

# A molecular and population genetics study of the Bluff Oyster (*Ostrea chilensis*)

Leighton James Thomas (BSc (Hons), MSc)

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

Victoria University of Wellington

*Te Whare Wānanga o te Ūpoko o te Ika a Māui*

2015

# Abstract

---

New Zealand has a long history of isolation and has evolved a unique biota. Spanning from a sub-tropical climate in the North to a sub-Antarctic climate in the far south New Zealand provides an interesting opportunity to study the processes that lead to evolution. This thesis attempts to study the evolution of *Ostrea chilensis* at a population genetics and molecular level.

## Chapter Two

Microsatellite DNA loci represent an ideal marker for population genetics studies due to high levels of length polymorphism between individuals. Genomic sequencing technologies offer the potential to quickly identify thousands of loci, from which PCR primers can be developed and screened for polymorphisms. I aimed to develop PCR primers to amplify length polymorphic microsatellite loci and to use the genomic data set to elucidate patterns and processes of microsatellite evolution. DNA was extracted from a single *Ostrea chilensis* individual and used for a 1/8 plate sequencing run on a Roche 454. The subsequent quality checked DNA database was annotated for microsatellite loci. 6,208 dinucleotide, 7,326 trinucleotide repeats, 2,414 tetranucleotide, 33 hexanucleotide and 356 pentanucleotides were annotated on the partial genome. Four microsatellite loci were successfully amplified and genotyped. The loci have a low number of alleles compared to other bivalve studies and two have significant departures from HWE ( $F_{st} = 0.126$  and  $-0.348$ ). There were a number of highly significant BLAST hits ( $< 1 \times 10^{-20}$ ) with repetitive Oyster DNA sequences obtained from GenBank. Due to difficulties the microsatellite loci were abandoned as markers for later population genetic analysis. This work, however, provides the ground work for further developments of PCR primers for polymorphic microsatellite DNA and provides some observations of molecular evolution of repetitive DNA, which will lead to a greater understanding of these sequences.

## Chapter Three

This chapter forms the first population genetic study of *Ostrea chilensis* using New Zealand and Chilean populations. The life history traits of *O. chilensis* are thought to reduce the dispersal of the species. Using randomly amplified polymorphic DNA (RAPD) I aimed to test the population genetic structure with the null hypothesis that there is panmixia (i.e. high

levels of gene flow and no barriers between populations). I then aimed to see if there was an isolation-by-distance (IBD) profile. Barriers to gene flow at around 42°S have been identified in a number of studies around New Zealand. I aimed to see if those barriers were present in this study. Significant spatial genetic differentiation was found among populations ( $F_{st} = 0.194$ ,  $p < 0.00001$ ). Over all spatial scales a significant IBD was not found, until ‘outlier’ (those with two standard deviations from the mean  $F_{st}$ ) were removed, then a slight IBE was found ( $r_{xy} = 0.324$ ,  $p = 0.030$ ,  $r^2 = 0.1052$ ). In an AWclust analysis two main clusters were revealed, but they did not correspond to above and below the 42°S. It is possible that the brooding life style of *Ostrea chilensis* has resulted in reduced gene flow between populations. Much of the genetic structure was not congruent with geographic location; this apparent chaotic patchiness could be influenced by human mediated movements and/or environmental variables.

## Chapter Four

The genetic structure found in the previous chapter was analysed in the context of near shore environmental and geo-spatial variables. My aim was to elucidate the environmental variables that best explain the apparent genetic structure of *O. chilensis*. Using genetic data from the previous chapter a Generalised Linear Model (GLM) was used to test the effect of ten environmental variables and two geospatial variables on average genetic distance ( $F_{st}$  and  $\Phi_{Ist}$ ). A complimentary BEST analysis was used to test the effect of the same variables on the individual alleles for each population. Using the BEST analysis the model that best explained the apparent genetic structure ( $R_s = 0.263$ ) included the environmental variables Sediment (SED) and mean sea surface temperature (SSTGrad). The GLM showed a more complicated model including most of the variables tested. Previous studies have shown that bivalve dispersal patterns are associated with sediment and spawning times can be influenced by temperature. This line of enquiry is important as it could lead to the identification of candidate genes for selection.

## Chapter Five

Genomic datasets contain a wealth of data that can be used to further understand genome structure and arrangements. My aim was to discover mitochondrial gene fragments in the genomic data set, use them to assemble the *Ostrea chilensis* mitogenome, and then analyse the genome in a phylogenetic framework with other oyster mitogenomes. A custom BLAST database was created using the 454 genomic DNA data set; this was then BLASTED against

all available oyster mitogenomes on GenBank. The DNA sequences with good statistical support ( $<1 \times 10^{-20}$ ) were mapped against the mitogenome of *Ostrea edulis*. The *O. edulis* mitogenome was then used to annotate the resulting mitogenome. I was able to recover 10,086 bp of mitochondrial DNA – this represents around 65% of the full genome. The subsequent topology of the Bayesian phylogenetic tree was similar to that found in previous studies. PCR primers have been designed to sequence the gaps in the mitogenome. This will allow full annotation. Full genome annotation will aid further research into genome evolution.

## Chapter Six

In the mitogenome, evolution is thought to mainly result in synonymous substitutions, due to functional constraints. I aimed to describe the selection pressures acting on each protein coding gene of the oyster mitogenome by comparing rates of non-synonymous (dN) and synonymous (dS) substitutions. I then aimed to discover if there was a codon bias in the oyster mitogenomes. Genetic distance was assessed using p-distance. The ratio of dN/dS was calculated using the method of Nei and Gojobori (1986). The overall ratio of dN/dS was  $<1$  for all protein coding genes, this would suggest that the genes are under purifying selection (non-synonymous mutations are selected against). A consistent codon bias was found across all protein coding genes, this could indicate translational selection.

This multidisciplinary approach aimed to explain the patterns and processes of evolution in *O. chilensis*. This thesis research developed molecular tools, and provided information that will aid fisheries and aquaculture management.

# Acknowledgements

---

Science is a collaborative discipline this PhD thesis has been made possible by the support and guidance of many people.

I thank my primary supervisor Professor Jonathan Gardner, firstly, for giving me the opportunity to come to New Zealand - which has been a life changing experience - and secondly, for not only letting me take the research in directions that interested me, but also allowing me to make my own mistakes and learn from them. I am also grateful to my secondary supervisor, Dr. Peter Ritchie for offering friendly advice and valuable feedback on the content of this thesis. Thanks too, for Dr. Geoff Chambers' discussions about molecular evolution over coffee. Thanks to the examining committee Dr. James Bell, Dr. Shane Lavery and Professor Pat Gaffney for pushing this thesis to the standard it should be.

The work was made possible by funding from the Cawthron Institute. Collection of specimens was made possible with the help from Keith Michael and Helen Mussely.

Much of the valuable learning I have made since the onset of the thesis has been through sharing of ideas and practices with other students. For this I would like to thank Kosala Sirisena, Luke Thomas, Phil Sirvid, Heather Constable, Catarina de Silva, Craig Marsh, Alix Laiffeire, Megan Shaffer and Cong Zeng.

Thanks to the cheerful ladies of the 5<sup>th</sup> floor Biology office: Patricia Stein, Mary Murray, Sandra Taylor and Leslie Taylor. I would also like to thank the staff of Victoria University Recreation Centre especially Abbie for talking me out running around the hills of Wellington, which kept me mostly sane.

I would also like to thank the Calman Family: David and Diane, for providing needed distractions and adventures that always involved good wine and food. And Riana Calman, who has been an invaluable source of inspiration, I nor the thesis would not be the same without her.

Finally, I would like to thank my parents Roy and Carolyn Thomas. Thank you to my Dad for taking me on weekly trips to the library – in many ways my path into science began with you. And thanks to my mum for inspiring a spirit of adventure which I try to channel into many areas of life. This work is yours.

# Table of Contents

---

Abstract .....	ii
Acknowledgements .....	vv
Table of Contents .....	vivi
List of Figures .....	x
List of Tables .....	xii
List of Appendices .....	xiv
Declaration .....	
Chapter One: An Introduction to the study species: <i>Ostrea chilensis</i> .....	1
1.0 Aims .....	1
1.1 Introduction .....	1
1.1.2 Oyster beds and reefs, their ecosystem services and anthropogenic threats world wide .....	4
1.1.3 The Biogeography of New Zealand.....	5
1.1.4 An overview of marine population genetics and some key studies in New Zealand	6
1.1.5 Fishery of <i>Ostrea chilensis</i> in New Zealand and the threat of bonamiasis .....	7
1.1.6 Population genetics of <i>Ostrea edulis</i> , a closely related species with similar problems .....	10
1.1.8 Organisation of the thesis .....	13
1.2 References .....	15
Chapter Two: A study of the microsatellite loci of the Bluff Oyster ( <i>Ostrea chilensis</i> ).....	21
2.0 Aims .....	21
2.1 Introduction .....	21
2.1.1 Applications.....	23
2.1.2 Hypotheses.....	24
2.2 Methods .....	25
2.2.1 DNA extraction and 454 Pyrosequencing .....	25
2.2.2 Identification of microsatellite loci.....	25
2.2.3 Screening of reads for microsatellite sequences and development of PCR primers	26
2.2.4 PCR amplification .....	26
2.2.5 Genotyping .....	27
2.2.6 Summary Statistics .....	27

2.2.7 Search for orthologous repetitive loci .....	28
2.3 Results .....	28
2.3.1 Success of the 454 sequencing .....	28
2.3.2 Identification of microsatellite loci and a description of the length of array, motif and contribution to the genome for each repeat type.....	32
2.3.3 Identification of polymorphic microsatellite loci .....	33
2.3.4 Search for orthologous repetitive DNA in various oyster species.....	34
2. 4 Discussion .....	48
2. 5 Conclusion .....	52
2.6 References .....	53
Chapter Three: A Population Genetics Analysis of <i>Ostrea chilensis</i> Using Randomly Amplified Polymorphic DNA.....	66
3.0 Aims .....	66
3.1 Introduction .....	66
3.1.1 Spatial genetics in the marine environment.....	66
3.1.2 Previous work on the study species .....	69
3.1.3 The Oyster Fishery .....	70
3.1.4 Genetic Markers .....	70
3.1.5 Hypothesis .....	71
3.2 Methods.....	72
3.2.1 Sample collection .....	72
3.2.2 DNA Extraction.....	74
3.2.3 Primer Screening .....	75
3.2.4 PCR protocol .....	75
3.2.5 Analysis of RAPD-PCR products.....	75
3.3 Data Analysis.....	76
3.4 Results .....	80
3.4.1 Genetic diversity .....	80
3.5 Discussion .....	97
3.6 Summary.....	101
3.7 References .....	102
3.8 Appendix .....	108
Chapter Four: A Seascape Genetics Analysis of New Zealand Populations of <i>Ostrea chilensis</i> .....	116

4.0 Aims .....	116
4.1 Introduction .....	116
4.1.1 Recent studies in New Zealand .....	118
4.1.2 Hypothesis .....	119
4.2 Methods .....	120
4.2.1 Genetic data .....	120
4.2.2 Environmental and Geospatial data .....	120
4.2.3 Test of independence among variables .....	120
4.2.4 BEST (Biological Environmental Stepwise) .....	120
4.2.5 Generalised linear model .....	121
4.3 Results .....	121
4.3.1 BEST analysis .....	121
4.3.2 GLZ analysis .....	126
4.4 Discussion .....	131
4.5 Conclusion .....	134
4.6 References .....	135
4.7 Appendix .....	138
Chapter Five: Partial Assembly of the <i>Ostrea chilensis</i> Mitogenome and its Subsequent use in Phylogenetic Reconstruction .....	140
5.0 Aims .....	140
5.1 Introduction .....	140
5.1.1 Mitochondrial DNA .....	140
5.1.2 Oyster mitogenomes and phylogenetics .....	141
5.1.3 Hypotheses .....	144
5.2 Methods .....	145
5.2.1 Identification of mtDNA sequences .....	145
5.2.2 Phylogenetic reconstruction using oyster mitogenomes .....	145
5.2.3 Phylogeographic analysis of <i>Ostrea chilensis</i> Cytochrome Oxidase I sequences .....	146
5.3 Results .....	147
5.3.1 Mitochondrial Genome Construction .....	147
5.3.2 Best-fit models of gene evolution for oyster mitogenomes .....	147
5.3.3 Phylogenetic analysis based on all oyster mitogenomes .....	148
5.3.4 Phylogeographic analysis .....	148
5.4 Discussion .....	161

5.5 Conclusion.....	163
5.6 References .....	164
Chapter 6: Molecular Evolution in Oyster Mitochondrial Genes .....	168
6.0 Aims .....	168
6.1 Introduction .....	168
6.1.1 Hypotheses.....	170
6.2 Methods .....	171
6.2.1 Assessment of codon bias.....	171
6.2.2 Comparison of genetic distances .....	171
6.3 Results .....	172
6.3.1 Codon usage bias .....	172
6.3.2 Genetic distance.....	172
6.3.3 Test of Neutrality - dN/dS ratios of mitochondrial genes between <i>Ostrea</i> and <i>Crassostrea</i> spp. ....	173
6.4 Discussion .....	179
6.5 Conclusion.....	181
6.6 References .....	182
6.7 Appendix .....	184
Chapter Seven: General Discussion.....	193
7.0 Aims .....	193
7.1 Summary of Key Findings .....	193
7.2 Context .....	195
7.3 Limitations common to population genetic studies.....	197
7.4 Recommendations for further study .....	199
7.5 Implications .....	202
7.6 Conclusion.....	204
7.7 References .....	205

# List of Figures

---

<b>Figure 1.1 :</b> The New Zealand marine environment.....	2
<b>Figure 2.1:</b> Screen shot from GENEIOUS, length distribution of reads after before clipping .....	30
<b>Figure 2.2:</b> Screen shot from GENEIOUS, length distribution of reads after clipping .....	31
<b>Figure 2.3:</b> Screen shots of selected FastQC results. ....	32
<b>Figure 2.4a and Figure 2.4b</b> Electrophoresis gels .....	35
<b>Figure 2.4c and Figure 2.4d</b> Electrophoresis gels .....	36
<b>Figure 2.4e and Figure 2.4f</b> Electrophoresis gels .....	37
<b>Figure 2.4g and Figure 2.4h</b> Electrophoresis gels .....	38
<b>Figure 2.4i and Figure 2.4j</b> Electrophoresis gels.....	39
<b>Figure 2.5:</b> Contribution of each repeat type to genomic DNA.....	42
<b>Figure 2.6:</b> Frequency of repeat motif of each microsatellite type .....	43
<b>Figure 2.7</b> Length distributions of the microsatellite arrays .....	44
<b>Figure 2.8</b> Alignment of an orthologous microsatellite repeat shared between <i>Ostrea chilensis</i> and <i>Crassostrea virginica</i> . ....	47
<b>Figure 3.1</b> Sampling locations around New Zealand.....	73
<b>Figure 3.2</b> Location of Chilean samples relative to New Zealand.....	74
<b>Figures 3.3a to 3.3d</b> Mantel tests of Geographic distances vs genetic distances .....	88
<b>Figures 3.3e to 3.3h</b> Mantel tests of Geographic distances vs genetic distances .....	889
<b>Figure 3.4</b> Global data set. Nei's genetic distance Principle Co-ordinate analysis .....	900
<b>Figure 3.5a</b> Number of populations assigned by the AWclust analysis using global data set	93
<b>Figure 3.5b</b> Number of populations assigned by the AWclust analysis when outlier loci are removed.....	94
<b>Figure 3.6</b> Frequency band presence of outlier loci by population.....	95
<b>Figure 5.1</b> The tree on the left was produced using 19 protein coding genes extracted from the full mitogenomes. The tree on the right was produced using COI fragments from all published <i>Ostrea</i> , <i>Ostreola</i> , <i>Saccostrea</i> and <i>Crassostrea</i> . Support values are the result of 100 Bootstraps .....	143
<b>Figure 5.2</b> Total number (bp) of mitochondrial DNA sequences recovered compared to total genomic sequences from various next generation sequence datasets, adapted from Leese et al. (2012). Amount of mitochondrial DNA recovered for <i>Ostrea chilensis</i> is highlighted in red .....	150
<b>Figure 5.3</b> partial draft mitochondrial genome of <i>Ostrea chilensis</i> .....	154
<b>Figure 5.4</b> Trace plot of the MCMC chain for the mitogenome phylogenetic reconstruction .....	157
<b>Figure 5.5</b> Trace plot of the MCMC chain for phylogeography analysis .....	158
<b>Figure 5.6</b> Phylogenetic tree of produced using Bayesian methods. Branch supports are Bayesian posterior probabilities. Out group is <i>Argopecten irradians</i> .....	160
<b>Figure 5.7</b> Phylogenetic tree based on Bayesian analysis using partial cytochrome oxidase 1 gene (COI) alignments for <i>Ostrea chilensis</i> . <i>Ostrea stentina</i> and <i>Ostrea angasi</i> are used as outgroups.....	160
<b>Figure 6.1a (left)</b> Base composition of all CO1 genes compared for <i>Ostrea</i> and <i>Crassostrea</i> species. <b>Figure 6.1b (right)</b> pairwise comparison of all CO1 genes for all oyster species. Y-	

axis shows p-distance for either transitions or transversions the X axis shows the overall p-distance. ....	174
---	-----

