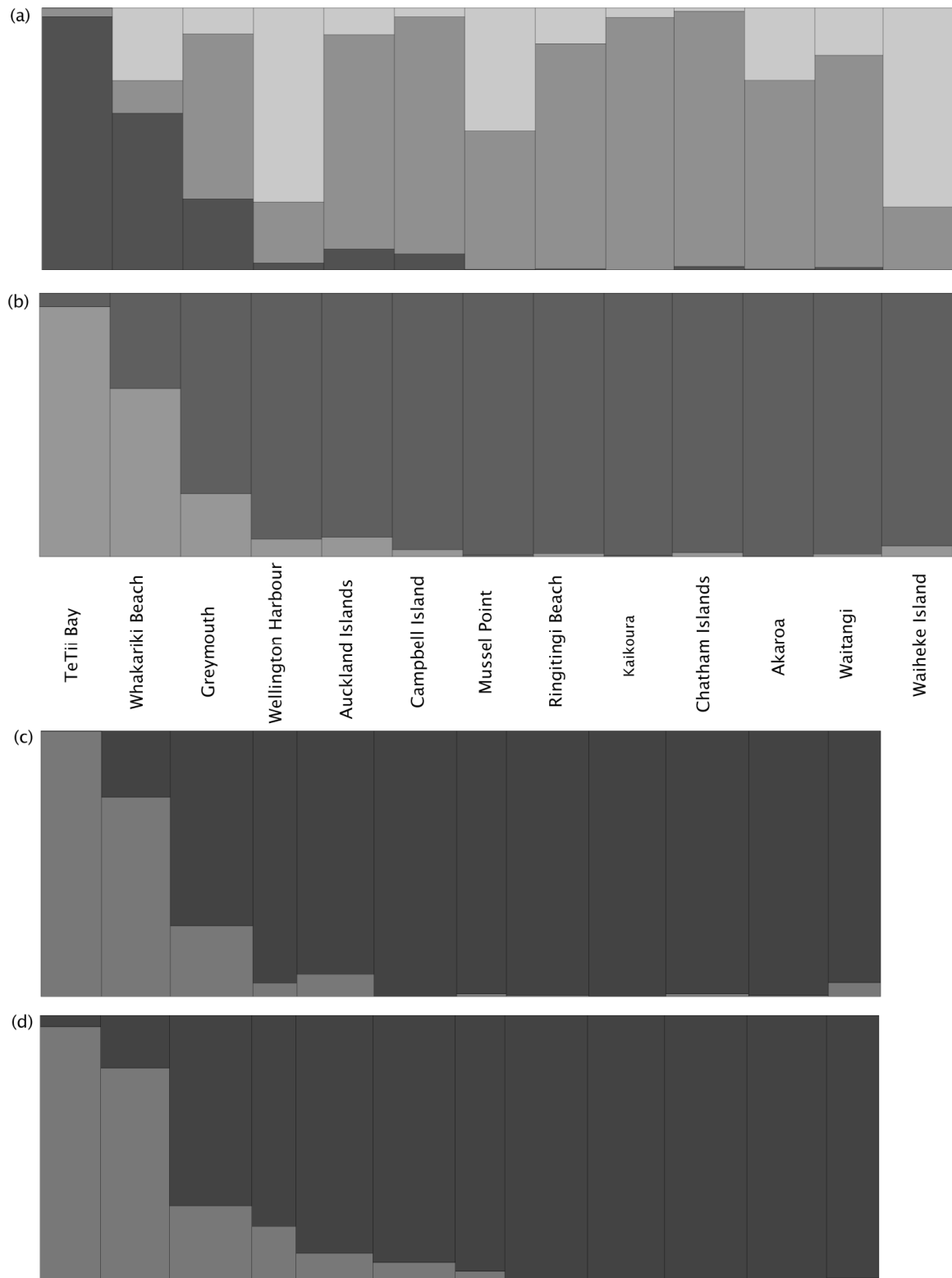
Figure 6.3 – Estimated population structure from Structure (Pritchard et al. 2000).
(a) all taxa and HWE loci (*K*=3), (b) all taxa and all loci (*K*=2), (c) Southern lineage *M. galloprovincialis* and HWE loci (*K*=2) and (d) Southern lineage *M. galloprovincialis* and all loci (*K*=2).

D_C distance NJ trees derived from data sets with all loci and all taxa (Figure 6.4a) and Southern lineage *M. galloprovincialis* all loci (data not shown) show slight differences in topology. A prominent cluster containing TeTii Bay and Whakariki Beach sampling locations is evident in both trees, whilst Campbell Island and Auckland Islands locations placed on a branch with other South Island locations when all taxa were analysed and on a branch with the Wellington Harbour and Waiheke Island locations when only Southern hemisphere *M. galloprovincialis* individuals were analysed. D_A trees for the same two data sets differed in topology from the D_C trees but not from each other (data shown for all taxa and all loci data set only, Figure 6.4b). Here, one cluster includes only TeTii Bay and Whakariki Beach, whilst the Greymouth location resolves within a cluster that includes mainly eastern South Island locations and the Chatham Islands. Like the D_C tree using the Southern hemisphere *M. galloprovincialis*, Campbell Island and Auckland Islands resolve onto the same branch as Wellington Harbour and Waiheke Island.
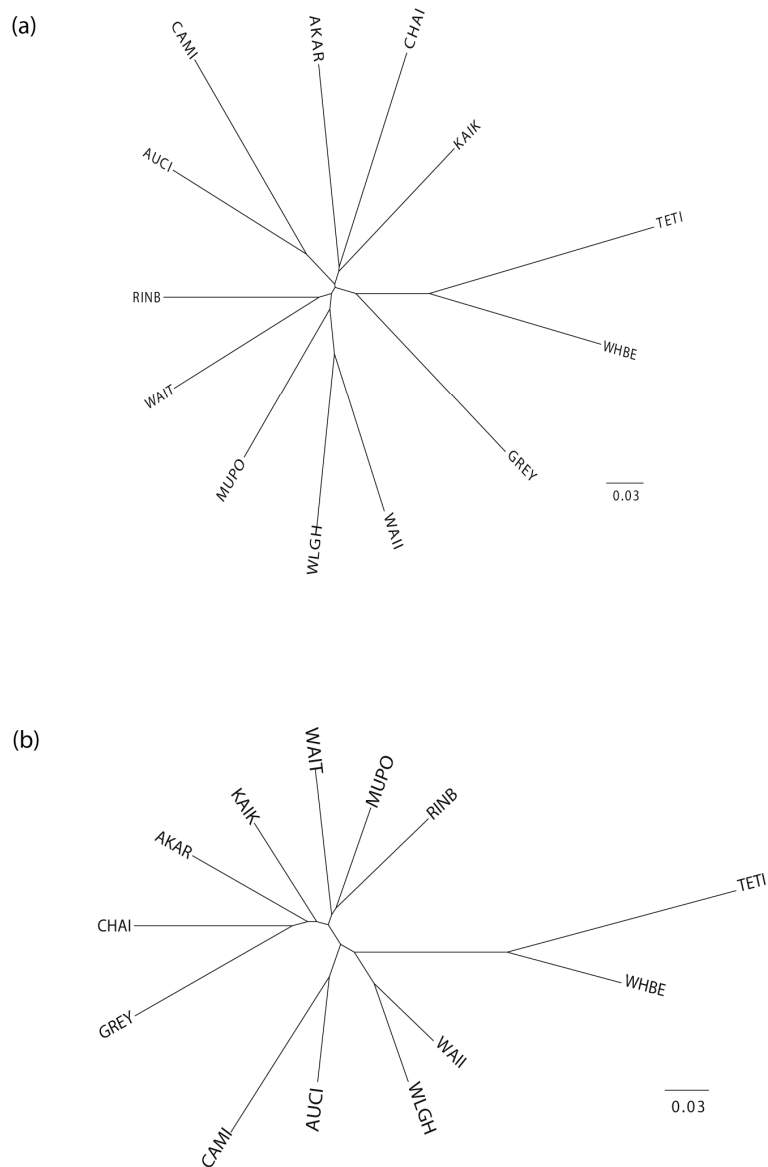
Figure 6.4 – Unrooted (a) D_C and (b) D_A distance neighbour joining trees for all loci and all taxa.

Two additional data sets including loci in HWE, one with all taxa (Figure 6.5b) and one with Southern lineage *M. galloprovincialis* only (data not shown), differed slightly from each other in topology between D_C and D_A measures. There is no difference between the data sets using all taxa and the Southern hemisphere lineage of *M. galloprovincialis* for each distance measure (data not shown). Between distance measures, Greymouth appears on a branch with TeTii Bay and Whakariki Beach in the D_C NJ tree and singularly on a long branch diverging from a mainly South Island sample location grouping in the D_A NJ tree. Other than the grouping of Ringitingi Beach, Waitangi, Akaroa, Chatham Islands and

Kaikoura at the top of both D_C and D_A NJ trees, the topologies of trees derived from two distance measures are quite different.

Topologies derived from HWE loci data differ greatly from the data sets including all loci and among measures. In the D_A NJ tree from HWE loci, TeTii Bay and Whakariki Beach form a more genetically distance population than on any other tree. The distance of the Greymouth sampling location on the D_A NJ tree from HWE loci is in line with BAPS analysis of the same data set and demonstrates that Greymouth is more closely related to the large cluster containing the majority of sample sites than to the TeTii Bay and Whakariki Beach cluster. This topology lends support to the classification of Greymouth part of the larger Southern population with intermediate gene flow between the Northwestern population.

(a)

(b)
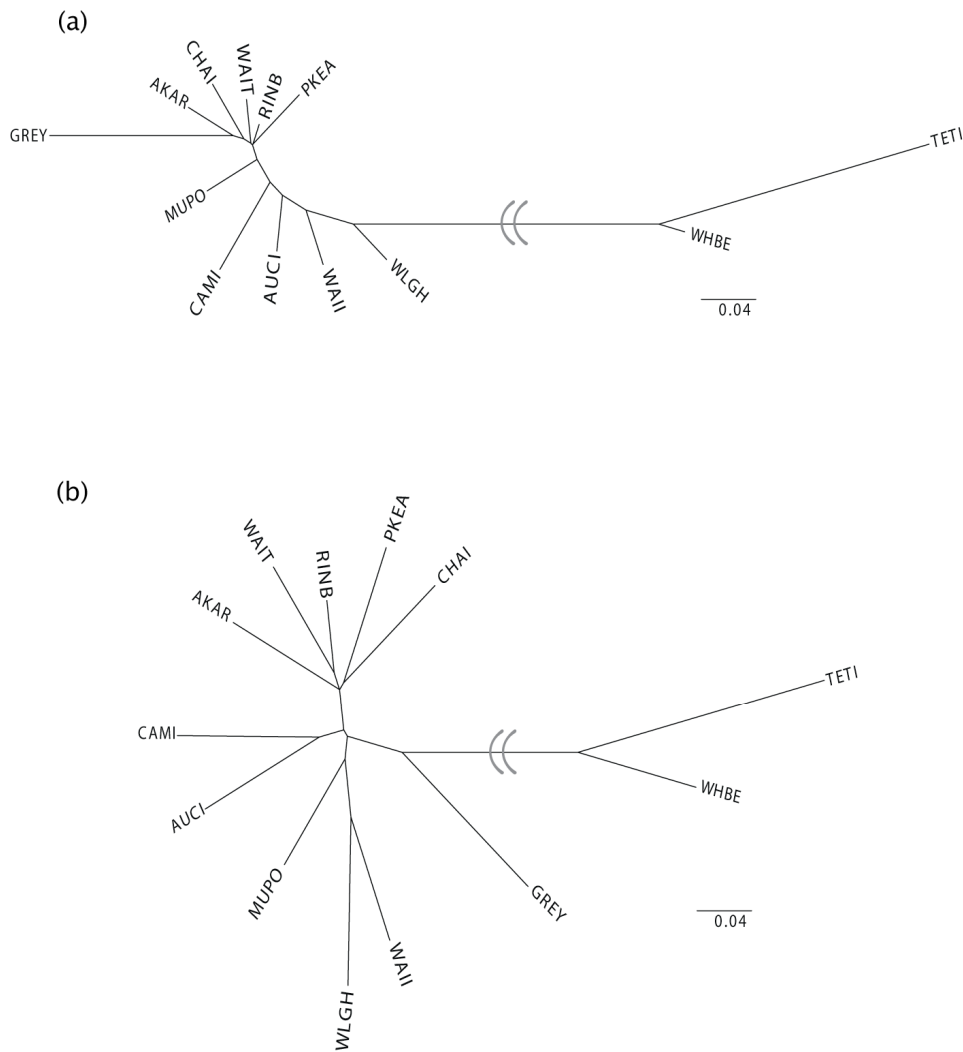
Figure 6.5 – Unrooted D_A (a) and D_C (b) distance neighbour-joining trees for loci in Hardy Weinberg equilibrium and all taxa. Population delineation is shown with curved break lines.