# List of Tables

---

<b>Table 2.1:</b> PCR Protocol adapted from Korbie and Mattick (2008) .....	27
<b>Table 2.2:</b> Problems and warning suggested by the program FastQC for the clipped 454 sequencing data set .....	29
<b>Table 2.3:</b> Primers that produced a clear band in the expected size region .....	40
<b>Table 2.4</b> Characterisation of microsatellite loci .....	45
<b>Table 2.5a</b> BLAST hits with published <i>Ostrea edulis</i> repetitive DNA on GenBank .....	45
<b>Table 2.5b</b> BLAST hits with published <i>Ostrea conchaphila</i> repetitive DNA on GenBank .....	46
<b>Table 2.5c</b> BLAST hits with published <i>Crassostrea gigas</i> repetitive DNA on GenBank .....	46
<b>Table 2.5d</b> BLAST hits with published <i>Crassostrea virginica</i> repetitive DNA on GenBank .....	46
<b>Table 2.5e</b> BLAST hits with published <i>Crassostrea hongkongensis</i> repetitive DNA on GenBank .....	47
<b>Table 3.1</b> Sampling size and localities .....	72
<b>Table 3.2</b> Summary of randomly amplified polymorphic DNA marker (RAPD) loci and Nei's gene diversity for each sampling site.....	83
<b>Table 3.3a</b> Global data set. $\Phi_{PT}$ values below diagonal based on 9999 permutations is shown below the diagonal. Number of migrants ( $N_m$ ) is shown above the diagonal. $\Phi_{PT}$ values given in bold were not statistically significant after false discovery correction for multiple testing.....	84
<b>Table 3.3b</b> Outlier loci removed. $\Phi_{PT}$ values below diagonal based on 9999 permutations is shown below the diagonal. Number of migrants ( $N_m$ ) is shown above the diagonal. $\Phi_{PT}$ values given in bold were not statistically significant after False discovery correction for multiple testing.....	85
<b>Table 3.4</b> Pairwise $F_{st}$ between sampling sites using global data set below the diagonal and with outlier loci removed above the diagonal.....	86
<b>Table 3.5</b> AMOVA using binary distance matrix using 4 regions (northern, eastern, southern and Chilean).....	87
<b>Table 3.6a</b> Global data set. Assignment test results for individuals from each population. Values are percentage of from each collection site .....	91
<b>Table 3.6b</b> Outlier loci removed. Assignment test results for individuals from each population. Values are percentage of from each collection site.. .....	92
<b>Table 3.10</b> Percentage of individuals assigned to each cluster. ....	94
<b>Table 3.11</b> Assignment to the most parsimonious cluster of $k=3$ produced without the outlier loci included.....	95
<b>Table 4.1a</b> 10 best fitting models the best analysis using the global data set .....	123
<b>Table 4.1b</b> 10 best fitting models of the BEST analysis with outlier loci removed .....	124
<b>Table 4.1c:</b> 10 best fitting models using the BEST analysis two random loci removed.....	124
<b>Table 4.2a</b> Results of the GLZ analysis using the global data set using $F_{st}$ .....	127
<b>Table 4.2b</b> Results of the GLZ analysis using the outlier removed data set using $F_{st}$ .....	128
<b>Table 4.3a</b> Results of the GLZ analysis using $\Phi_{st}$ and the global data set.....	129
<b>Table 4.3b</b> Results of the GLZ analysis using $\Phi_{st}$ and the global data set using phist .....	130

<b>Table 5.1</b> Mitogenomes used in the study and the number of hits when each genome was blasted against the 454 database .....	149
<b>Table 5.2</b> partial COX1 sequences used in the phylogeographic study .....	151
<b>Table 5.3</b> Summary Statistics for mitochondrial genome coverage.....	152
<b>Table 5.4</b> PCR primers designed to amplify DNA in the gap regions found in the mitochondrial genome .....	155
<b>Table 5.5</b> Comparison of best fit models for each of the protein coding genes.....	156
<b>Table 6.1</b> Summary statistics for the CO1 gene compared between all oysters .....	175
<b>Table 6.1b</b> Comparison of p-distance for the CO1 gene compared between different codon positions and at different taxonomic levels .....	175
<b>Table 6.2</b> Relative synonymous codon usage (RSCU) for the COI gene in two oyster genera and <i>D. melangastor</i> .....	176
<b>Table 6.1:</b> Codon usage bias for the cytochrome oxidase gene.....	177
<b>Table 6.4</b> Comparison of the codon bias for each protein-coding gene from the mitochondrial genome of four <i>Ostrea</i> species. ....	178

# List of Appendices

<b>Appendix 3.1</b> RAPD primers tested those in bold were kept for further analysis .....	108
<b>Appendix 3.2</b> Allele frequency estimates using the Bayesian method of Zhivotovsky (1999). 'a' represents the 'dominant' allele (presence of a band) and 'b' represents the 'recessive' allele (absence of a band).....	110
<b>Appendix 3.3</b> Individual locus $F_{st}$ . Loci highlighted in bold are outlier loci.....	113
<b>Appendix 3.4</b> Pairwise $\Phi_{PT}$ values for randomised individuals in four popualtions based on 9999 permutations is shown below the diagonal .....	114
<b>Appendix 3.5</b> AMOVA table for the four random populations .....	114
<b>Appendix 3.6</b> Pairwise $\Phi_{PT}$ values for randomised individuals in six popualtions based on 9999 permutations is shown below the diagonal.....	114
<b>Appendix 3.7</b> AMOVA table for the six random populations.....	114
<b>Appendix 3.8</b> AMOVA table for the eight random populations.....	115
<b>Appendix 3.9</b> Pairwise $\Phi_{PT}$ values for randomised individuals in eight popualtions based on 9999 permutations is shown below the diagonal .....	115
<b>Appendix 4.1</b> . Environmental variables from the Marine Environment Classification (MEC) .....	138
<b>Appendix 6.1</b> Summary statistics for the ATP6 mitochondrial gene .....	184
<b>Appendix 6.2</b> Summary statistics for the COI mitochondrial gene .....	185
<b>Appendix 6.3</b> Summary statistics for the COII mitochondrial gene.....	186
<b>Appendix 6.4</b> Summary statistics for the COIII mitochondrial gene .....	187
<b>Appendix 6.5</b> Summary statistics for the CYTB mitochondrial gene .....	188
<b>Appendix 6.6</b> Summary statistics for the NAD1 mitochondrial gene .....	189
<b>Appendix 6.7</b> Summary statistics for the NAD2 mitochondrial gene .....	190
<b>Appendix 6.8</b> Summary statistics for the NAD4 mitochondrial gene .....	191
<b>Appendix 6.9</b> Summary statistics for the ATP6 mitochondrial gene .....	192



# Declaration

---

I certify that this intellectual work is my own with logistical and theoretical support from my two supervisors, Jonathan Gardner and Peter Ritchie. The section below details the contributions to the research chapters in this thesis.

## Chapter Two

Study design: Author

Data collection: Author

Analyses: Author

Writing and editing: Author with review and edits from Jonathan Gardner and Peter Ritchie.

## Chapter Three

Study design: Author

Data collection: Author

Analyses: Author

Writing and editing: Author with review and edits from Jonathan Gardner and Peter Ritchie.

## Chapter Four

Study design: Author

Data collection: Author

Analyses: Author

Writing and editing: Author with review and edits from Jonathan Gardner and Peter Ritchie.

## Chapter Five

Study design: Author

Data collection: Author

Analyses: Author

Writing and editing: Author with review and edits from Jonathan Gardner and Peter Ritchie.

## Chapter Six

Study design: Author

Data collection: Author

Analyses: Author

Writing and editing: Author with review and edits from Jonathan Gardner and Peter Ritchie

# An Introduction to the study species: *Ostrea chilensis*

---

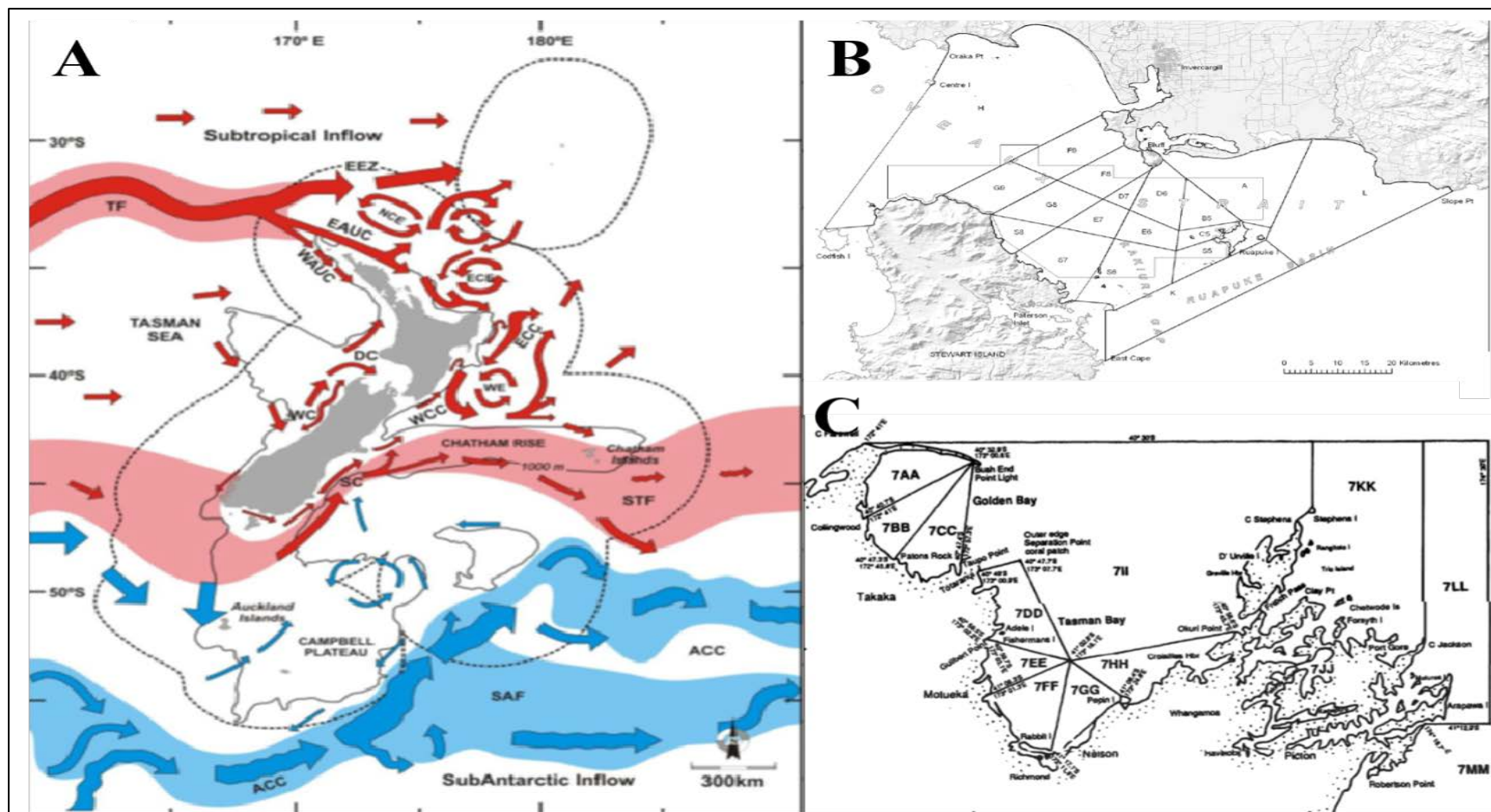
## 1.0 Aims

The main objective of this thesis research was to gain an insight into the genetic structure around New Zealand and Chile of the commercially important oyster species *Ostrea chilensis*. This chapter describes the biology of the species, the biogeography of New Zealand, the problems that are particular to oyster fisheries in New Zealand, and the general problems that are faced by oysters around the World.

## 1.1 Introduction

*Ostrea chilensis* is a commercially and culturally important marine bivalve mollusc of the family Ostreidae. It is known variously as: the flat oyster, the mud oyster, the dredge oyster, and the Bluff oyster among other names. The species is commercially fished in two main locations in New Zealand: the Foveaux Strait and the Marlborough sounds (Figure 1.1).

*Ostrea chilensis* is a protandrous hermaphrodite that brood its larvae for an incubation period lasting 2 to 9 weeks (Millar and Hollis, 1963; Toro and Morande, 1998). The fully developed larvae are able to settle within minutes of release, although a portion of them may remain in the plankton for up to 20 days (Cranfield 1968; Cranfield *et al.* 1968; DiSalvo *et al.* 1983; Cranfield and Michael, 1989). *Ostrea chilensis* larvae from incubating females were shown to be 0.45 – 0.49 mm long and to possess eye spots and a foot that other species develop only at the end of the pelagic phase (Millar and Hollis, 1963). An average of 70,000 eyed larvae can be expected per brooding female (Winter *et al.* 1984b). This is a much lower fecundity in comparison with other *Ostrea* species: the closely related flat oyster *Ostrea edulis* spawns as many as 1-2 million eggs per brood (Cole 1941; Walne, 1964). *Ostrea edulis* larvae are released from the female when they are 0.15 – 0.17 mm long and then undergo pelagic development, usually to a size of 6-14 mm that is influenced by water temperature and food supply, before settling and completing metamorphosis. The extended brooding strategy of *O. chilensis* is thought to increase the viability of embryos by reducing environmental stress; however, the process is associated with reduced fecundity in comparison to species whose



**Figure 1.1:** The New Zealand marine environment, showing the location of notable landforms, major current patterns and water masses, bathymetry and the New Zealand exclusive economic zone (EEZ). Abbreviations used for currents and water masses are as follows: Tasman Front (TF), East Auckland Current (EAUC), West Auckland Current (WAUC), North Cape Eddy (NCE), East Cape Eddy (ECE), East Coast Current (ECC), Wairarapa Eddy (WE), D'Urville Current (DC), Wairarapa Coastal Current (WCC), Westland Current (WC), Southland Current (SC), Subtropical Front (STF), Sub-Antarctic Front (SAF), Antarctic Circumpolar Current (ACC). Redrawn after Laing & Chiswell (2003) **B:** Dredge oyster management area (OYS5) – Foveaux Strait. **C:** Dredge oyster management area (OYS7) – Nelson/Marlborough Sounds.

gametes are dispersed into the environment. This brooding strategy is also thought to prevent long-distance larval dispersal (Toro and Chaparro, 1990). *O. chilensis* is, therefore, an interesting study species for which the effects of its life history can be compared and contrasted with other oyster species.

*Ostrea chilensis* are found in New Zealand and Chile. The species is distributed throughout the near shore waters of New Zealand, including the Chatham Islands, to depths of 550 m (Jeffs and Creese, 1996). Some local populations have been known to become functionally extinct (for example Manukau harbour, personal communication with Professor Andrew Jeffs). The precise distribution of populations around New Zealand has not been quantified. The species is also found in the shallow sub-tidal zone of Southern Chile (Buroker *et al.* 1983). The distribution for the Chilean populations is restricted to Chiloe Island and the Southern islands such as Las Guaitecas (Winter *et al.* 1984). The fossil record suggests that New Zealand is the centre of origin for the species (Beu and Maxwell, 1990). O’Foighil *et al.* (1999) used molecular and fossil data to conclude that the most likely explanation for their contemporary distribution is dispersal by rafting from New Zealand to Chile, and that the North Island is the source of the Chilean population.

Jozefowicz and O’Foighil (1998) investigated the phylogenetic relationship of the flat oyster family Ostreidae based on partial mitochondrial 16S rDNA gene sequences. *Ostrea chilensis* is closely related to the geographically disparate taxa: the European Flat Oyster *O. edulis* and the Australian Flat Oyster *O. angasi*. However, in that study, the position of *O. chilensis* was only weakly supported. In a more recent study (O’Foighil and Taylor, 2000), place *O. chilensis* in the same position, more oyster species were used to construct the phylogeny and they used a different gene region: a partial 28S sequence. It is likely the position of *O. chilensis* would be better resolved using more gene regions and a greater range of oyster species that were not available at the time of previous studies.

Within the genus *Ostrea*, brooding and settlement cycles are thought to be primarily influenced by changes in water temperature (Westerskov, 1980; Seale and Zacherl, 2009). In particular, cold seasonal water temperatures are believed to play a key role in synchronising larval production and settlement (Jeffs *et al.* 1996, 1997b), as they do in most species. Populations of *O. chilensis* from different geographical locations exhibit wide variation in temporal patterns of brooding and larval release. This is associated primarily with seasonal changes in water temperatures (Cranfield, 1968; Westerskov, 1980; DiSalvo *et al.* 1983; Jeffs

*et al.* 1996, 1997a). In New Zealand, the brooding period has been reported to range from a minimum of 16 days at temperatures of 18-20°C (Hollis, 1963), to 38 days at temperatures of 13-14°C (Westerskov, 1980), while studies in Chilean waters (that average around 13°C) have reported a brooding period between 40 and 62 days (Toro and Morande, 1998). In the warmer waters of the Manukau Harbour and Hauraki Gulf in Northern New Zealand, larval brooding and settlement among *O. chilensis* populations continues throughout the year (Jeffs *et al.* 1996; Jeffs *et al.* 1997b), whereas Southern New Zealand populations show a sharply defined and shorter breeding season during the summer months (Cranfield 1968; Westerskov 1980; Jeffs and Hickman 2000).