Results from Structure, Nei's genetic distance NJ tree (derived from HWE loci), Chord distance NJ tree, and to a lesser extent, BAPS, are concordant and display two genetically distinct populations (*K*=2) of blue mussels in New Zealand. The conclusions are made from the data set containing loci in HWE for the majority of sample locations as this is a very important assumption of Bayesian modeling algorithms. The Northwestern population consists of TeTii Bay and Whakariki Beach sampling locations. The Southern population includes the remainder of the sample sites. Waitangi is also included in this population, a surprising result considering this location is very close to TeTii Bay in the Bay of Islands (Northern New Zealand).

The two Bay of Islands sampling locations are ~15 km. apart but exhibit completely different genetic assignments. Resulting from the Southern hemisphere lineage *M. galloprovincialis* HWE loci data set Structure analysis, 99.8% of TeTii Bay (n=22) individuals belong to the Northwestern population, whereas 94.8% of Waitangi (n=19) individuals are grouped with the Southern population.

Taking into account the intermediate quality of Greymouth, the substructure of populations throughout New Zealand indicates a biogeographic split for *M. galloprovincialis* that occurs longitudinally around 174°E on the North Island, dividing sample locations into Eastern and Western groups. Sample locations on the South Island, other than Whakariki Beach, do not adhere to this split and are part of one population.

6.4.2 Allele Arrays, Hardy-Weinberg Equilibrium, F-statistics, linkage disequilibrium and Analysis of Molecular Variance (AMOVA)

Locus *Mgµ5*, derived from Presa et al. (2002), shows an imperfect allele array, meaning some alleles differ in length by non-integer multiples of the base repeat length and hence, mutations are not a simple loss or gain of single, perfect repeats. This allele array cannot be divided into sub-arrays (i.e. allele arrays containing alleles that differ in length by perfect integers of base repeat lengths) because the nucleotide composition of alleles at each sub-array is not known, and therefore, this locus was removed from the data set.

Global HWE tests reveal 'all loci' and 'all populations' conform to HWE (p=1.00000). Per population per locus analyses reveal that none of the populations conform to HWE across all loci, and conversely, none of the loci conform to HWE across all populations (Table 6.7). The loci *Mgµ5*, MGE003, MGE005 and MGE007 conform to HWE for >50% of the populations, and thus, these loci were used as a data subset for analyses dependent on the assumption of HWE (Structure, BAPS, etc). Significant

heterozygote deficits occur for all (significant) loci and population pairs except for locus *Mgμ5* at sample location WHBE (Whakariki Beach) where heterozygote excess occurs (P<0.01). The pooled data set with populations Northwestern and Southern do not conform to HWE expectations for 'all loci' (p<0.00000) and significant heterozygote deficiencies occur over all 'populations' (p<0.00000).

Tests for genotypic association of loci pairs reveal no significant disequilibrium, meaning all genotypes at all loci are independent (results not shown).

Table 6.7 – Hardy Weinberg Equilibrium test results. Observed allelic size ranges (in base pairs), observed heterozygosity (Ho), expected heterozygosity (He) and the significance value (P-val) for Hardy Weinberg Equilibrium tests for heterozygote deficiency for all loci at all sample locations separately and populations (in bold) corresponding to spatial structure inferred by Structure and distance (text in 50% grey) is significant at p<0.05. n/a indicates the locus was monomorphic for that sample location/populat

Locus		TeTii Bay	Waiheke Island	Wellington Harbour	Greymouth	Whakariki Beach	Mussel Point	Akaroa	Waitangi	Auckland Islands	Ringitingi Beach	Kaikoura	Campbell Island	Chatham Islands	**Southern**	**North West**
Mgu1	Observed size range (bp)	138–202	146–224	142–214	142–200	148–200	148–200	148–200	142–212	128–200	148–196	142–202	138–188	140–202	128–224	138–202
	Ho	0.83	0.80	0.53	0.70	0.73	0.52	0.53	0.61	0.43	0.50	0.59	0.50	0.43	0.41	0.62
	He	0.90	0.94	0.91	0.87	0.86	0.78	0.83	0.89	0.87	0.80	0.80	0.48	0.78	0.62	0.68
	HWE (*P-val*)	0.03	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.53	0.00	0.00	0.00
Mgu5	Observed size range (bp)	133–139	137–141	137–141	133–147	133–147	131–141	139–147	137–141	137–147	137–147	137–151	137–147	133–139	131–151	133– 147
	Ho	0.83	0.70	0.40	1.00	0.70	0.23	0.27	0.14	0.20	0.03	0.21	0.10	0.03	0.21	0.62
	He	0.57	0.52	0.44	0.64	0.57	0.39	0.35	0.19	0.19	0.22	0.20	0.09	0.03	0.24	0.47
	HWE (*P-val*)	0.01	0.04	0.76	0.00	0.56	0.01	0.07	0.25	1.00	0.00	1.00	1.00	1.00	1.00	0.00
MGE 003	Observed size range (bp)	376–379	373–472	370–382	364–379	370–379	376–379	373–379	373–379	373–382	373–379	367–382	367–475	376–475	367–475	370– 379
	Ho	0.20	0.17	0.30	0.52	0.47	0.30	0.33	0.21	0.23	0.13	0.42	0.37	0.33	0.23	0.30
	He	0.24	0.22	0.44	0.44	0.44	0.38	0.35	0.19	0.38	0.26	0.36	0.45	0.34	0.25	0.30
	HWE (*P-val*)	0.41	0.04	0.16	1.00	0.72	0.33	0.26	1.00	0.01	0.01	1.00	0.01	0.01	0.49	0.00
MGE 004	Observed size range (bp)	156–180	156–228	156–177	156–180	144–183	156–174	156–177	156–174	153–174	171–77	156–177	156–174	144–180	153–228	144–183
	Ho	0.37	0.47	0.67	0.63	0.50	0.30	0.70	0.24	0.63	0.43	0.37	0.80	0.69	0.44	0.37
	He	0.46	0.63	0.70	0.74	0.67	0.70	0.69	0.48	0.68	0.40	0.66	0.71	0.75	0.53	0.49
	HWE (*P-val*)	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.15	1.00	0.00	0.00	0.00	0.00	0.01
MGE 005	Observed size range (bp)	223–250	220–253	220–244	226–247	217–247	217–244	220–244	223–250	217–244	226–244	217–244	220–244	232–250	217–253	217-250
	Ho	0.67	0.63	0.33	0.60	0.63	0.43	0.57	0.69	0.59	0.27	0.69	0.67	0.63	0.43	0.47
	He	0.63	0.62	0.43	0.67	0.61	0.51	0.59	0.63	0.61	0.47	0.63	0.60	0.60	0.46	0.57
	HWE (*P-val*)	0.23	0.93	0.19	0.34	0.80	0.02	0.45	0.68	0.39	0.01	0.30	0.38	0.77	0.01	0.03
MGE 007	Observed size range (bp)	242–248	239–245	242–254	242–242	239–248	242–245	239–242	242–248	239–245	239–242	239–242	239–245	239–242	239–248	217– 250
	Ho	0.00	0.13	0.13	n/a	0.10	0.10	0.03	0.03	0.41	0.03	0.03	0.30	0.03	0.09	0.03
	He	0.13	0.19	0.34	n/a	0.22	0.10	0.03	0.03	0.40	0.03	0.03	0.27	0.03	0.11	0.14
	HWE (*P-val*)	0.00	0.24	0.00	n/a	0.00	1.00	1.00	1.00	0.57	1.00	1.00	1.00	1.00	0.00	0.02
MT203	Observed size range (bp)	165–199	175–217	117–209	171–219	175–211	173–207	175–245	173–219	155–215	171–209	177–217	175–195	175–205	165–211	117–245
	Ho	0.35	0.37	0.40	0.21	0.47	0.43	0.40	0.48	0.34	0.31	0.25	0.40	0.60	0.29	0.32
	He	0.90	0.91	0.92	0.84	0.91	0.91	0.92	0.90	0.88	0.89	0.85	0.76	0.80	0.70	0.70
	HWE (*P-val*)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MT282	Observed size range (bp)	338–362	334–388	334–354	340–378	342–364	340–372	346–372	344–364	334–370	324–366	338–370	334–370	340–376	338–364	334–388
	Ho	0.93	0.73	0.57	0.83	0.67	0.40	0.47	0.52	0.90	0.57	0.86	0.67	0.70	0.53	0.63
	He	0.86	0.89	0.84	0.84	0.85	0.85	0.78	0.80	0.87	0.88	0.81	0.80	0.81	0.67	0.67
	HWE (*P-val*)	0.00	0.03	0.00	0.51	0.01	0.00	0.00	0.00	0.01	0.00	0.19	0.07	0.20	0.05	0.00

F-estimates are presented in Tables 6.8 – 6.10. Inbreeding coefficients (F_{IS}) range from 0.089 at Campbell Island to 0.426 at Ringitingi Beach with an average of 0.2210 across all sample locations (Table 6.8).   Sample location significant pair-wise F_{ST} estimates range from 0.00650 to 0.24099 (Table 6.9). The clustering of sample locations into populations from HWE loci and Southern lineage *M. galloprovincialis* data supported by Structure and NJ trees has the pairwise F-estimate of 0.15935 (p=0.0000) (full data set). After removal of individuals with detected non-native alleles, the clustered populations reveal a pairwise F-estimate of 0.16728 (p=0.00000).

Table 6.8 – Sample location inbreeding coefficients (F_{IS}), all values are significant (p<0.05).

Sample Location	F_{IS}
TETI	0.106
WHBE	0.171
GREY	0.109
AKAR	0.279
AUCI	0.238
CAMI	0.089
CHAI	0.164
RINB	0.426
KAIK	0.217
MUPO	0.413
WAIT	0.295
WAII	0.190
WLGH	0.338

Table 6.9 – Pair-wise sample location F_{ST} estimates

	TETI	WHBE	GREY	AKAR	AUCI	CAMI	CHAI	RINB	KAIK	MUPO	WAIT	WAII	WLGH
TETI	-	0.06505	0.21707	0.23680	0.17143	0.23841	0.24099	0.23698	0.23492	0.22493	0.22418	0.18556	0.19189
WHBE	*	-	0.15963	0.16421	0.10983	0.15935	0.18945	0.17068	0.16250	0.14126	0.16966	0.12780	0.10818
GREY	*	*	-	0.02538	0.04498	0.06807	0.05474	0.07914	0.03162	0.02730	0.07017	0.04112	0.03645
AKAR	*	*	*	-	0.02433	0.05106	0.04560	0.04827	0.01272	0.00730	0.04577	0.03417	0.02544
AUCI	*	*	*	*	-	0.03425	0.05133	0.03731	0.02086	0.01389	0.03952	0.02623	0.02185
CAMI	*	*	*	*	*	-	0.05917	0.07997	0.03163	0.04896	0.08248	0.07258	0.06016
CHAI	*	*	*	*	*	*	-	0.08341	0.02535	0.04841	0.08132	0.06602	0.07395
RINB	*	*	*	*	*	*	*	-	0.03296	0.02412	0.02168	0.03233	0.04678
KAIK	*	*	*		*	*	*	*	-	0.00749	0.03678	0.03230	0.02966
MUPO	*	*	*			*	*	*		-	0.02508	0.01835	0.00650
WAIT	*	*	*	*	*	*	*	*	*	*	-	0.02370	0.04914
WAII	*	*	*	*	*	*	*	*	*	*	*	-	0.01035
WLGH	*	*	*	*	*	*	*	*	*		*		-

Estimates above the diagonal and significance of associated p-values below the diagonal (* is p<0.05).