Oyster larvae use a type of cement for attachment (a highly organised complex of micro-fibres), squeezing this from the byssal gland and then applying the left shell valve to the cement. Once settled, oysters are not capable of detaching. The attached larva undergoes metamorphosis, from a pelagic to a sessile existence, during this time there is complete loss of the foot. *O. chilensis* larvae will not settle on fine-grained sediment (Baker, 1995; Tamburri *et al.* 2008). They can however settle when the seafloor habitat has been stabilised by other species of macrofauna, and settlement is greatly increased in areas of complex biogenic reef (Cranfield *et al.* 2004).

### **1.1.2 Oyster beds and reefs, their ecosystem services and anthropogenic threats world wide**

Globally oyster beds are one of the most threatened marine habitats on earth; they have declined >90% from historic levels in 70% of the bays and marine ecoregions, in 37% of estuaries and 28% of ecoregions they are functionally extinct (Thush and Dayton, 2002). Functional extinction refers to the situation where a species becomes so rare they do not fulfil the ecosystem roles that have evolved in the system (Thush and Dayton, 2002). Of particular concern, given the global threats to oysters, is that biogenic reefs - such as those created by oysters - are recognised as having important ecosystem functions. Among these many functions are: the provision of habitat structure that can provide refuge from predation for juvenile fish (Diaz *et al.* 2003), the creation of feeding habitats for juveniles and adults of mobile species, attachment sites for species with a sessile stage, and also provision of nesting habitats (Coen *et al.* 2007). In New Zealand, accumulation of mollusc shell debris has been shown to increase biodiversity when compared with similar environmental conditions that lack the shell debris (Hewitt *et al.* 2005). Further ecosystem services include the removal of

suspended solids from surrounding waters and subsequently increasing water clarity. This process reduces turbidity and allows light to reach down to further depths enabling sea grass and other macro-algae to grow (Newell and Koch, 2004). This same filtration process when carried out by the hard clam (*Mercenaria mercenaria*) has been shown to reduce the incidence of harmful algal blooms of the brown tide algae (*Aureococcus anophagefferens*) (Cerrato, 2004). Finally, shellfish can be used to absorb wave energy, protecting shorelines and reducing erosion. Therefore, restoration of Oyster beds (*Crassostrea virginica*) has the potential for use in coastal defence strategies (Piazza *et al.* 2005).

Around the world numerous anthropogenic processes have led to a decline of oyster populations and the associated biogenic reefs. In Chesapeake Bay, USA *Crassostrea virginica* numbers have massively declined due to overfishing and mechanical destruction of the reef by bottom trawling (Rothschild *et al.* 1994). These fishing methods destroy the biogenic reef and lead to functional extinction of benthic habitats (Thrush and Dayton, 2004). In Tasman Bay, New Zealand the oyster (*Ostrea chilensis*) fishery has declined 40-fold in the last 15 years (New Zealand Ministry of Fisheries, 2009). The sea bed here is largely dominated by mud, sand and some shell hash (Broekhuizen *et al.* 2011), it has been suggested that the original levels of productivity of this fishery are unable to be recovered as there is a lack of hard substrate for oyster larvae to attach. Work is being done to provide additional hard substrate onto which the larval oysters may settle (Powell and Klinck, 2007).

### **1.1.3 The Biogeography of New Zealand**

New Zealand consists of three main islands (North Island, South Island and Stewart Island), a few hundred nearby islands, including Great Barrier, D'Urville and Resolution, and several offshore groups (Three Kings, Kermadecs, Chathams, Bounty, Antipodes, Campbell, Auckland, Snares) (Wallis and Trewick, 2009). This archipelago of more than 700 islands, spans from the subtropical north (Kermadec Islands, 29°S) to the sub-Antarctic south (Campbell Island, 52°S) (Ross *et al.* 2009).

New Zealand has a long history of isolation, Zealandia the land mass from which New Zealand is derived, separated from what are now Australia, Antarctica and South America at around 82 MYA (Wallis and Trewick, 2009). There has been much debate about the origin of New Zealand taxa. Current thought is that only a very small percentage of taxa are the result of this Gondwanan vicariance, more contemporary geological processes and dispersal are

thought to be the main factors involved in the speciation of New Zealand taxa (Wallis and Trewick, 2009).

New Zealand has high levels of endemism, an estimated 65,000 known and un-described marine species with around 44% endemism (MacDiarmid, 2007). The high level of endemism and geographical isolation provides an interesting and unique opportunity to study evolution and the micro-evolutionary events at the scale of populations. One aim of these biogeographic questions is to identify barriers to dispersal and describe their effects on the geographical distributions of species. New Zealand has a complex oceanography resulting from the interaction of subtropical and sub-Antarctic water masses, and deep water and surface currents (Figure 1.1). Currents have a strong influence on the connectivity among coastal species (White *et al.* 2010).

#### **1.1.4 An overview of marine population genetics and some key studies in New Zealand**

Geographical variation in genetic structuring results from an interaction of processes: mutation and genetic drift due to finite population size, and natural selection tend to create genetic heterogeneity; while gene flow between populations as a result of the movement of individuals, larvae or other stages is a force that will tend to create homogeneity in genetic structuring (Slatkin, 1987).

Historically, due to the continuity of the sea as a medium for transport, the lack of obvious geographical barriers, the large population sizes and a pelagic larval stage in many marine taxa, it was assumed that populations within the marine environment would have a panmictic structure. It has since been shown that some marine taxa have much greater levels of genetic differentiation than expected, and there are much higher levels of speciation than would be expected in these supposed sympatric populations (Cowen *et al.* 2006). Long distance movements may be impeded by the existence of hydrological processes such as oceanic currents or ecological barriers such as habitat patchiness (Rocha-Olivares and Vetter, 1999). In addition, biological factors such as non-passive larval movements may increase local settlement (Taylor and Hellberg, 2003) and local selection may lead to genetic structuring (Zardi *et al.* 2011).

Several studies of genetic structuring in marine invertebrates generally concluded that population subdivision is higher in species with direct development than those with planktonic development (e.g. Ayre and Hughes, 2000; Collin, 2001). *Ostrea chilensis* has

been found to settle almost immediately (Cranfield and Michael, 1989) this is thought to reduce its ability to disperse and hence increase levels of genetic structuring (Toro and Aguilla, 1996). Ross *et al.* (2009), in a review of the phylogeography of New Zealand's coastal benthos, a significant negative relationship between planktonic larval duration (PLD) and population subdivision ( $F_{st}$ ) was revealed. However, this is not an absolute rule in the marine environment. The stomapod *Haptosquilla pulchella*, which has a high estimated potential dispersal distance, exhibits significant genetic structure between oceanographic regions in Indonesia (Barber *et al.* 2000, 2002). In the same oceanographic region Kirkendale and Meyer (2004) found the opposite to be true in a study of the limpet *Patelloida profunda*, which has a short larval span and a presumed low dispersal potential when they discovered genetic homogeneity over a large area.

Reviews of genetic structuring of New Zealand coastal populations have been independently published by Ross *et al.* (2009) and Gardner *et al.* (2010). Gardner *et al.* (2010) found a variety of patterns out of the 58 studies examined: 28% exhibited no structure (presumed panmixia), 16% found an isolation by distance (IBD) model, 21% found divergence within and/or among populations, one study found the existence of East-West differentiation, but the majority of the studies (around 33%) found differentiation between a northern group (that includes the North Island and the northern part of the South Island) and South Island. This north and south genetic differentiation has been found in many studies of marine invertebrates. Apte and Gardner (2002), in their phylogeographic study of the green shell mussel, *Perna canaliculus*, postulated a barrier to gene flow occurring, in an area of intense upwelling on the north-east and north-west of the coasts of the South Island. This putative barrier to gene flow has been supported by a number of studies: the sea star *Patiriella regularis* (Ayers and Waters, 2005), three limpet species in the genus *Cellana* (Goldstein and Gemmell 2006) and most recently the snakeskin chiton *Sypharochiton pelliserpentis* (Veale and Lavery, 2011).

Each taxon has its own unique evolutionary history, as a result of vicariance or dispersal events. While it is unlikely there will be a unifying theory to explain the evolution of contemporary biodiversity in New Zealand; this study aimed to investigate the microevolutionary events that led to the formation of structure and connectivity between *O. chilensis*, this will fit into a broader description of the processes operating in New Zealand.

#### **1.1.5 Fishery of *Ostrea chilensis* in New Zealand and the threat of bonamiasis**

The study of marine connectivity is crucial to our understanding of the replenishment of fisheries and maintenance of diversity. Palumbi (2003) suggested the design of marine reserves whether for conservation or effective fisheries management, should have an understanding of: (1) direction larval transport in and out of the reserves, (2) whether the reserves are self-seeding or they depend on recruitment from other areas and (3) whether there is an exchange of recruits between sites. The degree of connectivity among populations has direct consequences for species evolution, development of disease resistance, and capacity of a metapopulation to adapt to climate change (Levin, 2006; Cowen and Sponaugle, 2009). Since an empirical study of larval movement is technically and logistically difficult population genetics can serve as a useful tool to estimate dispersal.

In New Zealand, the main *O. chilensis* fishery is in Foveaux Strait. There are small commercially fished stocks occurring in Tasman Bay and Golden Bay. The Foveaux Strait fishery is a high value, iconic fishery and has been fished for around 140 years. The total annual commercial catch of oysters from the Foveaux Strait fluctuated from 16 – 26 million individuals from 1907 to 1927, rising to 97 million in 1960. There was a significant decline in catch between 1962 and 1963 with an outbreak of the parasite *Bonamia exitosa* that kills mature dredge oysters. As a result, stock size and catch have both dwindled compared to historic levels.

The Foveaux Strait Oyster Fisheries Plan (2009) outlines a multi-disciplinary approach to management of the fishery. However, it did not include an analysis of the current population structuring or the levels of genetic diversity of oyster populations/stocks. This needs to be addressed given that a current assessment of genetic diversity can be used to assess the impact of future epizootics, or anthropogenic and environmental effects on genetic diversity and population structuring and also help inform current management policies.

The cause of bonamiasis: *Bonamia exitosa* is a haplosporidian parasite; microcells belonging to the genus *Bonamia* infect the haemocytes of different oyster species around the world: *B. ostreae* affects the flat oyster *Ostrea edulis* in Europe, the United States, Canada and Morocco; *B. angasi* and *B. roughleyi* are two species that are present in Australian oysters. There is little understanding of the phylogenetic relationships within the *Bonamia* genus, but work is being done to identify morphological and molecular characters that will be phylogenetically informative (López-Flores *et al.* 2007). *B. exitosa* proliferates utilising the oysters' energy reserves, especially late in the female reproductive cycle. Oysters are

exhausted to a point where they cannot maintain physiological functions or keep their shells closed, making them vulnerable to predation. The breakdown of dead oyster tissue releases infective particles that are dispersed by tidal currents. Infectious particles filtered from seawater by oysters during feeding pass through the gut to the blood stream and to the blood cells. *B. exitosa* multiplies rupturing haemocytes, exhausting the oysters' energy reserves and leading to its death. *In vitro* tests indicate that intensity of infection is exacerbated by stress and by high temperatures ( $>15^{\circ}\text{C}$ ), but is reduced by lower salinity. This suggests that changes in habitat, sea and climate, primary productivity, or stress from disturbance are also likely to be important triggers of epizootics (Cranfield *et al.* 2005).

Cranfield *et al.* (2005) found the presence of *B. exitosa* going back until 1964, which suggests that *B. exitosa* is endemic to the Foveaux Strait. Hine (1996) considered the annual infection of *O. chilensis* by *B. exitosa*, and the close life history of *B. exitosa*, the seasonal cycle and physiology of *O. chilensis* to be indicative of a long-term relationship between host and pathogen. If this is the case then it may be expected that *O. chilensis* has evolved some level of resistance, which could be of use for selective breeding strategies. Hine (1996) considered that the fishing practices had stressed the oysters, which reduced disease resistance; benthic modification and mechanical disturbance of the oysters could have caused this increased susceptibility (Cranfield *et al.* 2005). The epizootics were shown to commence in populations that have been fished intensely and ceased in populations around the margins that had been little fished until that time. A similar occurrence was found in Europe, where van Banning (1991) found much higher levels of infection in dredged beds compared to un-dredged beds of the European oyster *Ostrea edulis*.

Work is being carried out by the National Institute of Water and Atmospheric Research (NIWA), the Bluff Oyster Management Company and the Ministry of Fisheries, with the overall goal to better understand the relationship between bonamiasis and oysters. Their specific objectives were set out in the Foveaux Strait Oyster Fisheries Plan (2009). The Foveaux Strait Oyster Fisheries Plan (2009) outlines a multi-disciplinary approach to management of the fishery. The main focus is to understand the relationship with the *Bonamia* parasite, but other measures include: restocking certain areas with spat and trials of returning shell material to enhance beds. However, it did not include an analysis of the current population structuring or the levels of genetic diversity of oyster populations/stocks. This needs to be addressed given that a baseline assessment of genetic diversity can be used to assess the impact of future epizootics, or anthropogenic and environmental effects on

genetic diversity and population structuring and also help inform current management policies. It would also help inform levels of local diversity, which could be useful in spat re-stocking programmes.

Oyster aquaculture work in New Zealand began with the native rock oyster: *Saccostrea glomerata*, but this was largely replaced by the Pacific oyster (*Crassostrea gigas*) as it has a faster growth rate. Aquaculture New Zealand aims to increase the aquaculture industry to \$1 Billion by the year 2020. The focus is on three main species: Green shell mussels, Pacific Oyster and King Salmon. But the management plan also outlines the need of diversification into new aquaculture shellfish species of which, *O. chilensis* is a potential candidate. This thesis research will develop new genetic markers and assess level of population genetic diversity, which could support the industry goals for *O. chilensis* breeding projects.

#### **1.1.6 Population genetics of *Ostrea edulis*, a closely related species with similar problems**

Similar to *O. chilensis*, *Ostrea edulis* is a commercially important species that has been affected by a *Bonamia* parasite. It is much more fecund and has a shorter larval duration than *O. chilensis*; the affects of this life history strategy on the population structure can be compared. Many studies of the population genetics of *Ostrea edulis* in Europe have been made, and the some progress has been made into selectively breeding bonamiosis resistance in this species. For these reasons the work conducted on *O. edulis* is relevant to the research proposed in this thesis.

*O. edulis* is a hermaphroditic species that can be found in sub-tidal habitats. Its natural geographic ranges are along the European Atlantic coast from Norway to Morocco and all along the Mediterranean coastline, as well as in the Black Sea.

Like *O. chilensis* it is also affected by a *Bonamia* parasite. *Bonamia ostrea* was first reported in France in 1979 (Comps *et al.* 1980). More recently the *Bonamia* parasite which affects *O. chilensis* (*B. exitosa*) has been found in Spain (Abollo *et al.* 2008) and Italy (Narcisi *et al.* 2010). Dramatic stock decreases due to *Bonamia* outbreaks have been documented on the Atlantic coast of France (Brittany), the Netherlands, Spain, Denmark, Ireland and England. Due to fishing pressures and disease outbreaks *O. edulis* is said to be functionally extinct in Europe (Beck *et al.* 2011)

There is a larger body of work on the population genetics of *O. edulis* than *O. chilensis*. The nuclear genetic diversity and geographical structure of wild populations have been investigated using allozymes by Saavedra *et al.* (1995), although, the overall genetic differentiation was low (mean  $F_{st} = 0.062$ ), at one allozyme locus (arginine-kinase, (ARK\*)) there was a high degree of inter-population differentiation ( $F_{st} = 0.289$ ) resulting in variation along a geographical cline on either side of the Straits of Gibraltar. This pattern was interpreted as the result of secondary contact of Atlantic and Mediterranean populations during an interglacial period. The authors suggested that this cline is maintained by an environmental variable, such as salinity or temperature. The low overall inter-population differentiation is thought to be due to a number of factors: commercial exploitation (Johannesson *et al.* 1989), the prolonged larval phase (Saavedra *et al.* 1993) and restocking of depleted populations.

More recently, microsatellite loci have been used to study *O. edulis* populations. Launey *et al.* (2002) found high levels of polymorphism with an average of 18.5 alleles across five microsatellite loci. The Mediterranean populations were significantly more genetically diverse with a mean of 20.6 alleles per locus compared with the Atlantic population with 17.2 alleles per locus. Most of the microsatellite loci showed heterozygote deficiencies relative to Hardy-Weinberg equilibrium (HWE) this is a common phenomenon in bivalves (Hedgecock, 2007). The estimate of  $F_{st}$  was 0.019, indicating a heterogeneous distribution of genetic variability. Regression of  $F_{st}/(1 - F_{st})$  over coastal distance in kilometres showed a positive correlation between genetic and geographical distances supporting an isolation by distance model. This finding was congruent with the allozyme data of Saavedra *et al.* (1995) but the study failed to reveal a genetic discontinuity between Atlantic and Mediterranean populations.

Diaz-Almela *et al.* (2004) assessed the population structure of *O. edulis* using single-strand conformational polymorphism of a 313 base pair fragment of the mitochondrial 12S-rRNA gene using the same samples as Launey *et al.* (2002). Fourteen haplotypes were observed, with one being dominant in the Mediterranean samples and one in the Atlantic populations. The geographically extreme populations sampled in Norway and the Black Sea appeared to be differentiated and had a third group of haplotypes at high frequency. These data confirm the isolation by distance pattern found in the study by Launey *et al.* (2002) and Saavedra *et al.* (1993). The haplotype differences among populations were high, which might reflect the small effective population sizes in most locations. Also, a 10-fold difference was observed in

$F_{st}$  values between mitochondrial and nuclear genomes this could be due to an unbalanced sex ratio or sex-based differential reproductive success between males and females.

In Northern European populations (The British Isles, France, The Netherlands and Norway) of *O. edulis*, Sobolewska and Beaumont (2005) observed low levels of genetic subdivision (mean  $F_{st} = 0.039$ ) using microsatellite markers and suggested that human-mediated transport of oysters played a significant role in the observed pattern.

In the French Atlantic population, Tarris *et al.* (2009) tested predictions of the sweepstakes hypothesis. They assayed four microsatellite markers and a 12S mitochondrial DNA sequence fragment in four *O. edulis* recruitment cohorts. According to the sweepstakes hypothesis the temporal cohorts should be differentiated from each other (Hedgewood, 2011). However, while no differentiation was found, the authors did find differentiation between cohorts within the mitochondrial data set. This difference is thought to be due to lower effective population size of the mitochondrial genome, and hence its increased sensitivity to drift compared to the nuclear genome. The signal found in the mitochondrial DNA data might indicate a certain limitation in the number of contributing female parents in this species.

Since there has been much work completed on *O. edulis* it serves as a useful guideline for the work that could be done with *O. chilensis*. It will also be interesting to compare and contrast the effect of life history traits on the population genetics of the two species.

### **1.1.7 Aims of the present research**

There were two main goals of this thesis research: to investigate the population genetic structure of *O. chilensis* and to use a genomic data set to develop bioinformatic resources for the species. Firstly, the research aimed to describe the population genetic structure of *Ostrea chilensis*. This was achieved by: (1) testing for genetic structure at different spatial scales, (2) testing if the genetic structure could be explained by an isolation-by-distance profile, and (3) testing for putative barriers to gene flow. A further aim was to seek the environmental variables that best explain the apparent genetic structure. The overall findings were compared with other studies of *O. chilensis*, closely related oyster species, and other studies of New Zealand marine taxa aiming to compare patterns of apparent structure.

The relevance of this part of the study is that it will contribute to the understanding of the patterns and processes that are generating genetic structure in New Zealand marine taxa. It

will also provide information that will aid the management of oyster fisheries and future aquaculture projects through the identification of genetic stocks.

The second goal was to develop bioinformatic resources for *O. chilensis* and describe patterns of molecular evolution. This was achieved by: (1) identifying microsatellite loci in a genomic database and categorising them based on repeat motif, array length and contribution to the genome, (2) identification of potentially orthologous repetitive sequences, and (3) partial assembly of the *O. chilensis* mitogenome, using new gene regions from this partial genome the phylogenetic position of *O. chilensis* was tested.

The relevance of this part of the study is that it will be the first to develop useful bioinformatic resources for this commercially important species, and it will contribute to studies that have investigated molecular evolution.

### **1.1.7 Organisation of the thesis**

For ease of reference the citations and appendices are listed at the end of each chapter. This introductory chapter is followed by five data chapters and a final discussion.

Chapter Two used genomic sequencing to develop novel microsatellite markers for *O. chilensis*. The microsatellite sequences were categorised and compared with other studies of microsatellite loci. Finally, the microsatellite loci were compared to loci from oysters with an aim to find orthologous loci.

Chapter Three used randomly amplified polymorphic DNA (RAPD) to measure the genetic diversity, and determined the patterns of genetic structure among 11 *O. chilensis* sampling sites in Chile and New Zealand.

Chapter Four used the levels of genetic variation measured for *O. chilensis* samples in chapter three to test for associations with geospatial and environmental variables. This aimed to find the set of environmental or geospatial variables that best explained the apparent genetic structure.

Chapter Five further developed bioinformatic resources for *O. chilensis*: the partial mitochondrial genome was reconstructed from the genomic data set. This mitogenome was then used to analyse the phylogenetic position of *O. chilensis* by comparison to other oyster mitogenomes.

Chapter Six further analysed the evolutionary processes that act on mitochondrial genes by comparing protein coding genes from oyster mitogenomes.

Chapter Seven summarises the major findings of the research and presents a synthesis of the results from each chapter. Relevance to New Zealand fisheries and Aquaculture is discussed. Finally, developments in genomics are discussed with relevance to possible future studies.

## 1.2 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(9), 1617-1628.
- Ayers KL, Waters JM. (2005). Marine biogeographic disjunction in central New Zealand. *Marine Biology*, 147(4), 1045-1052.
- Ayre DJ, Hughes TP. (2000). Genotypic diversity and gene flow in brooding and spawning corals along the Great Barrier Reef, Australia. *Evolution*, 54(5), 1590-1605.
- Baker P. (1995). Review of ecology and fishery of the Olympia oyster, *Ostrea lurida* with annotated bibliography. *Journal of Shellfish Research*, 14(2), 501-518.
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK. (2000). Biogeography: a marine Wallace's line? *Nature*, 406(6797), 692-693.
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK. (2002). Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, 11(4), 659-674.
- Bell JJ, Constable HB, Hannan D, Ritchie PA, Zuccarello GC. (2010). *Multi-species coastal marine connectivity: a literature review with recommendations for further research*. Ministry of Fisheries.
- Beu AG. (1990). Molluscan generic diversity of New Zealand Neogene stages: extinction and biostratigraphic events. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 77(3), 279-288.
- Broekhuizen N, Lundquist CJ, Hadfield MG, Brown SN. (2011). Dispersal of oyster (*Ostrea chilensis*) larvae in Tasman Bay inferred using a verified particle tracking model that incorporates larval behavior. *Journal of Shellfish Research*, 30(3), 643-658.
- Buroker NE, Chanley P, Cranfield HJ, Dinamani P. (1983). Systematic status of two oyster populations of the genus *Tiostrea* from New Zealand and Chile. *Marine biology*, 77(2), 191-200.
- Cerrato RM, Caron DA, Lonsdale DJ, Rose JM, Schaffner RA. (2004). Effect of the northern quahog *Mercenaria mercenaria* on the development of blooms of the brown tide alga *Aureococcus anophagefferens*. *Marine Ecology Progress Series*, 281, 93-108.

- Coen LD, Brumbaugh RD, Bushek D, Grizzle R, Luckenbach MW, Posey MH, Tolley S. (2007). Ecosystem services related to oyster restoration. *Marine Ecology Progress Series*, 341, 303-307.
- Cole HA. (1941). The fecundity of *Ostrea edulis*. *Journal of the Marine Biological Association of the United Kingdom*, 25(02), 243-260.
- Collin, R. (2001). The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Calyptraeidae). *Molecular Ecology*, 10(9), 2249-2262.
- Cowen RK, Paris CB, Srinivasan A. (2006). Scaling of connectivity in marine populations. *Science*, 311(5760), 522-527.
- Cranfield HJ, Dunn A, Doonan IJ, Michael KP. (2005). *Bonamia exitiosa* epizootic in *Ostrea chilensis* from Foveaux Strait, southern New Zealand between 1986 and 1992. *ICES Journal of Marine Science: Journal du Conseil*, 62(1), 3-13.
- Cranfield HJ, Michael KP. (1989). Larvae of the incubatory oyster *Tiostrea chilensis* (Bivalvia: Ostreidae) in the plankton of central and southern New Zealand. *New Zealand journal of marine and freshwater research*, 23(1), 51-60.
- Cranfield HJ, Rowden AA, Smith DJ, Gordon DP, Michael KP. (2004). Macrofaunal assemblages of benthic habitat of different complexity and the proposition of a model of biogenic reef habitat regeneration in Foveaux Strait, New Zealand. *Journal of Sea Research*, 52(2), 109-125.
- Cranfield HJ. (1968). An unexploited population of oysters, *Ostrea lutaria* Hutton, from Foveaux Strait: Part I. Adult stocks and spatfall distribution. *New Zealand journal of marine and freshwater research*, 2(1), 3-22.
- Di Salvo LH, Alarcon E, Martinez E. (1983). Induced spat production from *Ostrea chilensis* Philippi 1845 in mid-winter. *Aquaculture*, 30(1), 357-362.
- Díaz D, Zabala M, Linares C, Hereu B, Abelló P. (2005). Increased predation of juvenile European spiny lobster (*Palinurus elephas*) in a marine protected area. *New Zealand Journal of Marine and Freshwater Research*, 39(2), 447-453.
- Foighil DÓ, Marshall BA, Hilbish TJ, Pino MA. (1999). Trans-Pacific range extension by rafting is inferred for the flat oyster *Ostrea chilensis*. *Biological Bulletin*, 122-126.

- Gardner JPA, Bell JJ, Constable HB, Hannan D, Ritchie PA, Zuccarello GC. (2010). Multi-species coastal marine connectivity: a literature review with recommendations for further research: Victoria University of Wellington report (ZBD2009-10) to the New Zealand Ministry of Fisheries. Wellington: New Zealand Ministry of Fisheries. p.50.
- Goldstien SJ, Schiel DR, Gemmell NJ. (2006). Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology*, 15(11), 3259-3268.
- Gordon DP, Beaumont J, MacDiarmid A, Robertson DA, Ahyong ST. (2010). Marine biodiversity of Aotearoa New Zealand. *PLoS One*, 5(8), e10905.
- Hedgecock D, Launey S, Pudovkin AI, Naciri Y, Lapegue S., Bonhomme F. (2007). Small effective number of parents ( $N_b$ ) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Marine Biology*, 150(6), 1173-1182.
- Hewitt JE, Thrush SF, Halliday J, Duffy C. (2005). The importance of small-scale habitat structure for maintaining beta diversity. *Ecology*, 86(6), 1619-1626.
- Hine PM. (1996). The ecology of *Bonamia* and decline of bivalve molluscs. *New Zealand Journal of Ecology*, 20(1), 109-116.
- Jeffs AG, Creese RG. (1996). Overview and bibliography of research on the Chilean oyster *Tiostrea chilensis* (Philippi, 1845) from New Zealand waters. *Journal of shellfish research*, 15(2), 305-311.
- Jeffs AG, Hickman RW. (2000). Reproductive activity in a pre-epizootic wild population of the Chilean oyster, *Ostrea chilensis*, from southern New Zealand. *Aquaculture*, 183(3), 241-253.
- Jeffs AG, Hooker SH, Creese RG. (1997). Variability in life history characters of the Chilean oyster *Tiostrea chilensis* (Philippi, 1845). *New Zealand Journal of Marine and Freshwater Research*, 31(4), 487-495.
- Jozefowicz CJ, Foighil DÓ. (1998). Phylogenetic analysis of southern hemisphere flat oysters based on partial mitochondrial 16S rDNA gene sequences. *Molecular phylogenetics and evolution*, 10(3), 426-435.
- Kirkendale LA, Meyer CP. (2004). Phylogeography of the *Patelloida profunda* group (Gastropoda: Lottidae): diversification in a dispersal-driven marine system. *Molecular Ecology*, 13(9), 2749-2762.

- Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y. (2002). Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism. *Journal of Heredity*, 93(5), 331-351.
- López-Flores I, Suárez-Santiago VN, Longet D, Saulnier D, Chollet B, Arzul I. (2007). Characterization of actin genes in *Bonamia ostreae* and their application to phylogeny of the Haplosporidia. *Parasitology*, 134(14), 1941-1948.
- Millar RH, Hollis PJ. (1963). Abbreviated pelagic life of Chilean and New Zealand oysters. *Nature* 197, 512 – 513
- Newell RI, Koch EW. (2004). Modelling seagrass density and distribution in response to changes in turbidity stemming from bivalve filtration and seagrass sediment stabilization. *Estuaries*, 27(5), 793-806.
- Ó Foighil D, Taylor DJ. (2000). Evolution of parental care and ovulation behaviour in oysters. *Molecular phylogenetics and evolution*, 15(2), 301-313.
- Piazza BP, Banks PD, La Peyre MK. (2005). The potential for created oyster shell reefs as a sustainable shoreline protection strategy in Louisiana. *Restoration Ecology*, 13(3), 499-506.
- Powell EN, Klinck JM. (2007). Is oyster shell a sustainable estuarine resource? *Journal of Shellfish Research*, 26(1), 181-194.
- Rocha-Olivares A, Vetter RD. (1999). Effects of oceanographic circulation on the gene flow, genetic structure, and phylogeography of the rosethorn rockfish (*Sebastes helvomaculatus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 56(5), 803-813.
- Rotschild B, Ault JS, Goulletquer P, Heral M. (1994). Decline of the Chesapeake Bay oyster population: a century of habitat destruction and overfishing. *Marine Ecology Progress Series*, 111, 29-39.
- Saavedra C, Zapata C, Guerra A, Alvarez G. (1993). Allozyme variation in European populations of the oyster *Ostrea edulis*. *Marine Biology*, 115(1), 85-95.
- Saavedra C, Zapata C, Alvarez G. (1995). Geographical patterns of variability at allozyme loci in the European oyster *Ostrea edulis*. *Marine biology*, 122(1), 95-104.

- Seale EM, Zacherl DC. (2009). Seasonal settlement of Olympia oyster larvae, *Ostrea lurida* Carpenter 1864 and its relationship to seawater temperature in two southern California estuaries. *Journal of Shellfish Research*, 28(1), 113-120.
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787-792.
- Tamburri MN, Luckenbach MW, Breitburg DL, Bonniwell SM. (2008). Settlement of *Crassostrea ariakensis* larvae: effects of substrate, biofilms, sediment and adult chemical cues. *Journal of Shellfish Research*, 27(3), 601-608.
- Taylor MS, Hellberg ME. (2003). Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* 299(5603), 107-109.
- Thrush SF, Dayton PK. (2002). Disturbance to marine benthic habitats by trawling and dredging: implications for marine biodiversity. *Annual Review of Ecology and Systematics*, 449-473.
- Toro JE, Aguila PR. (1996). Genetic differentiation of populations of the oyster *Ostrea chilensis* in southern Chile. *Aquatic Living Resources*, 9(01), 75-78.
- Toro JE, González CP. (2009). The genetic structure of the Chilean oyster (*Ostrea chilensis* Philippi, 1845) in natural populations of southern Chile based on RAPDs analysis. *Revista De Biología Marina Y Oceanografía*, 44(2), 467-476.
- Toro JE, Morandé PR. (1998). Effect of food ration and temperature on length of brooding period, larval development and size of pediveligers released in the Chilean oyster *Ostrea chilensis*. *Journal of the World Aquaculture Society*, 29(2), 267-270.
- Van Banning P. (1991). Observations on bonamiasis in the stock of the European flat oyster, *Ostrea edulis*, in the Netherlands, with special reference to the recent developments in Lake Grevelingen. *Aquaculture*, 93(3), 205-211.
- Veale AJ, Lavery SD. (2011). Phylogeography of the snakeskin chiton *Sypharochiton pelliserpentis* (Mollusca: Polyplacophora) around New Zealand: are seasonal near-shore upwelling events a dynamic barrier to gene flow? *Biological Journal of the Linnean Society*, 104(3), 552-563.

- Waldbusser GG, Voigt EP, Bergschneider H, Green MA, Newell RI. (2011). Biocalcification in the eastern oyster (*Crassostrea virginica*) in relation to long-term trends in Chesapeake Bay. *Estuaries and Coasts*, 34(2), 221-231.
- Wallis GP, Trewick SA. (2009). New Zealand phylogeography: evolution on a small continent. *Molecular Ecology*, 18(17), 3548-3580.
- Walne PR. (1964). Observations on the fertility of the oyster (*Ostrea edulis*). *Journal of the Marine Biological Association of the United Kingdom*, 44(02), 293-310.
- Westerskov K. (1980). *Aspects of the biology of the dredge oyster Ostrea lutaria Hutton, 1873* Ph. D Thesis, University of Otago, NZ.
- White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences* 277, 1685–1694.
- Winter JE, Acevedo MA, Navarro JM. (1984). Quempillen estuary, an experimental oyster cultivation station in southern Chile. Energy balance in *Ostrea chilensis*. *Marine ecology. Progress series*, 20(1-2), 151-164.
- Zardi GI, Nicastro KR, McQuaid CD, Hancke L, Helmuth B. (2011). The combination of selection and dispersal helps explain genetic structure in intertidal mussels. *Oecologia*, 165(4), 947-958.

# Chapter Two: A study of the microsatellite loci of the Bluff Oyster (*Ostrea chilensis*)

---

*“Microsatellites: simple sequences with complex evolution” – Hans Ellegren*

## 2.0 Aims

The aims of this chapter were to identify microsatellite DNA repeats in the partial genome sequence of *Ostrea chilensis*. The goal was to find length-polymorphic loci among the candidate microsatellite repeats that could be used for a population genetic study. Further aims were to use the sequence reads to answer the following questions: (1) what are the most common types of repeat? (2) What is the most common repeat motif across each type of repeat? (3) What are the length differences between different repeat motifs? And finally (4) what are the most common lengths of the array? The purpose of this is to seek to describe the processes that lead to microsatellite evolution, which could aid developments in the better use of microsatellites as genetic markers.