Table 6.10 – Locus specific F-estimates with standard errors obtained from jackknifing over populations.

LOCUS	F_{IT}	F_{ST}	F_{IS}
Mgu1	0.303 ± 0.037	0.027 ± 0.016	0.283 ± 0.041
Mgu5	0.017 ± 0.092	0.122 ± 0.030	-0.116 ± 0.128
MGE003	0.131 ± 0.059	0.008 ± 0.007	0.124 ± 0.059
MGE004	0.216 ± 0.056	0.051 ± 0.017	0.174 ± 0.061
MGE005	0.101 ± 0.060	0.077 ± 0.057	0.027 ± 0.038
MGE007	0.767 ± 0.175	0.669 ± 0.181	0.254 ± 0.145
MT203	0.573 ± 0.032	0.026 ± 0.009	0.562 ± 0.033
MT282	0.217 ± 0.054	0.027 ± 0.011	0.195 ± 0.054
ALL LOCI	0.299 ± 0.076	0.080 ± 0.044	0.239 ± 0.084
95% CI	0.163 to 0.439	0.031 to 0.197	0.085 to 0.374

95% confidence interval was obtained from bootstrapping.

AMOVA analyses of ungrouped sample locations show more genetic variation among individuals within locations (20.23%) than among locations (8.48%) (Table 6.11, Scenario IV). When sample locations are grouped into populations, there is more variation among populations (15.04%) than among sample locations within populations (3.21%). Associated hierarchical F-estimates are in Table 6.11.

Table 6.11 - AMOVA results from grouped, pooled and individual sample location (HWE and all loci data sets).

Scenario	Source of variation	DF	Sum of Squares	Variance components	Percent variation	F-estimates
	Among populations	2	89.729	0.40924	15.04	$F_{CT} = 0.15041$
	Among sample locations within populations	10	79.462	0.08726	3.21	$F_{SC} = 0.03775$
I	Among individuals within sample locations	375	948.707	0.51761	19.02	$F_{IS} = 0.23271$
	Within individuals	388	611.000	1.70670	62.73	$F_{IT} = 0.37272$
	Among sample locations	12	114.643	0.16602	19.70	$F_{ST} = 0.19696$
II	Among individuals within sample locations	375	229.702	0.00879	1.04	$F_{IS} = 0.19696$
	Within individuals	388	232.500	0.66810	79.26	$F_{IT} = 0.20738$
	Among sample locations	12	181.210	0.20723	8.48	$F_{ST} = 0.08480$
III	Among individuals within sample locations	375	1024.124	0.49436	20.23	$F_{IS} = 0.22103$
	Within individuals	388	676.000	1.74227	71.29	$F_{IT} = 0.28708$

In scenario I sample locations are grouped into three groups corresponding to clustering/distance results. Second and third analyses are for comparison with no grouping or pooling of data sets containing (III) loci in HWE and (IV) all loci. All F-estimates have p<0.001.

6.4.3 Migration estimates

Genepop estimates of migration rate are 7.17 individuals per generation among all sample locations and 8.15 individuals when locations were grouped into populations. GeneClass2 estimate of the number of first generation migrants is 19 individuals among sample locations (with >99% probability). Discrepancies between Genepop and GeneClass2 (by identifying the number of first generation migrants) migration estimates are due to the correction for sample size employed in Genepop (see Barton & Slatkin 1986) and the inclusion of historical migration rates (by the nature of using F_{ST} values). Individuals identified as first generation migrants are listed in Table 6.12 with the assigned location of origin and the associated cluster membership values from Structure. These were used to determine the migration rate among sample locations. Then it was determined if the location of origin was in a different population and migration rates between populations were calculated.

Assignment (probability>0.1) of first generation migrants to source populations occurred for 9 individuals among sample locations and 2 individuals among populations (7 individuals are classed as within population migrants). Pair-wise population migration estimates derived from Genepop indicate an average 8.15 migrating individuals per generation between Northwestern and Southern populations and the assignment of first generation migrants identified by GeneClass2 is zero individuals between the same populations.

Table 6.12 – Migration of individuals based on p<0.01 for migrant probability.

Individual	Geographic Origin	Structure Q (NW/SW/E clusters, K=3)	GeneClass Assignment locality of highest probability	Between / Within (B/W) population migration	GeneClass migrant probability (L = L_home)	GeneClass highest assignment probability	Migrant Status
TETI 5	TeTii Bay	0.348/0.641/0.011	TeTii Bay	W	0.9999	0.0001	US
WAII 27	Waiheke Island	n/a	Wellington Harbour	W	0.9985	0.1458	WM
WLGH 1	Wellington Harbour	0.101/0.114/0.784	Greymouth	W	0.9925	0.4801	BM
WLGH 16	Wellington Harbour	0.078/0.035/0.887	Wellington Harbour	W	0.9990	0.0003	US
WLGH 25	Wellington Harbour	n/a	Kaikoura	W	0.9997	0.2677	WM
WHBE 15	Whakariki Beach	n/a	Auckland Islands	B	0.9939	0.0079	US
MUPO 6	Mussel Point	0.046/0.032/0.992	Greymouth	W	0.9935	0.1607	BM
MUPO 25	Mussel Point	0.124/0.035/0.841	Auckland Islands	W	0.9971	0.1531	WM
MUPO 27	Mussel Point	0.037/0.043/0.919	Waiheke Island	W	0.9979	0.0555	US
AKAR 9	Akaroa	0.008/0.000/0.992	Kaikoura	W	0.9971	0.0653	US
AKAR 21	Akaroa	0.007/0.000/0.993	Kaikoura	W	0.9979	0.1332	WM
WAIT 2	Waitangi	0.017/0.133/0.851	Auckland Islands	W	0.9992	0.0182	US
AUCI 28	Auckland Islands	0.159/0.287/0.553	Auckland Islands	W	0.9964	0.0044	US
RINB 12	Ringitingi Beach	0.022/0.038/0.940	Auckland Islands	W	0.9921	0.2453	WM
RINB 29	Ringitingi Beach	0.006/0.010/0.984	Auckland Islands	W	0.9980	0.3926	WM
PKEA 9	Kaikoura	0.005/0.065/0.931	Auckland Islands	W	0.9983	0.0136	US
PKEA 14	Kaikoura	0.005/0.066/0.929	Greymouth	W	0.9908	0.0915	US
CHAI 5	Chatham Islands	0.010/0.014/0.976	Auckland Islands	W	1.0000	0.2148	WM
CHAI 30	Chatham Islands	0.006/0.106/0.888	Auckland Islands	W	0.9939	0.0109	US

Migration status is read as US – unassigned due to lack of statistical power (probability <0.1), WM – within population migrant, and BM – between population migrant. n/a reflects non-native taxonomic status.

6.4.4 Spatial genetic structure

6.4.4.1 Isolation by distance

The measure of genetic relatedness (M) (Slatkin 1993) correlated with geographic distance did not indicate a significant isolation by distance relationship for log-transformed

or normal data. This result was also evident for the data set including only HWE loci and Southern hemisphere lineage *M. galloprovincialis* (data not shown).

6.4.4.2 Seascape genetics

Spatial autocorrelation analysis did not reveal significant correlation of geographic and genetic variability across all distance classes. The average pair-wise genetic distance across the entire data set was 0.535168. Six of ten distance classes revealed a significant (p=0.000) relationship between *Ay* and geographical distance (Figure 6.6). *Ay* values ranged from 0.492196 to 0.584707 (*Ay*=0 when individuals are genetically identical and *Ay*=1 when individuals are dissimilar). Three of six significant autocorrelation values cluster in the centre of the correlogram between approx. 600 and 1200 km and a further two significant values between approx. 0 and 400 km (Figure 6.6).
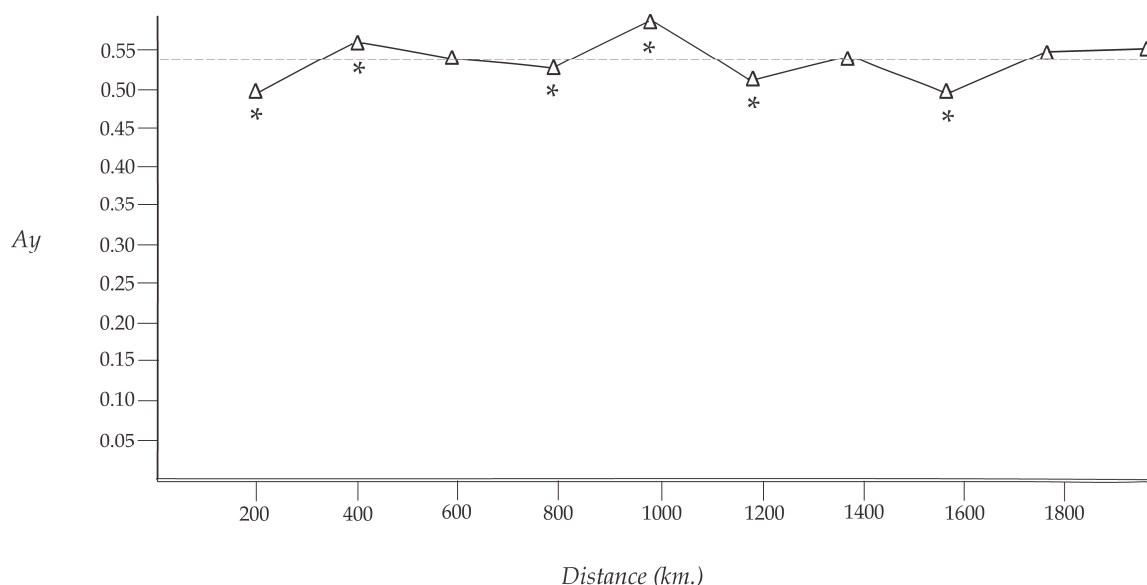

Distance (km.)

Figure 6.6 – Correlogram of autocorrelation coefficient (*Ay*) and distance (kilometers). * means significant results for that distance class (point is located at the maximum distance of the distance class). Average *Ay* (0.535168) across all sample locations denoted by dashed 50% grey line.