## 2.1 Introduction

*Ostrea chilensis* is a commercially important fisheries species in New Zealand, but there are currently limited genetic resources available for this species. This thesis research will be the first to identify and design microsatellite PCR-based markers using genomic fragments for this species.

For studies of micro-evolutionary patterns and processes, microsatellites represent an ideal marker for detecting population isolation, levels of gene flow and inbreeding; because these loci are highly polymorphic. The use of DNA sequencing on - even just a fraction - of a 454 DNA sequencer flow cell plate can provide a fast and efficient method of microsatellite loci discovery, for which primers can be designed to target the flanking region sequences.

Another method used to identify useful polymorphic microsatellite markers is the transfer of markers from a closely related species. This approach could use microsatellite PCR primers developed for a closely related species to search the 454 DNA sequence data set for matches in the species of question. An analysis of potentially transferable markers is particularly

useful when investigating the evolutionary processes that generate microsatellites (e.g. Chambers and MacAvoy, 2000). They are also useful in the study of mechanisms involved in population divergence and speciation (e.g. Noor and Feder, 2006). Cross-species transferability has had variable levels of success (Zhang and Hewitt, 2003; reviewed by Barbara *et al.* 2007). In bivalves, Kang *et al.* (2013) successfully used microsatellite loci in a range of mussel species, although the levels of success varied depending on the species and the loci tested. There are many other examples of successful cross amplification in other bivalves including for example: lampsilines (Galbraith *et al.* 2011) and scallops (Marín *et al.* 2012).

### 2.2.1 Microsatellites and their evolutionary patterns and processes

Microsatellites are tandem repeats of 1-6 nucleotides found in high frequency in the nuclear genomes of every higher organisms tested to date (Toth *et al.* 2000; Mayer *et al.* 2010). Within a population a microsatellite motif can vary in length from around 5 to 40 repeats. Dinucleotide repeats account for the majority of microsatellite repeats in most species (Li *et al.* 2002). Microsatellites are not often associated with gene coding regions (although there are many examples where they do occur in genes such as the association with human diseases e.g. colorectal and endometrial cancer genomes Kim *et al.* (2013)) and, therefore, they are thought to be selectively neutral ‘junk DNA’. However, Li *et al.* (2002) reviewed the putative function of microsatellites, such as chromatin organisation, regulation of gene activity and recombination, which might mean that they do, from time to time, experience selection. Moreover, if a microsatellite locus is in close proximity to a gene, or genes, under selection then it might appear to be non-neutral due to the ‘hitchhiking’ effect (Slatkin, 1995).

The primary mutational mechanisms leading to the generation of microsatellites were demonstrated empirically in *Escherichia coli* by Levinson and Gutman, (1987). They were shown to be a combination of polymerase slippage during replication, proofreading errors, and unequal crossing over. A mathematical model derived to explain the equilibrium distribution of allelic frequencies in a finite population, known as the step wise mutation model (SMM) (Kimura and Ohta, 1978) has been adapted to model microsatellite evolution. It postulates that mutations add or subtract one unit of the repetitive array independent of the length of the array. However, microsatellites have been shown to have an upper size limit and different mutabilities based on length, both of which are incompatible with the SMM (Ellegren, 2004). More sophisticated models have been developed that account for a genome

wide distribution of microsatellites being in a balance between length and point mutations (Kruglyak *et al.* 1998). The 'birth' (evolutionary origin) of microsatellites was demonstrated in the study of Messier *et al.* (1996). In this study microsatellite regions contained in the  $\alpha$ -globin pseudogene were examined in a phylogenetic context to show that a single nucleotide substitution occurred in the lineage leading to the African apes that created a tetra-nucleotide motif, which presumably allowed slippage to occur and then subsequently expanded. The 'death' of a microsatellite was shown in the study of Taylor *et al.* (1999). In this study they demonstrate that a substitution mutation can interrupt the repeat, which prevents slippage and therefore stabilises the region. This study demonstrates that point mutations break up microsatellite repeats and reduces the mutation rate of the locus. A number of factors have been investigated that could account for mutation rate differences in different microsatellite loci. These include: (1) length of the microsatellite sequence, (2) flanking sequence, (3) repeat motif, (3) repeat type and (4) genome location. Studies in these areas have shown conflicting results. For example, in human tetranucleotide repeats the length of the microsatellite has been shown to be positively correlated with increased mutation rate (Leopoldino and Pena, 2003). Conversely, a downward bias has been shown in long microsatellite repeats in *Drosophila melanogaster* (Harr and Schlötterer, 2000). There seems to be a relationship with the flanking sequence: allelic diversity was found to be negatively correlated with GC content of the flanking sequences in alligator microsatellites (Glen *et al.* 1996). Studies have detected a high similarity of the flanking sequences regions of different microsatellite loci in Lepidoptera (Meglécz *et al.* 2004) and molluscs (McInerney *et al.* 2010). The relationship between repeat motif and mutation rate is less clear. In *D. melanogaster* the (CA)<sub>n</sub> repeat appears to have the highest mutation rates and (AT)<sub>n</sub> appears to have the lowest mutation rate. Contrastingly, in humans and chimpanzees (CG)<sub>n</sub> has low and (AT)<sub>n</sub> has high mutability (Kelkar *et al.* 2008). In the study of Kelkar *et al.* (2008) mutability was found to increase with repeat type: mono-nucleotides had the lowest mutability and tetra-nucleotides had the highest mutability. Location in the genome of the microsatellite has also been shown to be a determinant of microsatellite mutability. Mononucleotides were shown to have a higher mutability on sex chromosomes than autosomes and inside rather than outside of *Alu* repeats (Kelkar *et al.* 2008).

### 2.1.2 Applications

The development of genetic markers is becoming increasingly important in aquaculture because they have applications for marker assisted selection, linkage mapping and in identifying quantitative trait loci (QTL). Although the full utilisation of genomics has not been seen in aquaculture, it is likely in the future developments of genomic resources will support improvements to fisheries management and development of new species and varieties for aquaculture.

### 2.1.3 Hypotheses

#### **Hypothesis One: Length of the microsatellite array.**

Short microsatellite arrays will be more common than long microsatellite arrays in the 454 DNA database. This will indicate that there is a constraint on the length of a microsatellite locus. This pattern will be the same regardless of microsatellite type.

#### **Hypothesis Two: Contribution of microsatellite repeat unit to genomic DNA and the most common repeat motif.**

It is expected that di-nucleotide repeats will be the most numerous in the genome of *O. chilensis*. This pattern is expected because it has been reported in humans and many other species. Most studies have shown the repeat motif to be richer in (AT)<sub>n</sub> than (CG)<sub>n</sub>, this is thought to occur as there are three hydrogen bonds between Cysteine and Guanine creating more stability than the two hydrogen bonds between Adenine and Thymine.

#### **Hypothesis Three: Search for potentially orthologous microsatellite loci**

It is expected that some of the microsatellite flanking sequences identified in the 454 DNA sequences generated from *O. chilensis* will share some homology with flanking sequences of microsatellite loci identified in closely related species. A blast hit of  $\leq E - 20$  between flanking sequences could mean the microsatellite loci is either orthologous or paralogous.

## **2.2 Methods**

### **2.2.1 DNA extraction and 454 Pyrosequencing**

Using oyster tissue from individuals collected from the Marlborough Sounds (collection locality is detailed in Table 3.1) a small amount of the preserved mantle tissue (around 5 mm<sup>3</sup>) was taken using flame/ethanol sterilized forceps and scissors. The DNA was extracted using a Genomic DNA extraction kit (Geneaid Biotech Ltd; Taiwan). Following the manufacturer's instructions, tissue was homogenised with a micropestle, then digested using Proteinase K and chaotropic salt, which lysed the cells and degraded the proteins. This chaotropic salt forms a bridge and allows DNA to bind to a glass fibre matrix in a spin column and while it is bound contaminants are removed using several washes of the column with a salt-ethanol buffer. The DNA was then eluted from the glass fibre matrix using a low salt buffer.

The extracted genomic DNA was electrophoresed in a 1.5% agarose gel and stained with ethidium bromide to check for a single band of DNA. Multiple bands and/or smears were indication that the genomic DNA had been degraded as a result of the preservation or extraction methodology and was not suitable for high-throughput sequencing. The quantity of DNA was estimated using a Nanodrop spectrophotometer. Genomic DNA from an individual that gave the best purity value (based on the ratio of absorbance at 260nm and 280nm) was sent for a 1/8 plate sequencing run on a Roche 454. The subsequent DNA database was examined using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

### **2.2.2 Identification of microsatellite loci**

The 454 sequence reads were searched for microsatellite-type repeats using the software PHOBOS (Version 3.3.11 Mayer 2006-2010) using the following settings: perfect search with minimum thresholds of 100% repeat perfection, three repetitive units of 2-6 bp motifs and a total length of 10 bp (for example this would be five repeat units for a dinucleotide). This number of repeats was chosen as it represents the number of repeats in typical microsatellites. The mononucleotide repeats were excluded because the 454 DNA sequencing technique has difficulty accurately reading homopolymers and misreads could generate spurious results.

PHOBOS standardises the output automatically, therefore, there are only four types of dinucleotide repeat: AC = (CA = GT = TG), AG = (GA = CT = TC), AT = (TA), and GC = (CG). In this chapter comparisons of other studies will be standardised this way. The frequencies of each microsatellite type, motif for each type, and length of the arrays were summarised in an Excel spreadsheet. A possible bias of this method is that there could be overlapping or duplicate DNA sequences of the same repeat regions.

### **2.2.3 Screening of reads for microsatellite sequences and development of PCR primers**

The sequences were screened for microsatellite loci using iQDD V1.3 (Meglecz *et al.* 2010). This is a PERL programme that uses Blast, Clustal and Primer3 to trim the sequences, detect microsatellites and design a range of suggested PCR primer pairs. These primer pairs were organised according to primer size in an Excel spreadsheet. The putative primers and microsatellite loci were mapped to the 454 data using GENEIOUS. The purpose of doing this is to make sure that the primers are not too close to the repeat region and to visually check the repeat motif to make sure it is not a compound repeat (which has not been identified as such by the program), or if there are multiple repeats between primer pairs, which can confuse later analysis.

### **2.2.4 PCR amplification**

Following the advice of Leese and Held (2011), the PCR protocol was optimised on a range of geographically disparate individuals to increase the chance of finding polymorphism in the length of the microsatellite array. Specifically, one individual from each of the sampled populations detailed in Table 3.1 was used as a testing panel for the PCR primers. Each primer pair and individual were subjected to three different “touchdown protocols” modified from Korbie and Mattick (2008) (Table 2.1). Primers that did not work with one of these three protocols were then tested with a gradient thermal cycling profile with annealing temperatures ranging from 65-50°C. Primers were labelled with a 5' M13 tag (Schuelke, 2000). This technique requires the addition of an extra step in the thermal cycling profile to allow annealing of a 3<sup>rd</sup> fluorescently-labelled primer.

PCR products were then electrophoresed in a 1% agarose gel (ethidium bromide stained) and visualised under UV light. Primer pairs that resulted in a single band at around the expected size region were further analysed. Those that still gave multiple bands were rejected because

the result of non-specific binding during PCR could make the loci difficult to size properly during the genotyping analysis.

**Table 2.1:** PCR Protocol adapted from Korbie and Mattick (2008)

Phase 1	Step	Temperature °C	Time
1	Denature	95	3 min
2	Denature	95	30 sec
3	Anneal	70/62.5/55	30 sec
4	Extension	72	30 sec
<i>Steps 2-4 are repeated x15 with a -0.5 °C drop each cycle</i>			
Phase 2	Step	Temperature °C	Time (sec)
1	Denature	95	30
2	Anneal	62.5/55/47.5	45
3	Extension	72	30
<i>Steps 1 – 3 are repeated x15</i>			
M13 Phase	Step	Temperature °C	Time (sec)
1	Denature	95	30
2	Anneal	53	45
3	Extension	72	30
Termination	Step	Temperature °C	Time
1	Extension	72	10 min
2	Halt	4	∞

### 2.2.5 Genotyping

The PCR product that amplified a single band in the expected size region when compared to a known molecular weight size standard (i.e. a ladder) was sent to Massey Genome Service for Sequencing Capillary Separation Service. The allele peaks were visually examined for conformity to the expected type of repeat unit and the expected size region using Peak Scanner v1.0 (Applied Biosystems). Following the advice of Leese and Held (2011) the loci that were difficult to interpret were amplified again with an extended final extension phase of 30 minutes.

### 2.2.6 Summary Statistics

In instances where alleles could be identified for a particular primer pair, the size of the two alleles was recorded in an Excel spreadsheet. The expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ), p-values of two-tailed tests for deviations from Hardy-Weinberg Equilibrium ( $P_{HWE}$ ) and Weir and Cockerham ( $F_{IS}$ ) were then calculated in the Excel plugin GenAlEx (Smouse and Peakal, 2006, 2012).

### 2.2.7 Search for orthologous repetitive loci

In order to study the evolution of microsatellite sequences in oysters and the possibility of transferring primers designed for microsatellite loci to *O. chilensis*, microsatellite containing sequences were compared to those available on GenBank. A custom 454 database was BLASTED against GenBank accessions that had the search terms ‘Oysters’ and ‘Repetitive DNA’, before this the repetitive sequences from the 454 DNA database were masked with a string of N’s with the GENEIOUS plugin PHOBOS. This was done in order to minimise irrelevant BLAST hits. The maximum number of hits was set to 1 with a cut of  $1 \times e^{-20}$ , so that only the one most probable hit was returned.