When set to find one biogeographic boundary, Monmonier's maximum distance algorithm identified a boundary between the Northwestern population (TeTii Bay and Whakariki Beach sample locations) and the Southern population (all other sample

locations), visualized on the Delaunay triangulation (Figure 6.7a) and on a map (Figure 6.7b). Results were the same across two sample data sets including all taxa and only Southern lineage *M. galloprovincialis*.
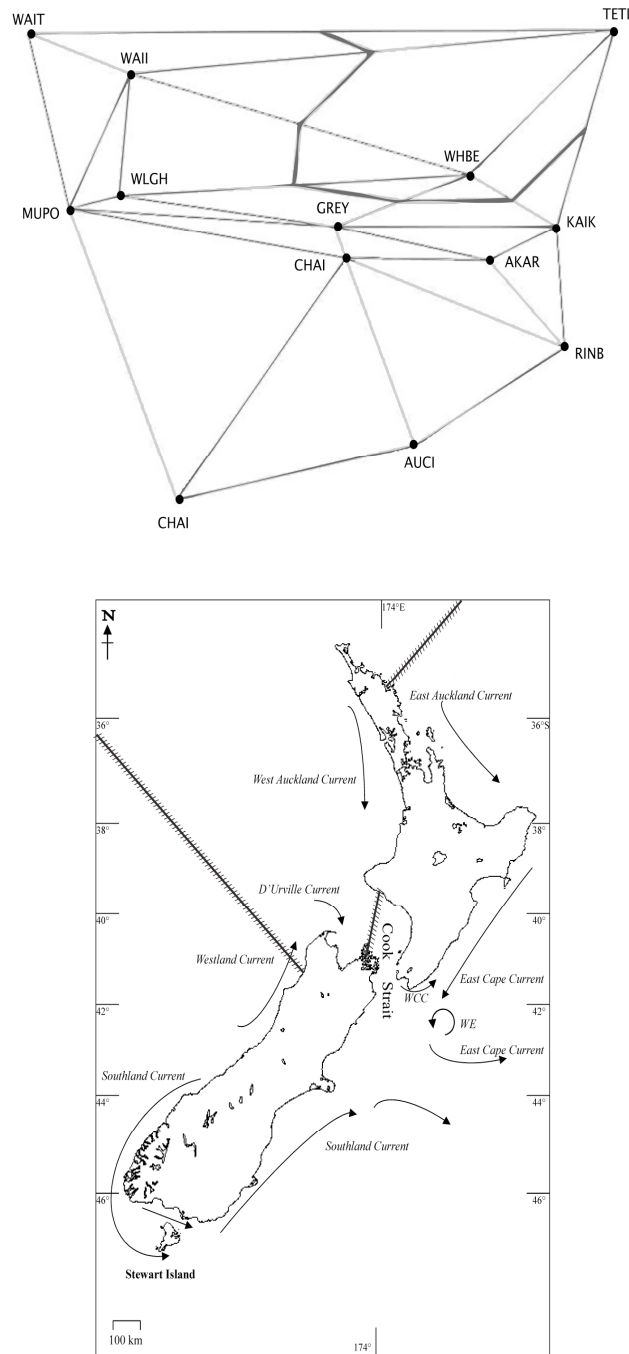

Figure 6.7 – Spatial genetic results.

(a) Delaunay triangulation where points represent sample locations with one biogeographic barrier overlayed (thicker black line). (b) biogeographic barriers visualized on a map (map without biogeographic barriers adapted from Helson & Gardner 2004).

The landscape shape interpolation graph also corroborates previously delineated populations and the biogeographic break. Genetic dissimilarity (Z-value) decreases in a South to North direction and is more prominent on the Western side, implying a general genetic discontinuity between South and North sample locations characterized by an isolation by distance relationship (Figure 6.8). A steep slope separates the Southwestern and Southeastern areas and also the Northwestern and Southwestern areas. There is a moderate slope between Northeastern and Northwestern areas. The steep slope between Northeastern and Southeastern areas indicates a population demarcation or biogeographic boundary although this demarcation has not been suggested by other inference methods.
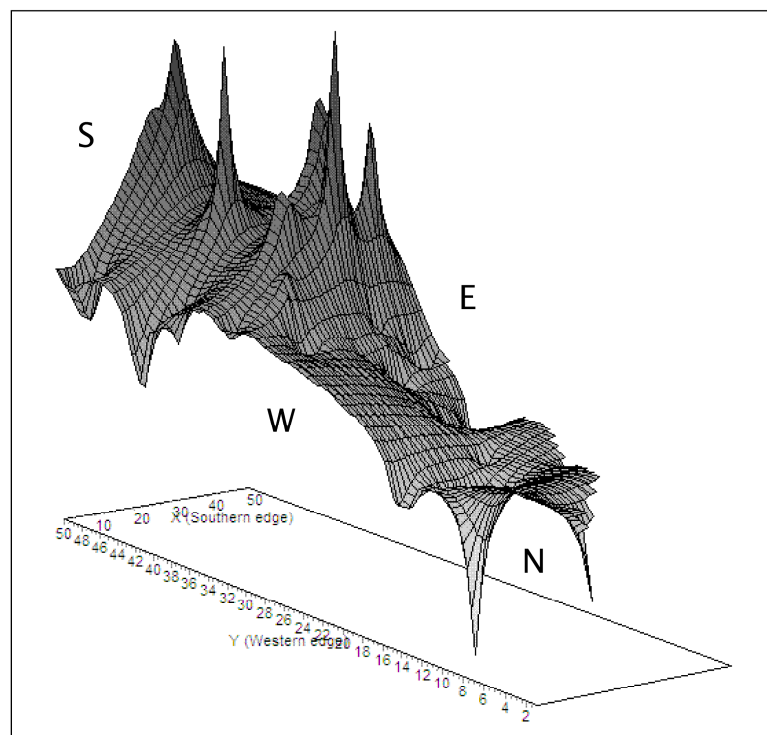

Figure 6.8 – Landscape Shape Interpolation graph.
(50X50 grid, *a*=1) from Alleles in Space (Miller 2003). X and Y-axes represent geographic distance and the Z-axis represents genetic distance.

The Southern population spans a large geographic distance (Bay of Islands to the subantarctic islands) and the genetic distance may be correlated to geographical distance on this spatial scale. Separate spatial autocorrelation analysis (AIS, Miller 2003) of the Southern population revealed an overall significant relationship ($p<0.05$) between geographical and genetic variation in 5 of the ten distance classes (0-200km., 200-400km., 800-1000km., 1200-1400km and 1400-1600km.) (data not shown). An isolation by distance Mantel test (Isolation by Distance Web Service, Bohonak 2002) also revealed a significant relationship between (log transformed) geographic distance and genetic variation (log transformed Slatkin's (1993) relatedness coefficient) ($Z = 118.0743$, $r = -0.5127$, $P<0.05$) (Figure 6.9).
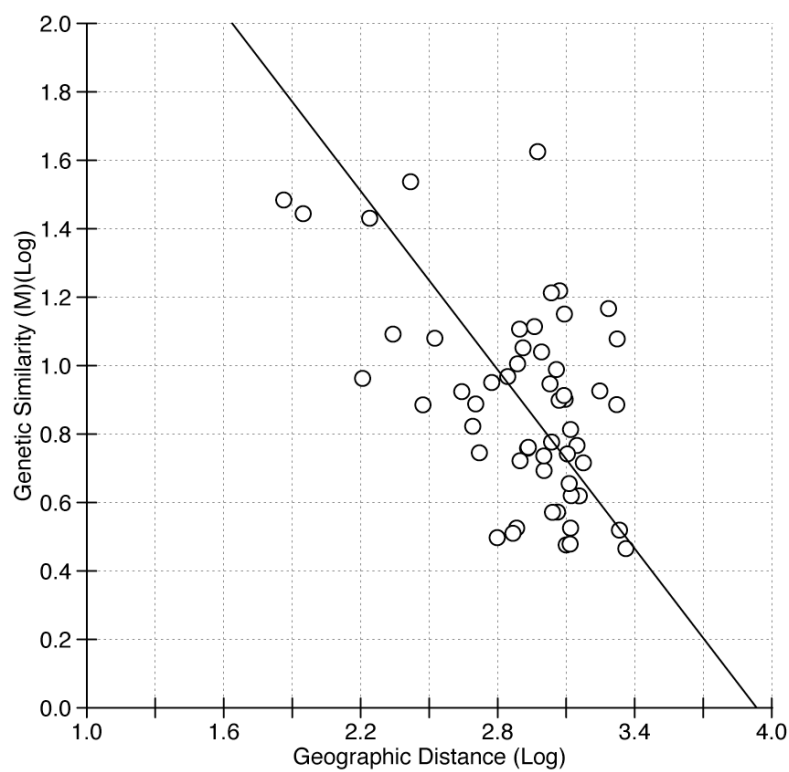

Figure 6.9 – Isolation by distance graph. Log transformed Slatkin's measure of genetic similarity (M) and geographic distance ($r=-0.5127$, $p<0.05$) for the Southern population.

Spatial patterns of genetic variation across the geographic landscape are concordant with previous estimates of population differentiation. Boundaries between Northwestern

and Southern populations are revealed in the Monmonier's maximum difference algorithm and reinforced by landscape shape interpolation analysis.

Environmental variation did not significantly correlate to the community dissimilarity matrix derived from all sample location allelic frequencies. The amount of freshwater input and winter sea surface temperature best explained allelic frequency variation across all sample locations (rho=0.2885), although this relationship was non-significant when assessed by the partial mantel test. The environmental variables sediment, spatial gradient annual mean sea surface temperature, winter solar radiation and winter sea surface temperature best explained allelic frequency variation within the Southern population (rho=0.5108). Spatial gradient annual sea surface temperature and winter solar radiation were significant variables when assessed by pair-wise partial mantel tests (p=0.04). These environmental variables join with geographic distance to best explain allelic frequency variation within the Southern population.

6.5 Discussion

6.5.1 Population determination, differentiation (F_{ST}) and AMOVA Analyses

Although Bayesian clustering and distance methods with the data set including all loci identifies two genetically distinct populations (with the exception of BAPS defining three), the assumption of HWE is paramount to the Bayesian algorithms and inferences from loci not meeting this assumption cannot be interpreted with high confidence (Pritchard et al. 2000). Bayesian analyses (Structure) based on loci in HWE support two genetically distinct populations of blue mussels throughout the New Zealand distribution. Chord and Nei's genetic distance based inference (HWE loci) support the Bayesian clustering results to determine two genetically differentiated populations.

The clustering of sample locations into two populations is supported by AMOVA analysis. When sample locations are ungrouped and unpooled, 19.70% of variation is found among sample locations (for loci in HWE). Grouping sample locations into populations results in a decrease of variation among populations (15.04%) but the variation between sample locations within populations remains lower at 3.21% and the addition of a fourth hierarchy reduces the total variation available to each step. The general trend of lower variation within populations than among populations supports the specific groupings of sample locations inferred from Bayesian clustering.

Significant F_{ST} of approximately 0.16 indicates the Northwestern population is greatly differentiated from the Southern population (Hartl & Clark 1997). Terrestrial population differentiation is highly variable depending on the level and robustness of barriers and can reach 100%, but widespread marine populations are often highly connected and F_{ST} values are significantly lower on the order of 1-2% (Palumbi 2003). The high F_{ST} value among Northwestern and Southern populations signifies a high degree of differentiation. This differentiation is more pronounced when data is restricted to only Southern lineage *M. galloprovincialis.*

6.5.2 Hardy-Weinberg Equilibrium and inbreeding (F_{IS})

Where populations and/or loci did not conform to HWE, significant heterozygote deficiencies were observed, a result concordant with many marine bivalves (Zouros & Foltz 1984) and common among *Mytilus* species (Skibinski et al. 1983, Zouros et al. 1988, Gaffney 1990). One of the most commonly presented explanations for this phenomenon is the Wahlund effect (the admixture of differentiated populations or cohorts within samples) and rings true in many *Mytilus* hybrid zones (Koehn 1991, McDonald et al. 1991, Vainola & Hvilsom 1991, Sarver & Foltz 1993, Quesada et al. 1995, Rawson & Hilbish 1995).

The presence of heterozygote deficiencies among Southern lineage *M. galloprovincialis* rules out a Wahlund effect (unless the current methods are not sufficient to identify introgressed individuals), and other factors potentially contributing to this genotypic array are (1) the presence of null alleles (Foltz 1986), (2) aneuploidy or parental imprinting (Chakraborty 1989) and/or (3) associative overdominance (Beaumont 1990).