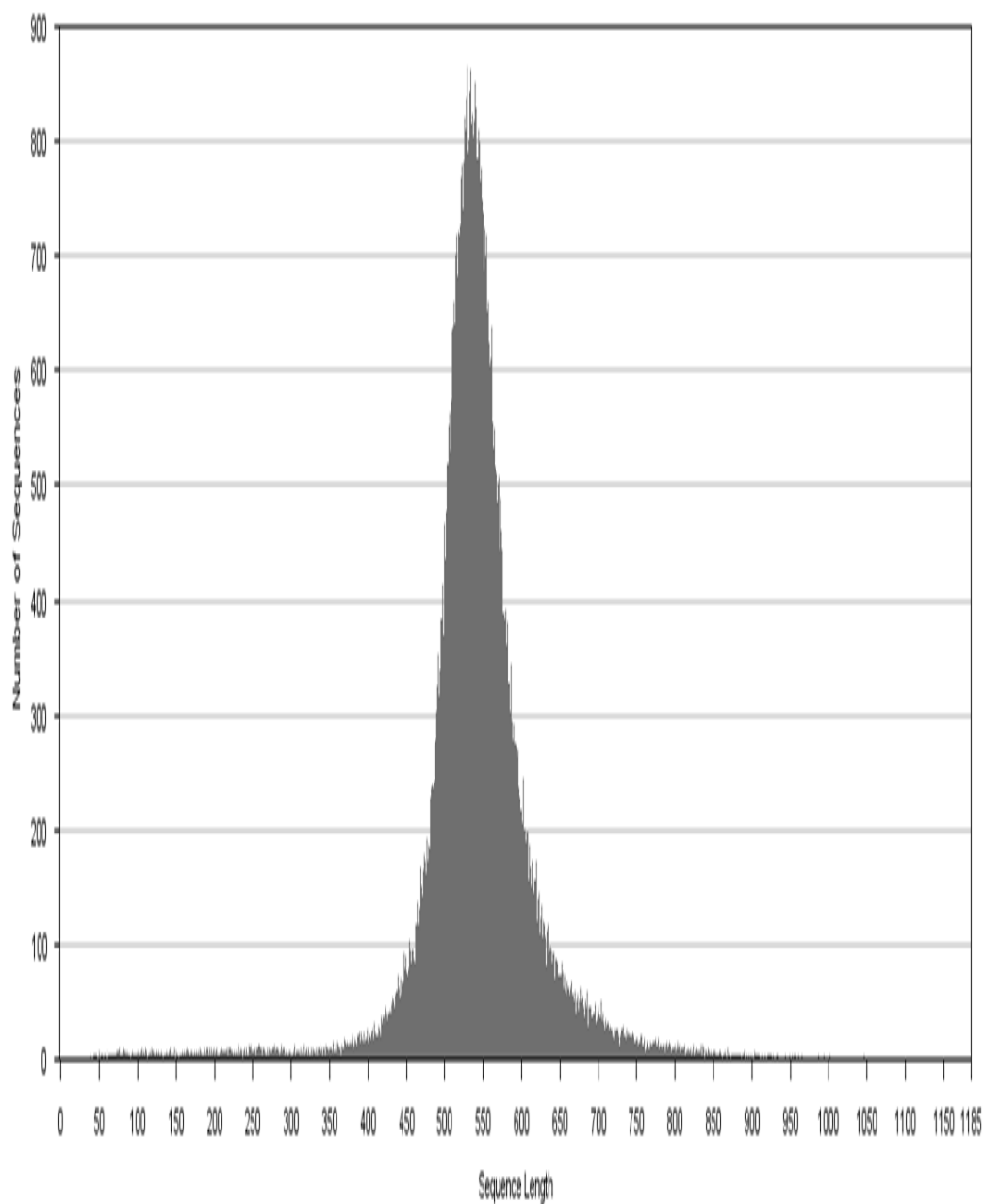
## 2.3 Results

### 2.3.1 Success of the 454 sequencing

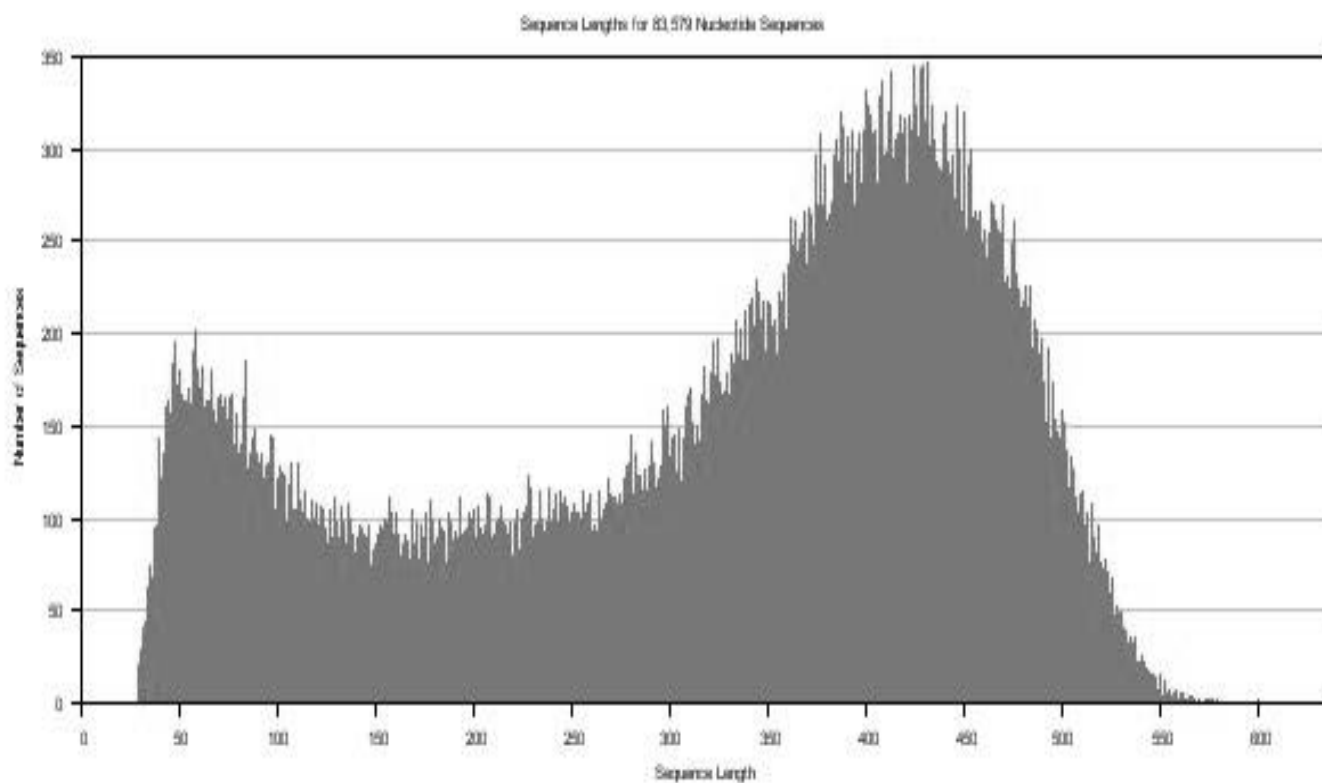
The run yielded 85,579 sequence reads that were between 39 and 1,185 bp long (Figure 2.1). The 454 sequencing data were clipped using GENEIOUS (based on clipping information in the \*.sff file) and tested for potential problems using FastQC: the results are displayed in Table 2.2. After clipping, the mean sequence length was 319 bp (Figure 2.1). Some small problems were found in the sequencing: the phred score (this is a quality score based on the peak shape and peak resolution of each base of a sequencing reaction) was low at the ends of the sequences, but these bases can be removed from the database, and the GC content had a second peak which was higher than the expected distribution which could indicate contamination (Figure 2.2). Contamination in sequences of interest can be further examined by BLASTING the sequences against GenBank.

**Table 2.2:** Problems and warning suggested by the program FastQC for the clipped 454 sequencing data set

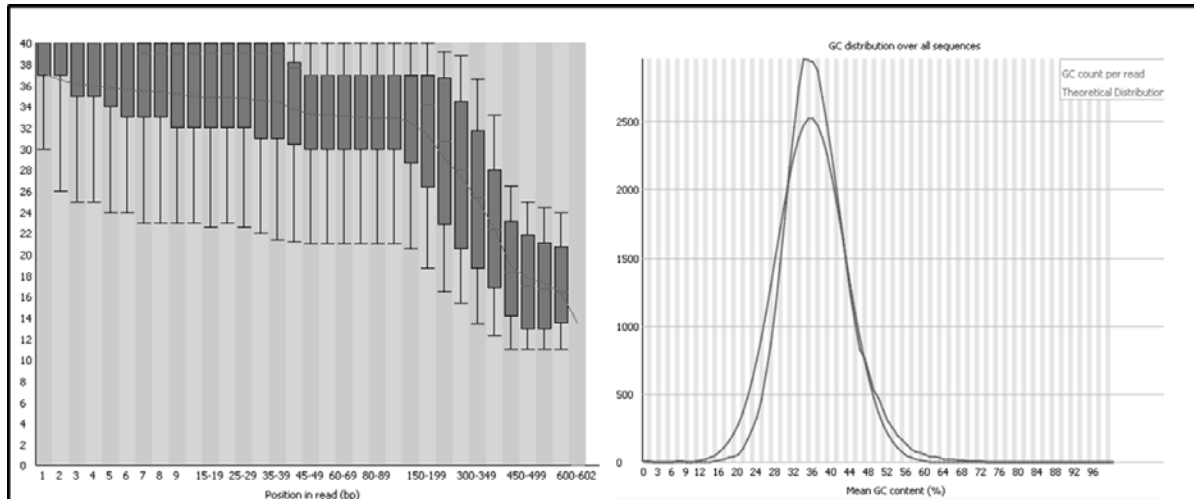
Fast QC Test	Clipped data	Comment
Basic Statistics	Good	GC content = 35%
Per base sequence quality	Problem	Figure 2.2, at around 300 bp there is a drop in the average phred score.
Per sequence quality scores	Good	However, most sequences have a good phred score
Per base sequence content	Problem	Only affected the first few bases and the last (500+ bp)
Per base GC content	Problem	Only affected the first few bases and the last (500+ bp)
Per sequence GC content	Warning	Secondary peak indicative of contamination
Per Base N content	Good	-
Sequence Length Distribution	Problem	Figure 2.2 no longer normally distributed (due to being clipped)
Sequence Duplication Level	Good	Only 18.59%
Over-represented sequences	Good	-
Kmer Content	Good	-



**Figure 2.1:** Screen shot from GENEIOUS, length distribution of reads before clipping



**Figure 2.2:** Screen shot from GENEIOUS, length distribution of reads after clipping



**Figure 2.3:** Screen shots of selected FastQC results. Left: Boxplot of Phred quality scores over all sequences, without cropping. Notice the changing proportions on the x axis. Right: Comparison between expected GC content per sequence (lower line) and the actual GC content.

### 2.3.2 Identification of microsatellite loci and a description of the length of array, motif and contribution to the genome for each repeat type

The software program PHOBOS identified 6,208 dinucleotide repeats, 7,326 trinucleotide repeats, 2,414 tetranucleotide repeats, 33 hexanucleotide repeats and 357 pentanucleotide repeats. These were annotated using the GENEIOUS software and placed in a separate database for further analysis.

To assess the contribution of the microsatellite regions to the genome size of *O. chilensis*, the total lengths of microsatellite regions were calculated (Figure 2.5). In terms of their contribution to the total DNA of the genome, trinucleotide and dinucleotide repeats made the biggest contributions to the genome. Further analysis on repeat motif and length of the array was compared for each repeat type. Results for di-, tri- and tetra-repeats are shown in Figure 2.6, hexa- and penta-nucleotides are shown in *Appendix 2.1* and *Appendix 2.2*. The length of each microsatellite repeat array was analysed for di-, tri- and tetra-repeats (Figure 2.7). There was a common trend that most of the microsatellites were  $\leq 30$  bp, with longer repeat arrays decreasing in frequency.

### **Dinucleotide repeats**

The most abundant dinucleotide repeat in *O. chilensis* was (AT)<sub>n</sub> (Figure 2.6), the other repeats have a similar abundance but there were very few repeats of (CG)<sub>n</sub>. The mean length of dinucleotide repeats was 12 bp with a range of 378 bp.

### **Trinucleotide repeats**

By far the most common tri-nucleotide repeat in *O. chilensis* was AAT (Figure 2.6). The mean length of trinucleotide repeats was 12 bp with a range of 512 bp.

### **Tetranucleotide repeats**

ACGT was the most common tetranucleotide repeat in *O. chilensis* (Figure 2.6), which seems unusual given that repeats containing G seem rare in the other repeat motifs. The mean length of tetranucleotide repeats was 12 bp with a range of 360 bp.

### **Hexanucleotide repeats**

The most common repeats were AACGTT, AAACGT, ACACAT, AACGGT, AACCGT, AACCTT, AAAAAC, AAAAAT. The mean length of hexanucleotide repeats was 15 bp with a range of 285 bp.

### **Pentanucleotide repeats**

The most common repeats were AACGT, ACCGT, ACTAT, AACTT, ACACT, AAACCT, AAAGT, ACTCT, AAACG. The mean length of pentanucleotide repeats was 12 bp with a range of 326 bp.

### **2.3.3 Identification of polymorphic microsatellite loci**

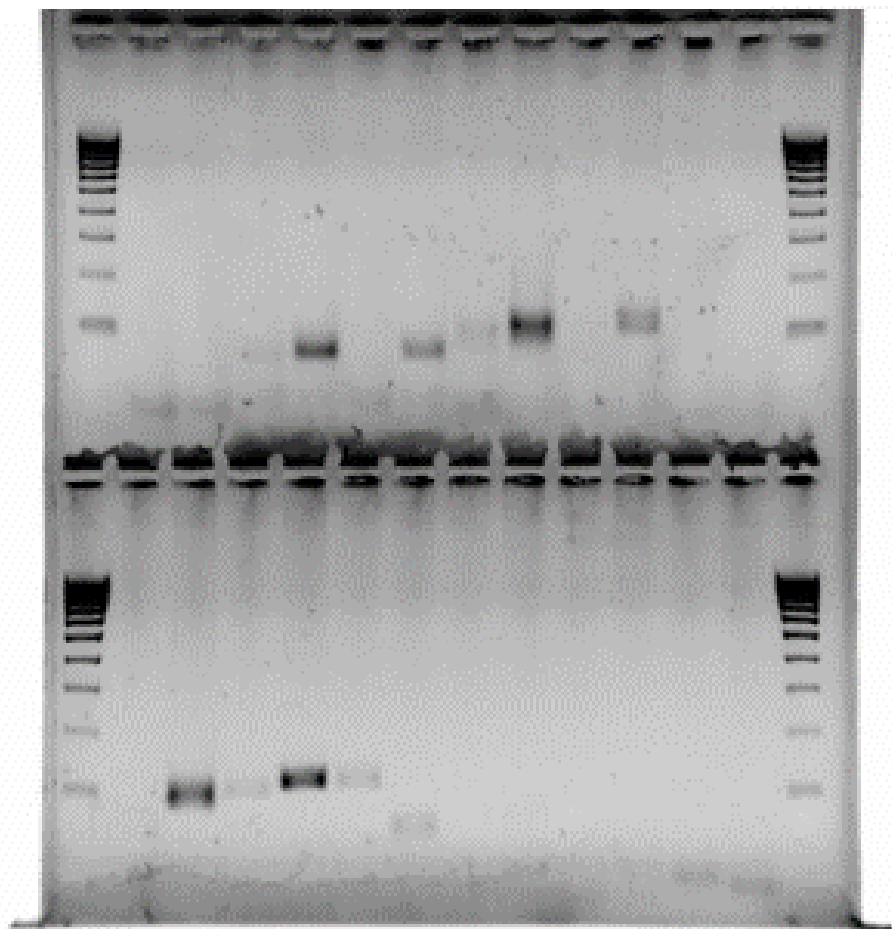
The analysis using the software programme iQDD designed primers for 223 microsatellite loci (data not shown). Upon visual inspection of the primer binding sites there were many instances of suggested primers that were not optimal. For example, there were found to be in some cases, multiple microsatellite regions within the primer binding site. In other cases the primer annealed to part of the microsatellite region. Visual inspection of the suggested primers was an essential step when using iQDD due to problems with primer binding location.

In total, 125 primers were tested for consistent PCR amplification. Of these, 36 were shown to consistently amplify a single clear band of DNA around the expected size (PCR primers for these loci are displayed in Table 2.3). Figures 2.4a to 2.4j show the outcome of a touchdown PCR with an initial annealing temperature of 62.5°C. It was found that if the primers did not produce a band of the expected size, then further optimisation of the PCR protocol would not improve the result; this included running a gradient PCR from 65°C to 50°C. Of the 36 primers, four primers were shown to amplify a product that was possible to genotype. The reasons that most of the amplified products were unsuccessfully genotyped were due to no product being found in the expected size range, excessive stutter which made genotyping impossible to score, or monomorphic alleles either as a homozygote or the same heterozygote.

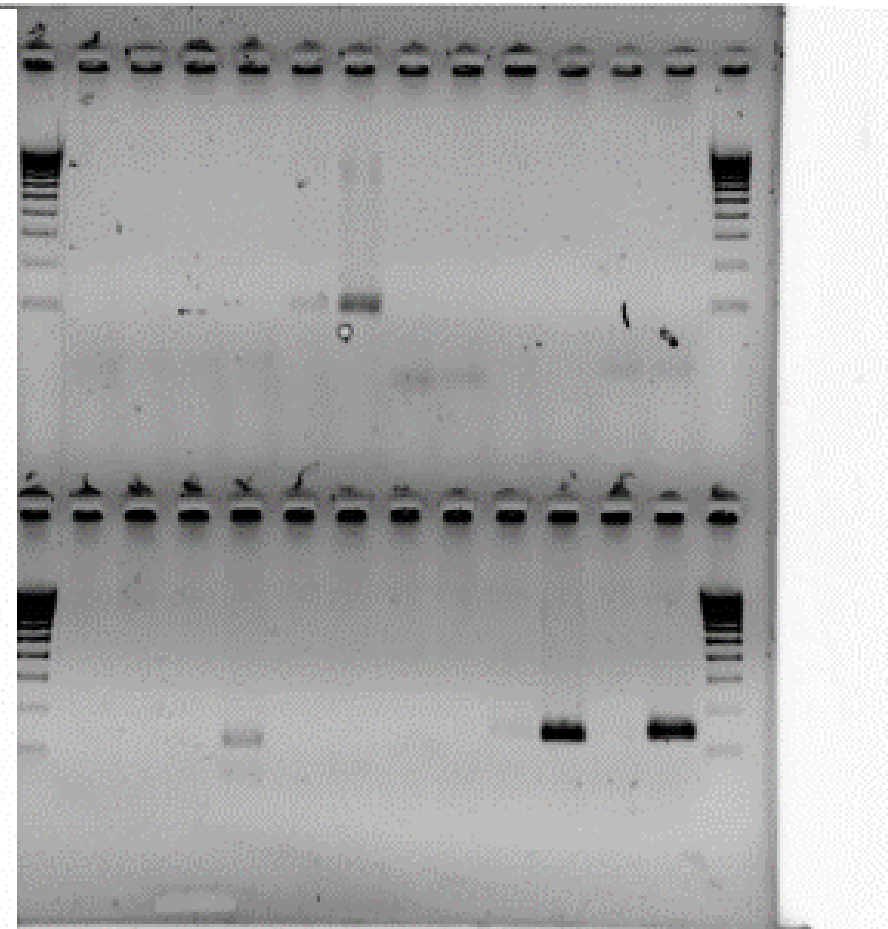
Four primer sets were developed for loci that give interpretable and consistent results. Summary statistics for these polymorphic loci are displayed in Table 2.4 Loci *Ost\_chi\_074* and *Ost\_chi\_084* were not in HWE. This is due to an excess of heterozygotes in *Ost\_chi\_084* and an excess of homozygotes in *Ost\_chi\_074*. However, these results should be interpreted with caution since the sample size was less than 50 individuals (Hedrick, 2000).