6.5.3 The Northwestern (NW) population

The first population of interest is the Northwestern population comprising the sample sites of TeTii Bay and Whakariki Beach. It's interesting to note the exclusion of Northern hemisphere lineage individuals at Whakariki Beach makes the genetic composition of the two locations almost identical indicating that some genetic dissimilarity between locations is due to the presence of non-native mussels at Whakariki Beach. The southeast flowing East Auckland Current dominates the northern coast (TeTii Bay) but intermittent northwesterly flow during summer months would provide increased connectivity between NW sites. Sharples & Grieg (1998) stipulated the northwesterly current occurs near the bottom of their sample sites in the Bay of Islands and surface waters maintain an, albeit reduced, southeasterly flow. Dispersal direction indicated by genetic homogeneity between sample sites and oceanographic currents supports larval transport through lower layers in the water column. Spatial positioning behaviour has been noted before in European *M. edulis* larvae where the majority of planktonic larvae position within 1.5m to 3m and veligers and pediveligers concentrate at 3m (corresponding to the top of the thermocline) (Dobretsov & Miron 2001).

Connected sample locations across the northeastern peninsula contradicts previous investigations for the brooding amphipod *P. lucasi* (Stephens & Hogg 2004) and the chiton *Sypharochiton pelliserpentis,* with a short-lived pelagic larval phase (Veale 2007 in Ross et

al. 2009). Differences in dispersal abilities between the aforementioned species and *M. galloprovincialis* enhance connectivity through upwelling in the North Cape and across the D'Urville Current that flows eastwards through the South Taranaki Bight into Cook Strait. Although NW and S populations are significantly differentiated (F_{ST} =0.16, p=0.00), individual allelic frequency proportions of 19.4% and 18.3% at Wellington Harbour and Whakariki Beach, respectively, exist in opposing populations indicating a contemporary or historical localized moderate degree of connectivity between these sites. This connectivity between Whakariki Beach and Wellington Harbour does not extend past Cook Strait to Mussel Point or Kaikoura as the proportion of NW alleles in these two sample locations is negligible at 0.0% and 6.0%, respectively.

6.5.4 The Southern (S) population

The Southern population warrants special attention because it is the largest in geographical range and individual sample size, spanning ~2300 km. and eleven sample locations (n=327 individuals) from Waitangi (Bay of Islands) to the subantarctic Campbell Island and onto the west coast of the South Island at Greymouth. An overall significant isolation by distance relationship was detected within the Southern population and spatial autocorrelation analysis indicates the distance classes of 0 to 400 km and 600 to 1200 km are significant. The number of within population but between sample location first generation migrants being two for 0 to 400km. (n=89, ratio 1: 44.5) and one for 600 to 1200km. (n=59, ratio 1:59) spatial classes supports an hypothesis of limited gene flow. Isolation by distance has not been identified in *Mytilus* species on spatial scales of less than 150 km before and it is interesting to note this relationship is found at very small geographic distances of 0 to 400 km.

Larval dispersal characteristics of *P. canaliculus* should be similar to *M. galloprovincialis* as both species' larvae are pelagic for similar amounts of time. A significant isolation by distance relationship along a 1700 km geographic range was identified in *P. canaliculus* derived from biochemical variation at 11 allozyme loci (Gardner et al. 1996). A northerly and southerly split with reduced gene flow also characterizes *P. canaliculus* populations (Gardner et al. 1996) and fine spatial sampling reveals the abrupt disjunction occurs at ~42°S latitude (Apte & Gardner 2002, Star et al. 2003).

Oceanographic features of the Eastern coast of the South Island are relatively benign South of 42°S latitude, with no sustained upwelling or strong directional currents and weak downwelling. The general northward and onshore flow of water could entrain larvae close to parental populations wherefore self-recruitment would play a large role in maintaining the isolation by distance relationship. In the absence of upwelling, reduced nutrient input limiting phytoplankton growth could also limit pelagic larval survivorship over more than a few days, which would lead to limited dispersal within small spatial scales and again maintain the isolation by distance relationship.

Indeed, bottom-up community processes such as limited primary productivity associated with oceanographic features are correlated with reduced mussel recruitment and predation rates on South Island East coast areas of persistent downwelling South of 42°S (Menge et al. 1999, 2003). Reduced recruitment associated with bottom-up processes due to oceanographic features is a likely contributing factor to the observed isolation by distance relationship over small spatial scales along the East coast.

6.5.5 Absence of genetic differentiation across an upwelling zone on the northeast coast of the South Island

Two marine benthic invertebrate taxa with long lived (>10 days) pelagic larvae, *P. regularis* and *P. canaliculus*, show significant genetic structuring surrounding a zone of upwelling near 42°S latitude (east coast) and the results from the current study indicate this relationship is not evident for *M. galloprovincialis*. Significant upwelling in this area was cited as the most plausible cause for reduced gene flow between Northern and Southern populations. The existence of a Southern population for *M. galloprovincialis* that is connected throughout the entire East Coast of New Zealand (from the Bay of Islands to Campbell Island) suggests the barrier to gene flow for other species does not impact on the connectivity of blue mussels along this coastline.

Temporal variation in gamete release among taxa and the relative strength of the offshore water transport may be factors in maintaining genetic connectivity across this boundary. Larval vertical positioning in the water column could also prevent the offshore flow of larvae across the Chatham Rise (Shanks & Brink, 2005) by moving away from turbulent surface layers to deeper layers involved in cross-shelf flow (Bradford-Grieve et al. 1993).

6.5.6 Genetic discontinuity in the Bay of Islands

One cannot proceed much further without addressing the major genetic discontinuity that exists between two sample locations in the Bay of Islands, TeTii Bay and Waitaingi, separated by only ~15km. but with 95.8% of TeTii Bay individual genotypes belonging to the Northwestern population and 98.9% of Waitangi individual genotypes belonging to the Southern population. The abrupt and significant barrier to gene flow over a very small distance of 15km. directs attention to factors other than oceanographic circulation features that are working to prevent gene flow between these two sites.

The first hypothesis to explain restricted gene flow in this region is the presence of non-native alleles at one or both locations. Introgressive hybridization between lineages is a known phenomenon and microsatellite loci may have introgressed through hitchhiking or linkage with other genes whilst failing to exhibit non-native alleles in the taxonomic assays used because Northern alleles were not detected at either Bay of Islands site with the single nuclear gene. If both sites were inundated with non-native mussels, founder effects could lead to the observed differentiation. Conversely, just one of the sites could be affected by the presence of non-native alleles. The Bay of Islands is supported in phylogenetic (Chapter 4) and genealogical (Chapter 5) reconstructions as a divergent clade geographically separated from other sample locations in New Zealand. Locations included in these reconstructions are TeTii Bay, Dove's Bay and Russell Harbour (Chapter 4, Figure 4.3, haplotypes gNZ46, gNZ47 and gNZ49) and unfortunately there were no individuals sequenced from Waitangi to investigate the historical nature of this differentiation. If we consider Russell Harbour as a proxy for Waitangi (3.5 km. apart), genealogical reconstructions do not indicate a historical divergence. The observed divergence between TeTii Bay and Waitangi then seems fairly contemporary, which supports the hypothesis of non-native allele introgression.

Another hypothesis to explain restricted gene flow in this region is spatial behavioural differences between larvae and the relative influence of self-recruitment. It was hypothesized from oceanographic circulation patterns and connectivity between TeTii Bay and Whakariki Beach that TeTii Bay mussels spatially orient themselves lower in the water column to catch the northwest flowing current in the same area. Conversely, Waitangi mussels could spatially orient themselves higher in the water column to catch the southeast flowing East Auckland Current.

Differences in the timing of gamete release between sample locations could also partially explain genetic differentiation. In a study of East coast North American *M. edulis*, two proximal sample sites of mussels had contrasting spawning regimes due to the availability of food during maximum energy assimilation times (correlated to water temperature but generally starting in spring) (Newell et al. 1982). One location had high food availability in spring when the mussel is actively feeding and able to build up energy stores for gametogenesis whilst the other location had high food availability during winter when the mussel is not actively feeding and cannot assimilate as much energy (Newell et al. 1982). Slight temporal differences in gamete release between sample locations due to site-specific variation in food availability could potentially entrain larvae in different currents as the strength of the northwest bottom current is correlated to intermittent downwelling periods.

Mytilus species once thought to disperse over tens to hundreds of kilometers have presently been shown to have moderate to high self-recruitment within 20-30km from their natal population (Becker et al. 2007). Comparisons among species along the California coast show self-recruitment is localized in *M. galloprovincialis* but can be influential across the entire distribution in *M. californianus* (Becker et al. 2007). Differences among species are attributed to sheltered vs. exposed habitats, respectively. This could relate to the gamut of habitats utilized by *M. galloprovincialis* in New Zealand, from sheltered bays (Bay of Islands) to exposed rocky shores (subantarctic islands). Self-recruitment as a function of habitat type or other variable would be an interesting avenue of research for New Zealand *Mytilus* populations.

6.5.7 Estimates of migration rate

Gene-frequency differences between populations are maintained by limited migration, and ten migrants per generation are enough to make gene-frequency differences non significant (Palumbi 2003). Historical estimates of migration rate are on average higher than estimates obtained by identifying the number of first generation migrants. The overall historical migration rate among populations is 8.15 individuals per generation. The historical migration rate is close to the cutoff considered necessary for panmixia, therefore contemporary processes are causing the observed gene frequency differences. The West coast was covered with sea ice during the last glacial maximum ~17 Kya (Stevens et al. 1995) and recolonisation could have occurred during interglacial periods where sea levels around New Zealand were higher (Stevens et al. 1995).

6.5.8 Phylogeographic breaks in New Zealand

The phylogeographic break for *M. galloprovincialis* identified by Monmonier's Algorithm (Monmonier 1972) (AIS) correspond exactly to the genetic structure detected by Bayesian (Structure) and distance-based methods. The Northern end of this break occurs in the middle of the Bay of Islands (~174°E), represented by almost complete genetic differentiation of two sample locations in this area. Between sites, differences in vertical positioning of larvae in the water column and/or temporal variation in gamete release coinciding with downwelling events that change current direction are two main hypotheses driving the observed patterns of genetic differentiation. The South end of this break is located in the Taranaki Bight West of Cook Strait and prevents sustained gene flow between east and west coasts of the South Islands. Moderate levels of genetic homogeneity between Whakariki Beach and Wellingon Harbour sample locations suggest this barrier is semipermeable to gene flow.

The current study does not identify the known phylogeographic break for the Northeast coast of the South Island (Apte & Gardner 2002, Star et al. 2003, Waters & Roy 2004, Ayers & Waters 2005, Ross et al. 2009) and again, vertical positioning in the water column is a strong hypothesis to explain connectivity over a persistent upwelling zone (42°S) in *M. galloprovincialis* where *P. canaliculus* shows population divergence with limited gene flow (Apte & Gardner 2002, Star et al. 2003).

6.5.9 Seascape genetics

Environmental variation within the described variables did not significantly explain genetic variation when applied across all sample locations but within the Southern population, spatial gradient annual mean sea surface temperature (SSTgrad) and winter solar radiation significantly affect allelic frequency variation. A validation survey of marine classification variables shows the relative importance of each variable in models of chlorophyll *a*, benthos and fish populations (Snelder et al. 2004). SSTgrad, the smoothed magnitude of the spatial gradient of annual mean sea surface temperature, does not correlate to chlorphyll *a*, is moderately correlated to fish populations and has a small degree of correlation with benthic models involving *Echinocardium* species (Snelder et al. 2004).