#### **2.3.4 Search for orthologous repetitive DNA in various oyster species**

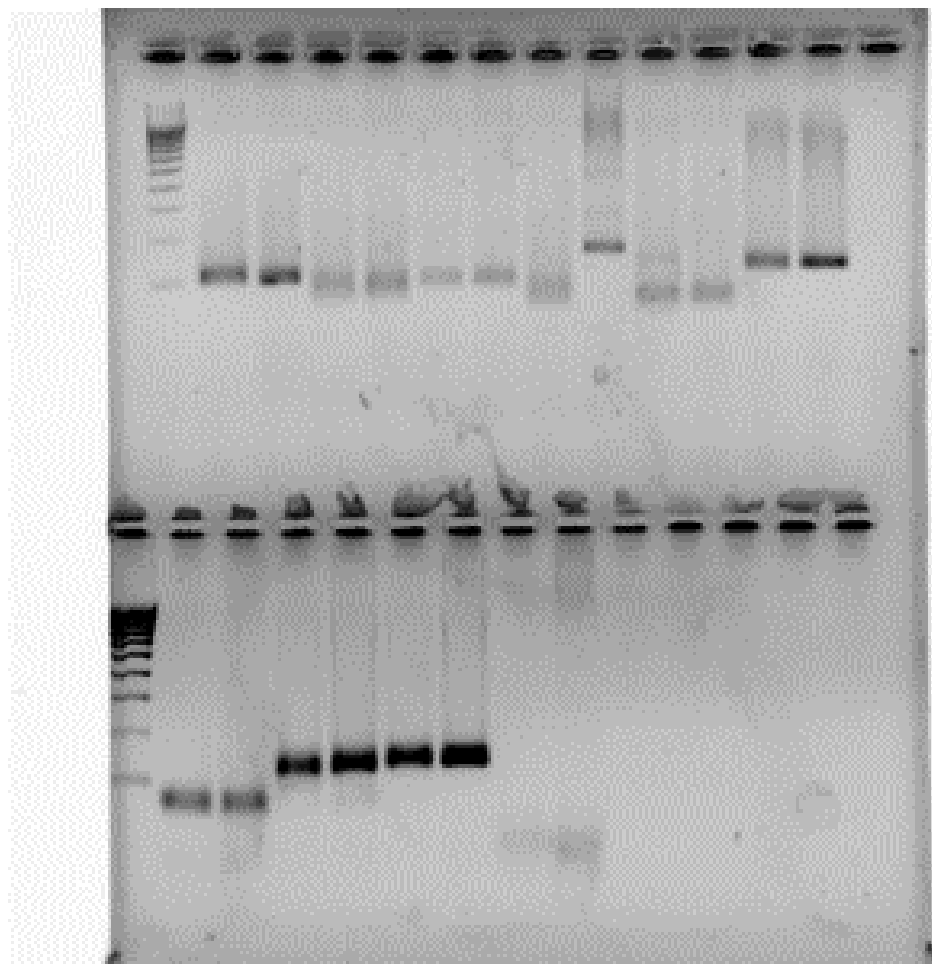
Results of the BLAST search of the 454 DNA database against all oyster repetitive sequences are displayed in Table 2.5a to 2.5e. Many of the blast hits were downstream or upstream of the microsatellite repeat region so in many cases it was impossible to compare the repeat unit of the microsatellite to the database sequences, but given the very low e-values overall it is unlikely that these hits were matched by chance and they are likely to reflect conserved sequence regions which are homologous between species. It should also be noted that the ‘best’ hits were restricted to single matches, therefore looking for more than one match with potentially lower e-values could reveal more coverage of the microsatellite regions and hence allow the comparison of repeat motifs. There was however one exception, where the query and reference sequences overlapped to show a common microsatellite loci (Figure 2.8). The figure illustrates an alignment between a sequence from the 454 data set and the microsatellite region *Cvi2i20.fa* from *Crassostrea virginica*. Here it is possible to see the BLAST hit is in the conserved flanking region and the query and reference sequences overlap to show a common microsatellite motif, in *O. chilensis* the repeat appears to have become a compound repeat.



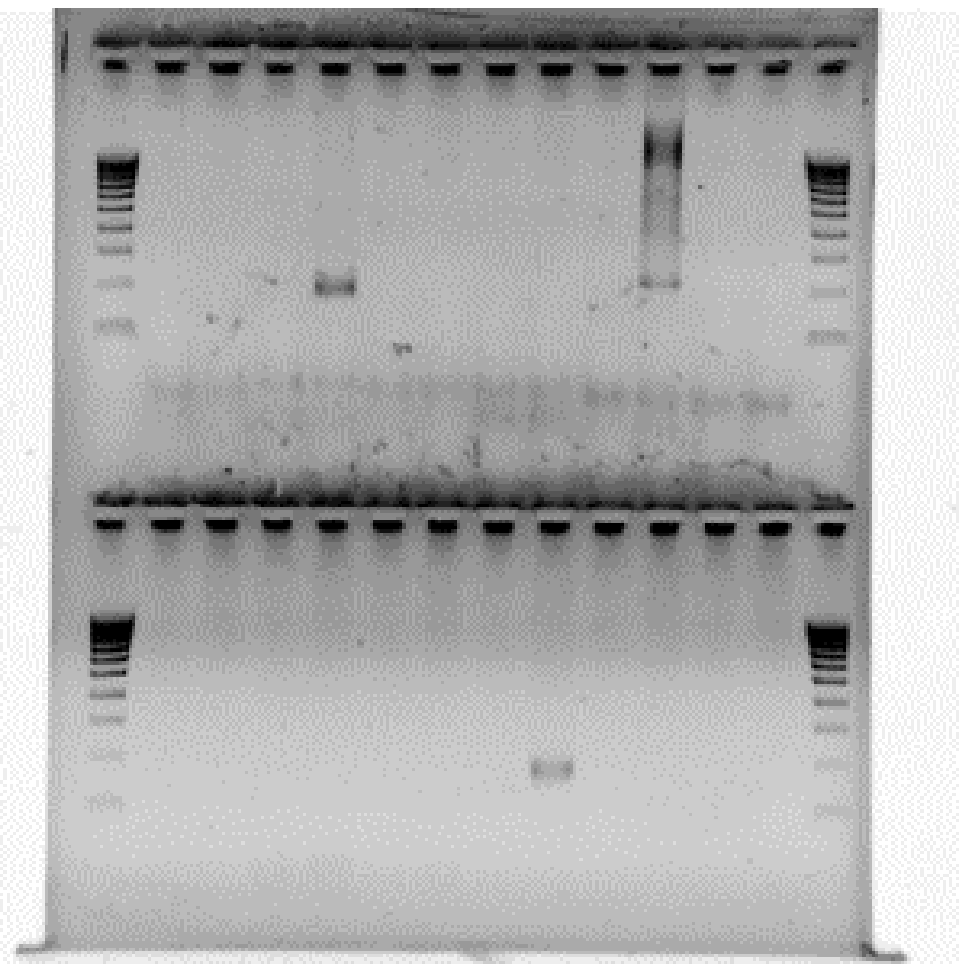
**Figure 2. 4a** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_001; Lanes 4 and 5: Ost\_chi 001.1, Lanes 6 and 7: Ost\_chi 001.9, Lanes 8 and 9 Ost\_chi 001.5, Lanes 10 and 11 Ost\_chi 002, Lanes 13 and 14 Ost\_chi 002.3, Lanes 16 and 17 Ost\_chi 002.4, Lanes 18 and 19 Ost\_chi 003.1, Lanes 20 and 21 Ost\_chi 004, Lanes 22 and 23 Ost\_chi 004.2, Lanes 24 and 25 Ost\_chi 006, Lanes 26 and 27



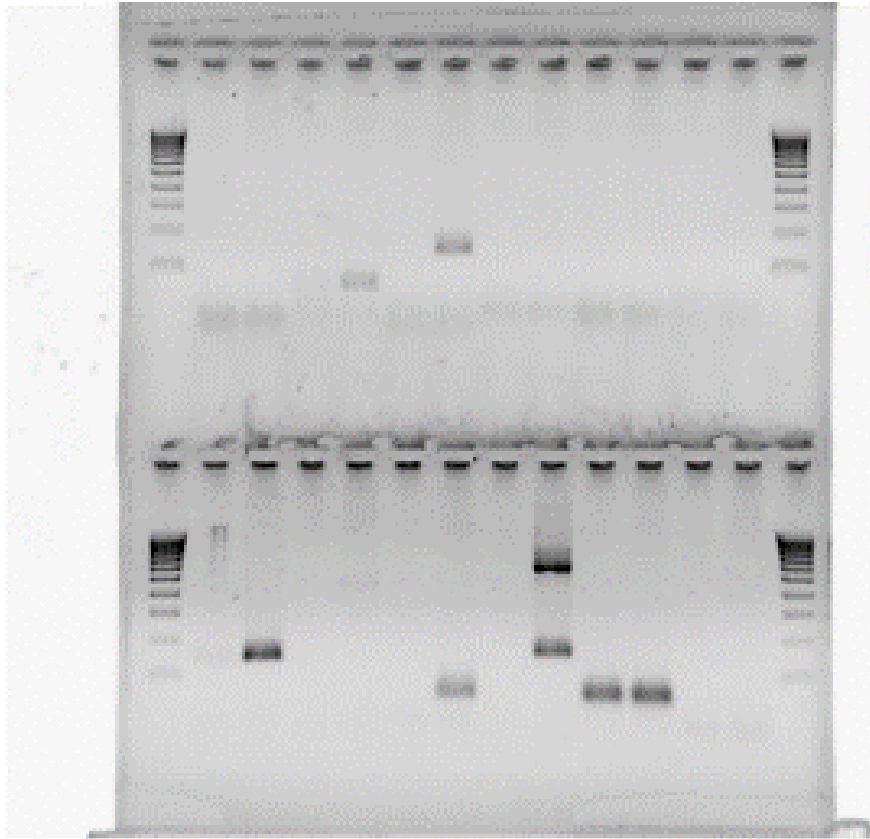
**Figure 2.4b** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_006.2; Lanes 4 and 5: Ost\_chi 007.1, Lanes 6 and 7: Ost\_chi 007.2, Lanes 8 and 9 Ost\_chi 008, Lanes 10 and 11 Ost\_chi 009, Lanes 13 and 14 Ost\_chi 010, Lanes 16 and 17 Ost\_chi 011, Lanes 18 and 19 Ost\_chi 012, Lanes 20 and 21 Ost\_chi 024, Lanes 22 and 23 Ost\_chi 025, Lanes 24 and 25 Ost\_chi 026 Lanes 26 and 27 Ost\_chi 028



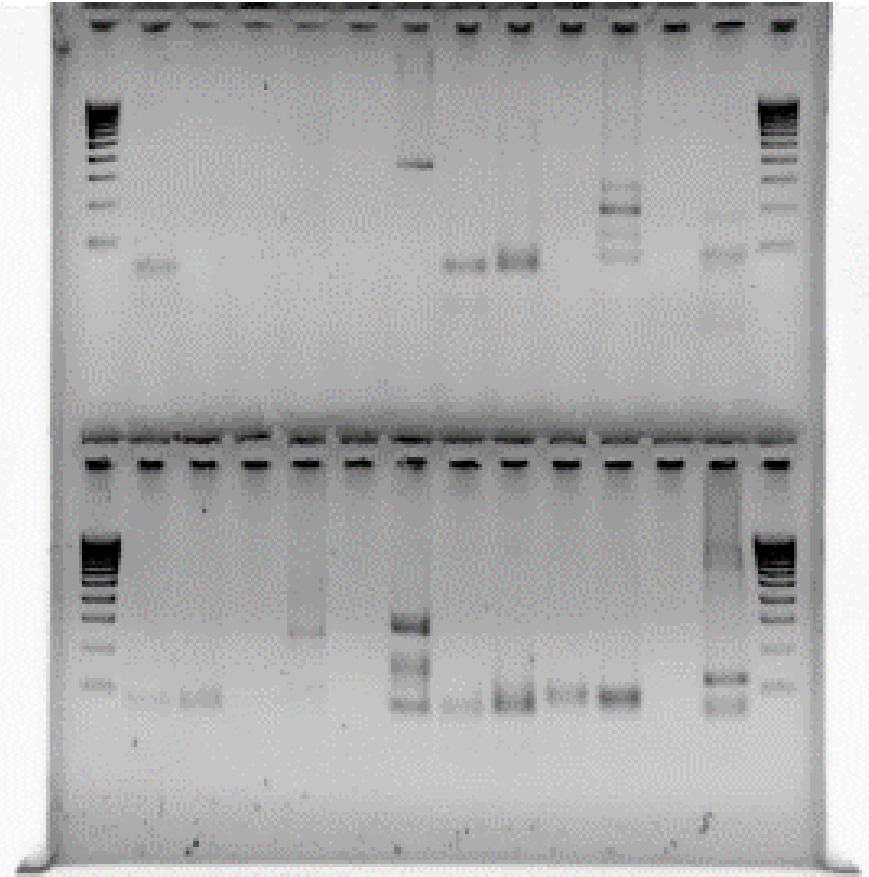
**Figure 2.4c** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_013.1; Lanes 4 and 5: Ost\_chi 015.2, Lanes 6 and 7: Ost\_chi 016, Lanes 8 and 9 Ost\_chi 017, Lanes 10 and 11 Ost\_chi 017.3, Lanes 13 and 14 Ost\_chi 018.4, Lanes 16 and 17 Ost\_chi 019, Lanes 18 and 19 Ost\_chi 022, Lanes 20 and 21 Ost\_chi 022.1, Lanes 22 and 23 Ost\_chi 023,



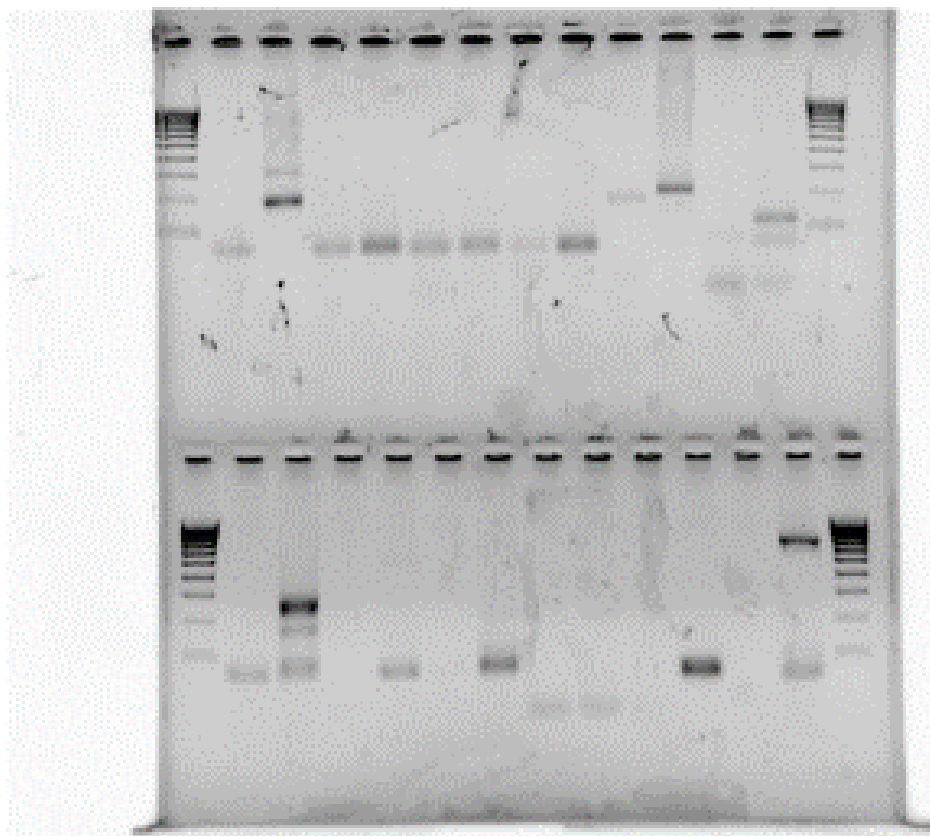
**Figure 2.4d** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_029; Lanes 4 and 5: Ost\_chi 030.1, Lanes 6 and 7: Ost\_chi 033, Lanes 8 and 9 Ost\_chi 035, Lanes 10 and 11 Ost\_chi 037, Lanes 13 and 14 Ost\_chi 038, Lanes 16 and 17 Ost\_chi 039, Lanes 18 and 19 Ost\_chi 042, Lanes 20 and 21 Ost\_chi 042.1, Lanes 22 and 23 Ost\_chi 043.1, Lanes 24 and 25 Ost\_chi 044.1 Lanes 26 and 27 Ost\_chi 043