Winter solar radiation is calculated as the extra atmospheric solar radiation in June modified by mean annual cloud cover and was designed to accurately represent areas with differences in summer and winter estimates but similar annual means (Snelder et al. 2004). Validation of this measure shows it is important to chlorophyll *a* and fish models with moderate correlation to shelf (combination of three benthic shelf presence/absence surveys includes all benthic species) models (Snelder et al. 2004). For suspension feeding *Mytilus* species, the production of chlorophyll *a* is an important determinant in fitness and

fecundity (Newell et al. 1982). The spawning cycle(s) of *Mytilus* in New Zealand is currently unknown, but in the Northern hemisphere, increased winter primary productivity (in relation to summer increases) has shown to reduce energy assimilation and lead to unsynchronized or intermittent spawning (Newell et al. 1982). Annual mean solar radiation was not correlated with allelic frequencies therefore areas with different summer and winter solar radiation amounts seem to be more affected. When expressed as a percent deviation from the mean, winter solar radiation was most reduced at Waitangi (56% of the mean) and the highest at Campbell Island (82% of the mean). An increase in winter solar radiation at Southern latitudes could drive a response like *Mytilus* in the Northern hemisphere (Newell et al. 1982).

6.5.10 Conclusions

The evidence of genetic differentiation among *M. galloprovincialis* sample locations throughout New Zealand indicates many processes that are working together to create and maintain phylogeographic boundaries with varying amounts of gene flow or different historical connectivity patterns. The importance of larval behaviour, temporal variation in gamete release and oceanographic circulation patterns are highlighted as contributing to observed genetic population connectivity. Further investigation is warranted to determine the driving factors behind two highly differentiated populations existing within ~15km. of each other (The Bay of Islands). Characterization of the genome to determine the degree of introgression with non-native mussels may be sufficient to explain this divergence.

Great population differentiation (F_{ST} = 0.16) (Hartl & Clark 1997) is more indicative of limited dispersing taxa, suggesting the barriers to gene flow are, or have been historically, strong enough to prevent substantial genetic exchange. Historical estimates of

gene flow between populations is just under the amount considered necessary to maintain panmixia but contemporary estimates based on first generation migrants show a different story with a migration rate of zero. This indicates contemporary processes are affecting gene flow between the studied populations more than they did in the past.

Small spatial scale isolation by distance within the Southern population suggests local or self-recruitment is pervasive and supports modern observations of strong self-recruitment within 20-30 km of parental populations (Becker et al. 2007) and maximum dispersal distances of 64 km (Gilg & Hilbish 2003) to 94 km (McQuaid & Phillips 2000).

The observations and hypotheses provided in this study highlight the importance of small-scale spatiotemporal processes in creating and maintaining population differentiation. However, further hypothesis-driven investigation is warranted to assess the relative contributions of spatiotemporal processes on the population dynamics of *Mytilus* in New Zealand.

6.6 List of References

Apte S, Gardner JPA (2002) Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology* 11: 1617-1628.

Apte S, Holland BS, Goodwin LS, Gardner JPAG (2000) Jumping ship: a stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions* 2: 75-79.

Ayers KL, Waters JM (2005) Marine biogeographic disjunction in central New Zealand. *Marine Biology* 147: 1045-1052.

Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56: 409-416.

Bay LK, Crozier RH, Caley MJ (2006) The relationship between population genetic structure and pelagic larval duration in coral reef fishes of the Great Barrier Reef. *Marine Biology* 149: 1247-1256.

Bayne BL, (1965) Growth and delay of metamorphosis of the larvae of *Mytilus edulis* (L). *Ophelia* 2: 1-47.

Beaumont A, Gjedrem T, Moran P (2006) Blue Mussel – *M. edulis* and Mediterranean mussel – *M. galloprovincialis*. In: Crosetti D, Lapègue S, Olesen I, Svaasand T (eds) *Genetic effects of domestication, culture and breeding of fish and shellfish, and their impacts on wild populations. GENIMPACT project: Evaluation of genetic impact of aquaculture activities on native populations*. A European network. WP1 workshop Genetics of domestication, breeding and enhancement of performance of fish and shellfish, Viterbo, Italy, p 6.

Becker BJ, Levin LA, Fodrie FJ, MacMillan PA (2007) Complex larval connectivity patterns among marine invertebrate populations. *Proceedings of the National Academy of Sciences USA* 104: 3267-3272.

Bergeron P, Bourget E (1986) Shore topography and spatial partitioning of crevice refuges by sessile epibenthos in an ice-disturbed environment. *Marine Ecology Progress Series* 28: 129-145.

Bohonak AJ (2002) IBD (Isolation by Distance): A program for analysis of isolation by distance. Journal of *Heredity* 93: 153-154.

Booth JD (1974) Observations on the hydrology of Bay of Islands, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 8: 671-689.

Bradford-Grieve JM, Murdoch RC, Chapman BE (1993) Composition of macrozooplankton assemblages associated with the formation and decay of pulses within an upwelling plume in greater Cook Strait, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 27: 1–22.

Brasher DJ, Ovenden JR, Booth JD, White RWG (1992) Genetic subdivision of Australian and New Zealand populations of *Jasus verreauxi* (Decapoda, Palinuridae) – preliminary evidence from the mitochondrial genome. *New Zealand Journal of Marine and Freshwater Research* 26: 53-58.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* 19: 233-257.

Chakraborty R (1989) Can molecular imprinting explain heterozygote deficiency and hybrid vigor? *Genetics* 96: 113-l 17.

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B* 126: 455-476.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9: 539.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science* 311: 522-527.

Dobretsov SV, Miron G (2001) Larval and post-larval vertical distribution of the mussel *Mytilus edulis* in the White Sea. *Marine Ecology Progress Series* 218: 179–187.

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

Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.

Foltz DW (1986) Null alleles as a possible cause of heterozygote deficiencies in the oyster *Crassostreas virginica* and other bivalves. *Evolution* 40: 869 – 870.

Gaffney PM (1990) Enzyme heterozygosity, growth rate, and viability in *Mytilus edulis*: another look. *Evolution* 44: 204-210.

Gardeström J, Pereyra RT, André C (2008) Characterization of six microsatellite loci in the Baltic Blue mussel *Mytilus trossulus* and cross-species amplification in North Sea *Mytilus edulis*. *Conservation Genetics* 9: 1003-1005.

Gardner JPA (1994) The structure and dynamics of naturally occurring hybrid *Mytilus edulis* Linnaeus, 1758 and *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia, Mollusca) populations: review and interpretation. *Archiv für Hydrobiologie, Monographische Beitrage* 99: 37-71.

Gardner JPA, Pande A, Eyles RF, Wear RG (1996) Biochemical genetic variation among populations of the greenshell mussel, *Perna canaliculus*, fromNew Zealand: preliminary findings. *Biochemical Systematics and Ecology* 24: 763–774.

Gardner JPA (2004) A historical perspective of the genus *Mytilus* (Bivalvia: Mollusca) in NZ: multivariate morphometric analysis of fossil, midden and contemporary blue mussels. *Biological Journal of the Linnaean Society* 82: 329-344.

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

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology* 84: 2989-2998.

Gosselin LA, Chia FS (1995) Characterizing temperate rocky shores from the perspective of an early juvenile snail: the main threats to survival of newly hatched *Nucella emaginata*. *Marine Biology* 122: 625-635.

Gosselin LA, Qian PY (1997) Juvenile mortality in benthic marine invertebrates. *Marine Ecology Progress Series* 146: 265–282.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86: 485–486.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.

Hamilton SL, Regetz J, Warner RR (2008) Postsettlement survival linked to larval life in a marine fish. *Proceedings of the National Academy of Sciences USA* 105: 1561-1566.

Hartl DL, Clark AG (1997) Principles of Population Genetics, 3rd edition. Sinauer Associates Inc., Sunderland, Massachusetts, p 542.

Heath RA (1982) A review of the physical oceanographyof the seas around New Zealand. *New Zealand Journal of Marine and Freshwater Research* 19: 79–124.

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

Helson J, Gardner JPA (2004) Contrasting patterns of mussel abundance at neighbouring sites: does recruitment limitation explain the absence of mussels on Cook Strait (New Zealand) shores? *Journal of Experimental Marine Biology and Ecology* 312: 285-298.

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

Hunt HL, Schielbing RE (1996) Physical and biological factors influencing mussel (*Mytilus trossulus*, *M. edulis*) settlement on a wave-exposed rocky shore. *Marine Ecology Progress Series* 142: 135-145.

Kimura M, Crow J (1964) The Number of Alleles that Can Be Maintained in a Finite Population. *Genetics* 49: 725–738.

Koehn RK (1991) The genetics and taxonomy of species in the genus *Mytilus*. *Aquaculture* 94: 125-145.

Koehn RK, Milkman R, and Mitton JB (1976) Population genetics of marine pelecypods. IV. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution* 30: 2–32.

Koehn RK, Bayne BL, Moore MN, Siebenaller JF (1980) Salinity related physiological and genetic differences between populations of *Mytilus edulis*. *Biological Journal of the Linnean Society* 14: 319-334.

Ladoukakis ED, Saavedra C, Magoulas A, Zouros E (2002) Mitochondrial DNA variationin a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Molecular Ecology* 11: 755-769.

Lane DJW, Beaumont AR, Hunter JR (1985) Byssus drifting and the drifting threads of the young post-larval mussel *Mytilus edulis*. *Marine Biology* 84: 301-308.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18: 189-197.

McDonald JH, Siebenaller JF (1989) Similar geographic variation at the LAP locus in the mussel *Mytilus trossulus* and *M. edulis*. *Evolution* 43: 228-231.

McDonald J, Seed R, Koehn R (1991) Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern hemispheres. *Marine Biology*, 111: 323-333.

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

Menge BA, Daley BA, Lubchenco J, Sanford E, Dahlhoff E, Halpin PM, Hudson G, Burnaford JL (1999) Top-down and bottom-up regulation of New Zealand rocky intertidal communities. *Ecological Monographs* 69: 297–330.

Menge BA, Lubchenko J, Bracken MES, Chan F, Foley MM, Freidenburg TL, Gaines SD, Hudson G, Krenz C, Leslie H, Menge DNL, Russell R, Webster MS (2003) Coastal oceanography sets the pace of rocky intertidal community dynamics. *Proceedings of the National Academy of Sciences USA* 100: 12229-12234.

Metaxas M (2001) Behaviour in flow: perspectives on the distribution and dispersion of meroplanktonic larvae in the water column. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 86-98.

Miller MP (2005) Alleles in Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity* 96: 722-724.

Monmonier MS (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* 5: 245–261.

Nei M, Tajima F, and Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19: 153–170.

Newell RIE, Hilbish TJ, Koehn RK, Newell CJ (1982) Temporal variation in the reproductive cycle of *Mytilus edulis* L. (Bivalvia, Mytilidae) from localities on the east coast of the United States. Biological Bulletin 162: 299-310.

Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2010) Community ecology package 'vegan' available from http://vegan.r-forge.r-project.org/.

Ovenden JR, Brasher DJ, White RWG (1992) Mitochondrial-DNA analyses of the red rock

lobster *Jasus edwardsii* supports an apparent absence of population subdivision throughout Australasia. *Marine Biology* 112: 319–326.

Paetkau D, Slade R, Burden M, Estoup A (2004) Direct, real-time estimation of migration rateusing assignment methods: a simulation-based exploration of accuracy and power. i 13: 55-65.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* 13: S146–S158.

Pawlik JR (1992) Chemical ecology of the settlement of benthic marine invertebrates. *Oceanography and Marine Biology: An Annual Review* 30: 233-335.

Pechenik JA, Eyster LS, Widdows J, Bayne BL (1990) The influence of food concentration and temperature on growth and morphological differentiation of blue mussel *Mytilus edulis* L. larvae. *Journal of Experimental Marine Biology and Ecology* 136: 47–64.

Pineda J (1994) Internal tidal bores in the nearshore: Warm-water fronts, seaward gravity currents and the onshore transport of neustonic larvae. *Journal of Marine Research* 52: 427-458.

Piry S, Alapetite A, Cornuet, J-M, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of *Heredity* 95: 536-539.

Presa P, Pérez M, Diz AP (2002) Polymorphic microsatellite markers for blue mussels (*Mytilus* spp.). *Conservation Genetics* 3: 441-443.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Quesada H, Beynon CM, Skibinski DOF (1995) A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* (Lmk.): Pleistocene vicariance biogeography and secondary intergradation. *Molecular Biology and Evolution* 12: 521-524.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* 94: 9197-9201.

Rawson PD, Hilbish TJ (1995) Distribution of male and female mtDNA lineages in populations of blue mussels, *Mytilus trossulus* and *M. galloprovincialis*, along the Pacific coast of North America. *Marine Biology* 124: 245-250.

Rawson PD, Agrawal V, Hilbish TJ (1999) Hybridization between the blue mussels *Mytilus galloprovincialis* and *M. trossulus* along the Pacific coast of North America: evidence for limited introgression. *Marine Biology* 134: 201-211.

Rawson PD, Joyner K, Hilbish TJ (1996) Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity* 77: 599-607.

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

Riginos C, Hickerson MJ, Henzler CM, Cunningham CW (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution* 58: 2438-2451.

Rodriguez SR, Ojeda FP, Inestrosa NC (1993) Settlement of benthic marine invertebrates. *Marine Ecology Progress Series* 97: 193–207.

Roemmich D, Sutton P (1998) The mean and variability of ocean circulation past northern New Zealand: determining the representativeness of hydrographic climatologies. *Journal of Geophysical Research* 103: 13041–13054.

Ross PM, Hogg ID, Pilditch CA, Lundquist CJ (2009) Phylogeography of New Zealand's coastal benthos. *New Zealand Journal of Marine and Freshwater Research* 43: 1009-1027.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics* 140: 1413–1419.

Sarver SK, Foltz DW (1993) Genetic population structure of a species complex of blue mussels. *Marine Biology* 117: 105 – 112.

Schneider S, Roessli D, and Excoffier L (2000) ARLEQUIN, Version 2.000. A software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries* 9: 363-377.

Shanks AL, Brink L (2005) Upwelling, downwelling, and cross-shelf transport of bivalve larvae: test of a hypothesis. *Marine Ecology Progress Series* 302: 1–12.

Sharples J, Greig MJN (1998) Tidal currents, mean flows and upwelling on the north-east shelf of New Zealand. *New Zealand Journal of Marine and Freshwater Research* 32: 215-231.

Shirtcliffe TGL, Moore MI, Cole AG, Viner AB, Badwin R, Chapman B (1990) Dynamics of the CapeFarewell upwelling plume, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 24: 555–568.

Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* 260: 83-96.

Siegel DA, Mitarai S, Costello CJ, Gaines SD, Kendall BE, Warner RR, Winters KB (2008) The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences USA* 105: 8974-8979.

Sigurdsson JB, Titman CW, Davies PA (1976) The dispersal of young post-larval bivalve molluscs by byssus threads. *Nature* 262: 386-387.

Skalamera JP, Renaud F, Raymond M, de Meeûs T. 1999. No evidence for differentiation in the mussel *Mytilus galloprovincialis* between lagoons and the seaside. *Marine Ecology Progress Series* 178: 251-258.

Skibinski DOF, Beardmore JA, Cross TF (1983) Aspects of the population genetics of *Mytilus* (Mytitidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society* 19: 137-183.

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

Smietanka B, Zbawicka M, Wolowicz M, Wenne R (2004) Mitochondrial DNA lineages in the European populations of mussels (*Mytilus* spp.). *Marine Biology* 146: 79-92.

Smietanka B, Burzynski A, Wenne R (2009) Molecular population genetics of male and female mitochondrial genomes in European mussels *Mytilus*. *Marine Biology* 156: 913-925.

Snelder T, Leathwick J, Weatherhead M, Wild M (2004) Pilot Marine Environment Classification of New Zealand's Exclusive Economic Zone and the Hauraki Gulf. NIWA Report. CHC2004-003. 56 p.

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology 1: Methodology. *Biological Journal of the Linnean Society* 10: 199-228.

Stanton BR (1976) Circulation and hydrology off the west coast of the South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 10: 445-467.

Stanton BR (1981) An oceanographic survey of the Tasman Front. *New Zealand Journal of Marine and Freshwater Research* 15: 289-297.

Star B, Apte S, Gardner JPA (2003) Genetic structuring among populations of the greenshell mussel *Perna canaliculus* revealed by analysis of randomly amplified polymorphic DNA. *Marine Ecology Progress Series* 249: 171–182.

Stevens MI, Hogg ID (2004) Population genetic structure of New Zealand's endemic corophiid amphipods: evidence for allopatric speciation. *Biological Journal of the Linnean Society* 81: 119–133.

Stevens G, McGlone M, McCulloch B (1995) PrehistoricNew Zealand. Auckland, Reed Publishing, 128 p.

Swearer SE, Shima JS, Hellberg ME, Thorrold SR, Jones GP, et al. (2002) Evidence of self-recruitment in demersal marine populations. *Bulletin of Marine Science* 70: 251–71.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389-399.

Tang J, Hanage WP, Fraser C, Corander J (2009) Identifying currents in the gene pool for bacterial populations using an integrative approach. PLoS Computational Biology 5: e1000455. doi:10.1371/journal.pcbi.1000455.

Thorson G (1950) Reproductive and larval ecology of marine bottom invertebrates. *Biological Review* 25: 1–45.

Todd CD (1998) Larval supply and recruitment of benthic invertebrates: Do larvae always disperse as much as we believe? *Hydrobiologia* 375/376: 1–21.

Toro JE, Castro GC, Ojeda JA, Vergara AM (2006) Allozymic variation and differentiation in the Chilean blue mussel, *Mytilus* chilensis, along it's natural distribution. *Genetics and Molecular Biology* 29: 174-179.

Vainola R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations. *Biological Journal of the Linnean Society* 43: 127-148.

Waters JM, Roy MS (2003) Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography* 30: 1787–1796.

Waters JM, Roy MS (2004) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology* 13: 2797–2806.

Westerbom M, Kilpi M, Mustonen O (2002) Blue mussels, *Mytilus edulis*, at the edge of the range: population structure, growth and biomass along a salinity gradient in the north-eastern Baltic Sea. *Marine Biology* 140: 991-999.

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

Young BL (1991) *Spartina* axil zones: preferred settlement sites of barnacles. *Journal of Experimental Marine Biology and Ecology* 151: 71-82.

Yu H, Li Q (2007) Development of EST-SSRs in the Mediterranean blue mussel, *Mytilus galloprovincialis*. *Molecular Ecology Notes* 7: 1308-1310.

Zouros E, Foltz DW (1984) Possible explanations of heterzygote deficieny in bivalve molluscs. *Malacologia* 25: 583-591.

Zouros E, Romero-Dorey M, Mallet AL (1988) Heterozygosity and growth in marine bivalves: further data and possible explanations. *Evolution* 42: 1332-1341.

Chapter 7: Conclusions

This thesis has covered a variety of spatiotemporal scales and aspects of blue mussel molecular ecology in the Southern hemisphere. The systematics of this cosmopolitan species is becoming better understood as more genetic data becomes available and is constantly changing with the influence of human-mediated dispersal. Our understanding of blue mussel molecular ecology from the (southern) hemispheric biogeographic scale to the fine-scale population differentiation in New Zealand has been so far behind similar interests for Northern hemisphere species, the main goal of this thesis is to bring our basic knowledge of Southern hemisphere blue mussel ecology a little closer to that of other *M. edulis* complex species.

7.1 A new member of the 'Mytilus edulis species complex': Mytilus meridianus

Sub-fossil midden blue mussel shells in New Zealand and Australia indicate this genus is native to Australasian regions and not all populations are recently introduced by humans. Previous investigations along with the current thesis support the endemicity of blue mussels in the Southern hemisphere, although their taxonomic status and evolutionary relationships with Northern hemisphere congeners remain ambiguous based on single gene histories and possibly biased genetic markers. Mitochondrial DNA variation has been used throughout this thesis to: (1) infer taxonomic status through an RFLP assay, (2) measure genetic distance between species, (3) infer evolutionary relationships among *M. edulis* complex species (4) uncover the biogeographic history of *Mytilus* in the Southern hemisphere and (5) evaluate the impact of known phylogeographical disjunctions on population subdivision of blue mussels within South America, New Zealand and Australia.

Measures of genetic distance alone are evidence for the unique character of Southern hemisphere mussels; on average, the amount of variation between hemispheres is

double that described among species in the Northern hemisphere, a property conserved between two mitochondrial genes. Variation contained within the 16s mtDNA gene has been used for phylogenetic reconstruction and forms the basis of the first RFLP assay differentiating between *M. galloprovincialis* hemispheric lineages. Based on available genetic evidence (outlined below) to date it is suggested that the Southern hemisphere lineage of *M. galloprovincialis* be named a new sibling species of the *M. edulis* species complex, *M. meridianus* (meaning 'South').

7.2 Reconstructing the evolutionary history of the Southern hemisphere blue mussels.

Resolving historical evolutionary relationships of blue mussels is obscured by allopatric speciation followed by secondary contact in the Northern hemisphere between North Atlantic and Mediterranean populations of *M. galloprovincialis* and *M. edulis*; extensive introgression between these species is evidenced by the inability of phylogenetic reconstruction to separate them into monophyletic clades. After *M. galloprovincialis* diverged in the Mediterranean Sea (estimated at ~2 mya), opportunity for dispersal to the North Atlantic was presented in a series of glacial cycles with shrinking and expanding coastal habitat availability. A phylogenetic approach to the taxonomic classification of Southern hemisphere smooth-shelled blue mussels theoretically removes bias associated with using molecular markers developed *a priori* for Northern hemisphere sibling species.

Phylogenetic attempts to place the Southern hemisphere lineage of *M. galloprovincialis* into the larger evolutionary framework of the remaining sibling species have resulted in three main hypotheses; (1) Southern hemisphere mussels diverged before the Northern hemisphere *M. edulis* and *M. galloprovincialis* (Gérard et al., 2008; Daguin & Borsa, 2000), (2) Southern hemisphere mussels form a monophyletic sister clade of the Mediterranean lineage of *M. galloprovincialis* that diverged after the North Atlantic *M.*

edulis/*M. galloprovincialis* haplogroup (Hilbish et al., 2000), and (3) Southern hemisphere mussels form a monophyletic sister clade of the North Atlantic shared *M. edulis* / *M. galloprovincialis* haplogroup that diverged after the Mediterranean *M. galloprovincialis* haplogroup (current investigation).

The evolutionary relationship of sister clades uncovered in this investigation indicates the Southern hemisphere lineage of *M. galloprovincialis* is derived from the common ancestor to the Atlantic haplogroup containing *M. galloprovincialis* and *M. edulis*. The order of divergence events intimates the Southern *M. galloprovincialis* lineage diverged after the Mediterranean *M. galloprovincialis* lineage. Thus, the Southern hemisphere lineage is most likely derived from Mediterranean lineage *M. galloprovincialis* following some secondary contact with North Atlantic *M. edulis* but before major introgressive hybridisation occurred to the extent that is witnessed at present day. Derived from coalescent based and not fossil calibrated phylogenetic analyses, the Southern *M. galloprovincialis* lineage is estimated to have diverged from the Northern hemisphere between 2.04 and 2.66 mya, indicating that *M. galloprovincialis* diverged in the Mediterranean Sea well before this time. This suggests the divergence time of *M. edulis* and Mediterranean *M. galloprovincialis* should be reviewed under new evidence that biotic interchange occurred across the Bering Strait earlier than once thought (4.8 − 7.3 mya, Marinkovich & Gladdenkov, 2006) and the Mediterranean Sea has had a permanent marine connection with the Atlantic Ocean since 5.3 mya (Loget and Van Den Driessche, 2006).

Cold water adaptation is used to enforce this hypothesis under the parsimony assumption that it is more likely an adaptation has evolved only once along an evolutionary pathway as opposed to more than once. Blue mussels in the Southern hemisphere are generally restricted to cold temperate waters at or higher than ~30° latitude, which mirrors the distribution of the North Atlantic *M. galloprovincialis* / *M.*

edulis haplogroup in the Northern hemisphere. It will be assumed that as *M. galloprovincialis* evolved in the Mediterranean this adaptation was lost. It is more parsimonious to assume that *M. galloprovincialis* from the warm Mediterranean waters gained cold-water adaptation through introgressive hybridisation with *M. edulis* rather than evolving it again on their own. The Southern lineage of *M. galloprovincialis* exhibits this adaptation through immediate shared ancestry with the North Atlantic haplogroup and not *de novo* through Mediterranean *M. galloprovincialis* haplogroup.

It is not surprising that *M. edulis* alleles/allelomorphs are found at variable frequencies across previously studied Southern hemisphere regions (McDonald & Koehn, 1988; McDonald et al., 1991), most notably the Kerguelen Islands and Tasmania (Daguin & Borsa, 2000; Borsa et al., 2007), but this is most likely due to the ancestral presence of these genes subject to patterns of drift and not secondary contact after colonisation. *M. edulis* alleles in the Southern hemisphere *M. galloprovincialis* genome supports the phylogenies presented in this investigation. The sister status of Southern *M. galloprovincialis* and *M. edulis*/*M. galloprovincialis* 'North Atlantic' haplogroups indicates shared ancestry of these taxa and the proposed ancestor of Southern *M. galloprovincialis* as being somewhat introgressed with *M. edulis* genes can explain the observed allelic profiles of geographically isolated Southern hemisphere locations under the influence of genetic drift since colonisation.

7.3 Differentiation within the Southern hemisphere

Within the Southern hemisphere, congruent reciprocally monophyletic clades in COI and COIII single gene topologies that are geographically restricted to South America (and Kerguelen Islands), New Zealand and Australia indicate intercontinental barriers to gene flow. The monophyletic property of these clades also infers that different lineages

within the Southern hemisphere are present. Across South America, New Zealand and Australia, the strongest barrier to gene flow is the Tasman Sea, a distance of ~ 2000 km. Trans-Tasman dispersal is not detected at the temporal scale intrinsic to the cytochrome oxidase complex genes, but connection of New Zealand and South American haplotypes under a 95% parsimony criterion indicates mussel populations on these continents experienced Trans-Pacific dispersal at one point that has since diminished causing these populations to fragment. Extensive ongoing gene flow between South America and the Kerguelen Islands (COI gene) indicates an eastward dispersal route most likely mediated by the West Wind Drift. The vast distance separating South America from the Kerguelen Islands (~5000 km.) is not a barrier to dispersal in this case, but geographically intermediate stepping stone populations on oceanic islands may mediate connectivity.

Irresolvable basal polytomies of South American haplotypes, topological incongruence among Bayesian phylogenetic reconstruction programs and overlapping confidence intervals of estimated divergence dates for deep phylogenetic tree nodes create uncertainty around the trans-equatorial pathway and continental order of Southern hemisphere colonisation. Past gene flow between South America and New Zealand may have also lead to their variable basal (clade) characteristic and erased the evolutionary signal of initial Southern hemisphere colonisation.

Genetic characteristics as contributing to barriers to gene flow among Southern hemisphere countries are explored through hybridisation dynamics and cytonuclear disequilibria among hybrid regions in these countries. Hybridisation is evident in regions of South America, New Zealand and Australia where Northern and Southern hemisphere *M. galloprovincialis* coexist, however, introgression among hybrids is not as widespread. Although characterizing the degree of introgression would become more accurate with additional nuclear loci, single nuclear and mitochondrial loci suggest introgression only

takes place in Australasia and is absent from Chile despite the long term occurrence of non-native mussels and first generation hybrids here. This barrier to gene flow could also account for at least some of the differentiation among geographically separated clades within the Southern hemisphere. Reciprocal monophyly of the South American clade and interlineage introgression barriers in Chile suggest mussels of this region constitute an evolutionarily significant unit (Crandall et al., 2001) within the Southern hemisphere lineage. New Zealand and Australia also represent regions of reciprocal monophyly within the larger Southern hemisphere clade. Of course, reciprocal monophyly of the Southern hemisphere *M. galloprovincialis* lineage and within the geographical regions of South America, New Zealand and Australia relies on the assumption that all Southern hemisphere sampled haplotypes resolving on Northern hemisphere clades are human mediated introductions supported by taxonomic identification from the 16s RFLP assay.

7.4 Non-native mussel presence

It is evident that Northern hemisphere *M. galloprovincialis* has been introduced into Southern hemisphere countries through anthropogenic activities, several investigations (including the current study) have identified mussels collected in New Zealand, Australia, South America and the Kerguelen Islands as having great taxonomic affinity with Northern hemisphere *M. galloprovincialis* and *M. edulis*. The additional knowledge of a unique Southern hemisphere lineage that is genetically differentiated from conspecifics in the Northern hemisphere and the successful introduction of *M. galloprovincialis* into other areas of the world (i.e. South Africa and the North-East Pacific Ocean) underlies the assessment of these rogue individuals as being recently introduced.

Furthermore, the development of two new molecular markers in the current investigation specifically targeted at global blue mussel mitochondrial and nuclear DNA

variation paves the way for new research into the magnitude and distribution of non-native mussels in the Southern hemisphere. Comparisons of native and non-native nuclear and mitochondrial allele frequencies gain insight into cytonuclear disequilibria and hybridisation dynamics in Southern hemisphere regions with non-native mussel presence.

7.5 Interlineage hybridisation and cytonuclear disequilibria

The frequency of non-native mussel nuclear and mitochondrial alleles is highest in Chile (34.5%) and New Zealand (42.7%), respectively. Hybridisation is detected in Chile, New Zealand and Australia with introgression only occurring in Australasia, where high frequencies of non-native alleles in New Zealand indicate a great magnitude or frequency of introduction events. A substantial presence of non-native alleles coupled with extensive introgression and a detected interlineage cross gender bias to Northern hemisphere maternal parents does not bode well for the perpetuation of this endemic lineage in New Zealand if introductions continue in a pattern similar to historical events that have contributed to current dynamics. Normalized disequilibria in the mid-range of marginal values indicate these taxonomic lines are already being blurred through introgression. With this in mind, it is apt to name and promote a new Southern hemisphere blue mussel species soon in order to prioritize conservation objectives and management efforts in New Zealand.

Although possible theories have been put forward in this investigation, further research is warranted to; (1) characterize degree of introgressive hybridisation with additional nuclear loci, (2) determine hybridisation dynamics at local scales with fine-scale, population level sampling and (3) determine source populations of already introduced mussels. The last point cannot be stressed enough, the vast number of potential Northern and Southern hemisphere (i.e. South Africa) source populations of introductions

could be, in part, responsible for differences in introgression dynamics between Chile and Australasia through genetic drift and/or selection.

7.6 Focus on New Zealand: population genetics from microsatellites

Genealogical reconstructions of COIII and COI sequences show a historical population subdivision within New Zealand mussels geographically restricted to the Bay of Islands. Microsatellite analysis, although not completely complementary in sampling locations, also indicates a Bay of Islands subdivision, with the exception of Waitangi clustering with the remaining national distribution (panmictic across the country). Within the Bay of Islands, Waitangi and TeTii Bay are genetically differentiated based on microsatellite allele frequencies and over 90% of individuals at each location cluster into separate populations whilst mtDNA indicates a single population in the Bay of Islands that is subdivided from the remaining national distribution. If TeTii Bay is genetically distinct from other locations by microsatellite allele frequency differences, this would constitute a contemporary differentiation. An historical differentiation is also detected with two mitochondrial loci of variable mutation rates. TeTii Bay and other Bay of Islands locations (Waitangi was not sampled so a direct comparison of historical vs. contemporary divergence is not possible) like Russell Harbour and Dove's Bay, cluster on the same geographically and genetically distinct clade, which is historically divergent from the remaining sampled locations as shown by deep coalescent events in the phylogenetic reconstruction. Russell Harbour is located just 3.5 km. from Waitangi and may be used as a proxy location for consideration of historical divergence within the Bay of Islands. Comparing the historical pattern of the Bay of Islands as a single population differentiated from remaining NZ locations to the contemporary pattern of differentiation within the Bay

of Islands, we can conclude that some shifts of factors affecting connectivity and dispersal have occurred in this region in the recent past.

Another likely contributor to contemporary (microsatellite) differentiation is the presence of non-native conspecifics and introgressive hybridisation at one of the sampled Bay of Islands locations. Although non-native individuals based on taxonomic classification from two individual genes were removed from analyses, introgression not identified by these methodologies could still exist. For example, the two genes used to determine taxonomic status could have ancestry in Southern lineage *M. galloprovincialis*, effectively displaying a native mussel status, whilst the gene regions associated with the microsatellite loci in the same individual could display non-native identity by descent.

There are two clues that undermine the hypothesis of artificial differentiation caused by non-native identity of loci. Firstly, if the product of non-native mussel presence and introgressive hybridisation were creating artificial population subdivision, one would still have to investigate why introgressive hybridisation dynamics are different at two geographically proximate locations. The number of individuals identified from these two target genes as non-native at TeTii Bay and Waitangi were 8 and 10, respectively, which aren't significantly different from eachother. To artificially create the observed population subdivision, Waitangi individuals would have had to undergo major introgressive hybridisation far beyond what has been detected by this study and TeTii Bay individuals would have had to undergone either introgression in a pattern uncomplimentary to Waitangi or no introgression at all. In the future, allele frequency arrays should be constructed for Northern hemisphere *M. galloprovincialis* free from the influence of Southern hemisphere conspecific introgressive hybridisation to determine if some loci can reliably infer hemisphere status and decipher introgression patterns.

The second clue undermining an hypothesis of introgressive hybridisation is as follows; TeTii Bay is part of the Bay of Islands genealogical clade that shows historical differentiation from the rest of New Zealand and those representative haplotypes are derived from several of the same individuals as microsatellite allele frequencies were typed from. If we can assume that TeTii Bay represents at least some of the genealogical differentiation present in the Bay of Islands, the majority of non-native introgressive hybridisation causing artificial population subdivision should be present at Waitangi. With non-native introgressive hybridisation changing the genetic identity of Waitangi individuals, one would expect this location to form an outlying population subdivision from microsatellite allele frequencies that is completely different from other sample locations, when in fact, Waitangi clusters with the remaining New Zealand sample locations inclusive of offshore islands. Replacing TeTii Bay for Waitangi as the location undergoing introgressive hybridiztion in this scenario would mean Whakariki Beach is also experiencing similar dynamics (with only five individuals identified as non-native).

These two clues undermine the hypothesis of artificial population subdivision being created by the presence of introgressive hybridisation. To explain the observed variation of microsatellite allele frequencies, almost any scenario invoking the presence of atypical alleles at one Bay of Islands location and not the other supports an alternate hypothesis of contemporary changes in factors affecting connectivity and dispersal of mussels in the Bay of Islands.

A promising part of the puzzle for contemporary change is the ephemeral current that shifts net water movement from the SE direction of the prevailing East Auckland Current to the NW. If larvae from TeTii Bay were entrained in this ephemeral current, they would have to cross the North Cape and travel southward to Whakariki Beach to maintain gene flow inferred by microsatellite allele frequencies. This dispersal route goes

against the prevailing current (from the Bay of Islands to the North Cape only) for the majority of the year and infers a specific temporal component to spawning regime.

Adhering to population subdivision identified from microsatellite allele frequencies, the Northwest population (Waitangi and Whakariki Beach) is relatively highly differentiated from the other panmictic population (Fst=0.17), indicating a moderate to strong barrier to gene flow.