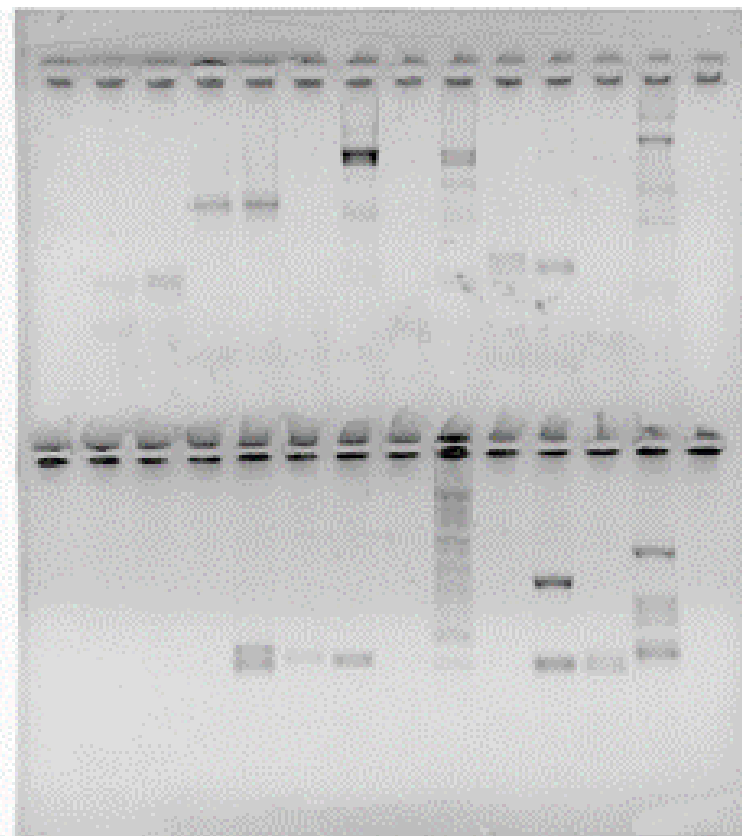
**Figure 2.4e** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_045; Lanes 4 and 5: Ost\_chi 045.1, Lanes 6 and 7: Ost\_chi 047.1, Lanes 8 and 9 Ost\_chi 049.1, Lanes 10 and 11 Ost\_chi 050, Lanes 13 and 14 Ost\_chi 051, Lanes 16 and 17 Ost\_chi 051.1, Lanes 18 and 19 Ost\_chi 053, Lanes 20 and 21 Ost\_chi 053.1, Lanes 22 and 23 Ost\_chi 053.2, Lanes 24 and 25 Ost\_chi 058



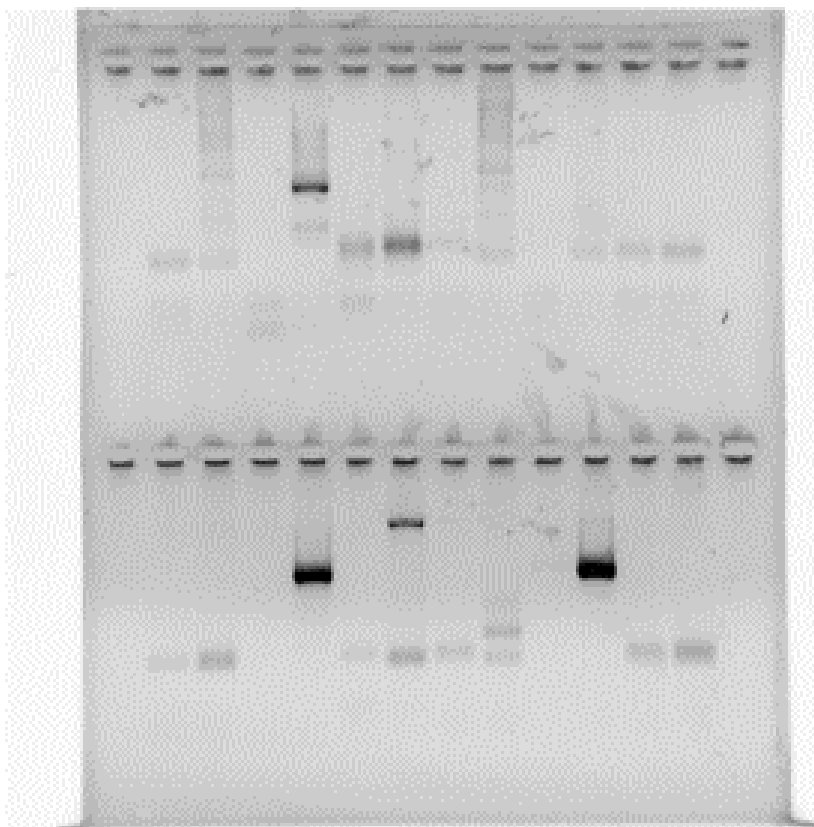
**Figure 2.4f** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_060; Lanes 4 and 5: Ost\_chi 061, Lanes 6 and 7: Ost\_chi 062, Lanes 8 and 9 Ost\_chi 063, Lanes 10 and 11 Ost\_chi 064.1, Lanes 13 and 14 Ost\_chi 064.2, Lanes 16 and 17 Ost\_chi 65, Lanes 18 and 19 Ost\_chi 065.1, Lanes 20 and 21 Ost\_chi 066, Lanes 22 and 23 Ost\_chi 066.3, Lanes 24 and 25 Ost\_chi 067, Lanes 26 and 27 Ost\_chi 068



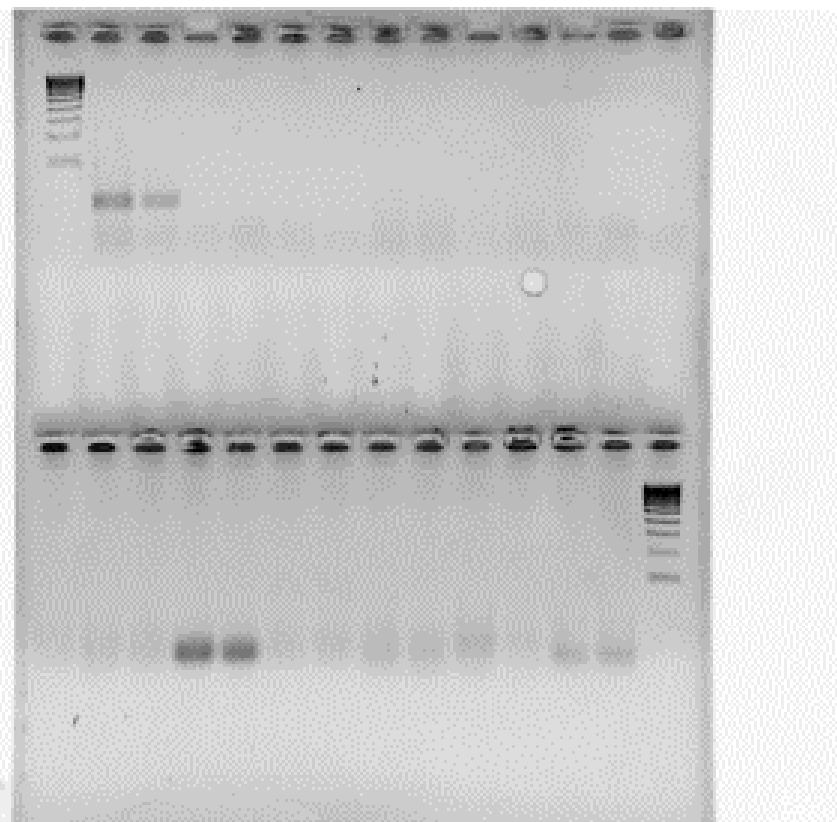
**Figure 2.4g** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_068.1; Lanes 4 and 5: Ost\_chi 068.3, Lanes 6 and 7: Ost\_chi 070, Lanes 8 and 9 Ost\_chi 070.1, Lanes 10 and 11 Ost\_chi 074, Lanes 13 and 14 Ost\_chi 075, Lanes 16 and 17 Ost\_chi 076.1, Lanes 18 and 19 Ost\_chi 077, Lanes 20 and 21 Ost\_chi 078, Lanes 22 and 23 Ost\_chi 079, Lanes 24 and 25 Ost\_chi 080, Lanes 26 and 27 Ost\_chi 082



**Figure 2.4h** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_083; Lanes 4 and 5: Ost\_chi 084, Lanes 6 and 7: Ost\_chi 087, Lanes 8 and 9 Ost\_chi 087.1, Lanes 10 and 11 Ost\_chi 087.2, Lanes 13 and 14 Ost\_chi 088.1, Lanes 16 and 17 Ost\_chi 089.2, Lanes 18 and 19 Ost\_chi 084.3, Lanes 20 and 21 Ost\_chi 090, Lanes 22 and 23 Ost\_chi 091, Lanes 24 and 25 Ost\_chi 092, Lanes 26 and 27 Ost\_chi 095



**Figure 2.4i** Electrophoresis gel. Lanes 1, 14, 15, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_096; Lanes 4 and 5: Ost\_chi 096.1, Lanes 6 and 7: Ost\_chi 097.1, Lanes 8 and 9 Ost\_chi 103, Lanes 10 and 11 Ost\_chi 104, Lanes 13 and 14 Ost\_chi 104.1, Lanes 16 and 17 Ost\_chi 104.2, Lanes 18 and 19 Ost\_chi 105, Lanes 20 and 21 Ost\_chi 106, Lanes 22 and 23 Ost\_chi 109, Lanes 24 and 25 Ost\_chi 110, Lanes 26 and 27 Ost\_chi 111



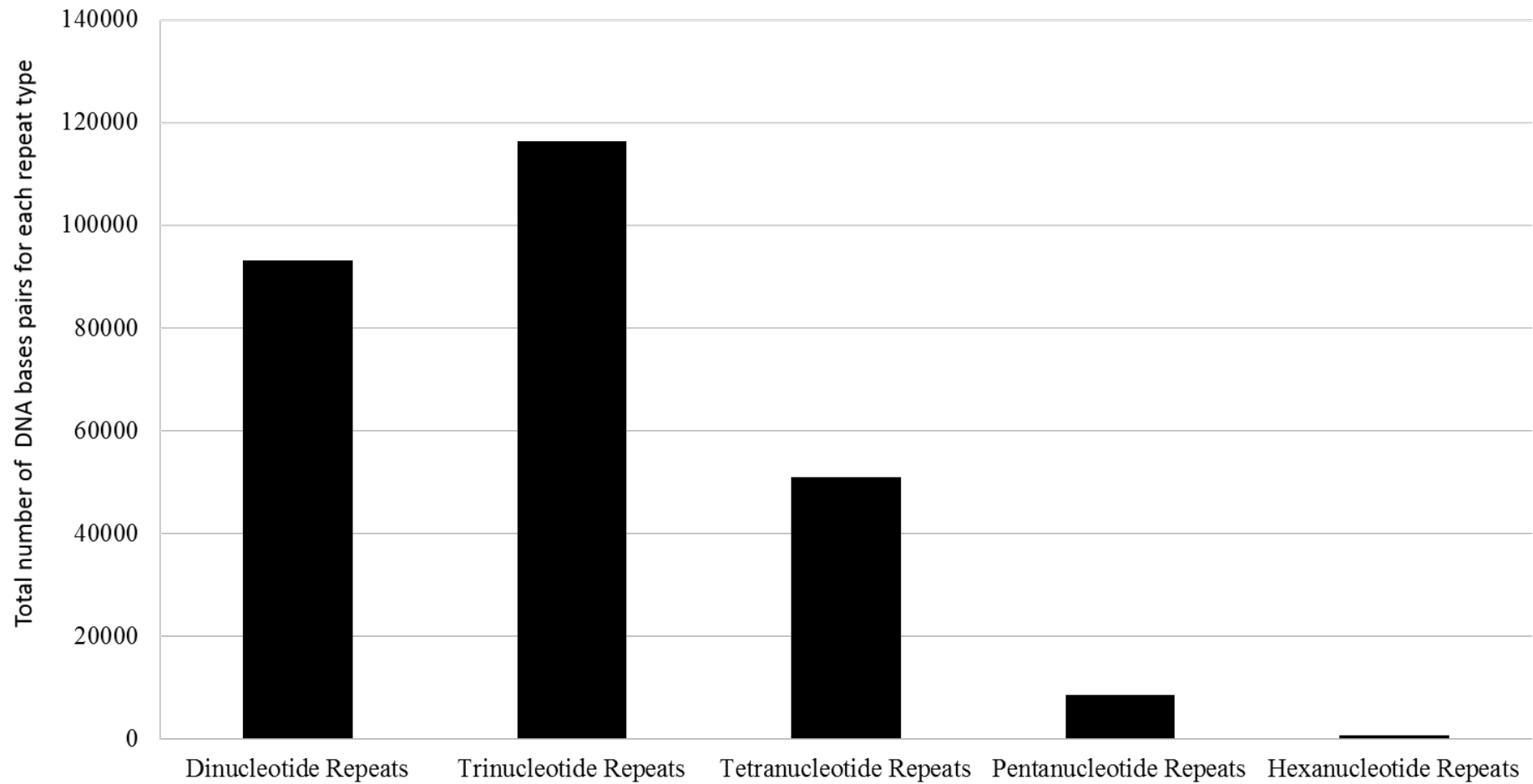
**Figure 2.4j** Electrophoresis gel. Lanes 1, 28 Hyper ladder IV. Lanes 2 and 3: Ost\_chi\_113; Lanes 4 and 5: Ost\_chi 114, Lanes 6 and 7: Ost\_chi 116, Lanes 8 and 9 Ost\_chi 117, Lanes 10 and 11 Ost\_chi 118, Lanes 12 and 13 Ost\_chi 118.1, Lanes 14 and 15 Ost\_chi 119, Lanes 16 and 17 Ost\_chi 120,

**Table 2.3:** Primers that produced a clear band in the expected size region

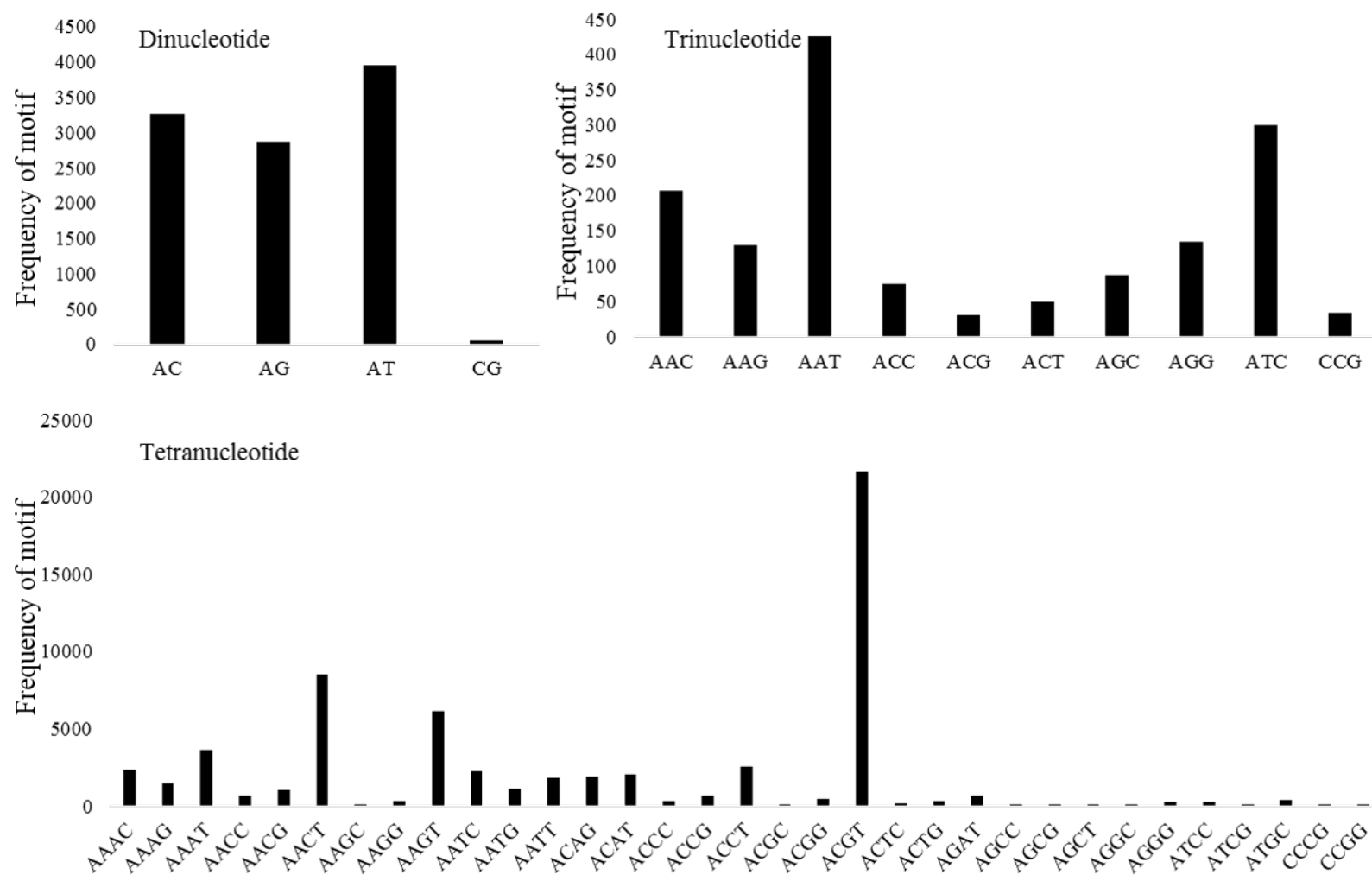
Primer	5' to 3' Forward Primer Sequence	5' to 3' Primer Sequence Reverse	Repeat Motif	Estimated product size (bp)
<i>Ost_chi_001.1</i>	CTTTGGTCGAAATTGGCCT	GCCGACCACTCTACGAGTAA	(GA) <sub>9</sub>	~ 60
<i>Ost_chi_001.5</i>	TACAGTGGCGGATTTAAGCG	AAACAAGAGACCACGATTACCTT	(CACC) <sub>5</sub>	~95
<i>Ost_chi_001.9</i>	AGTAGGTAACCTTAAGTACGAGTAACCG	CGTTACGTTTCGTTACGGTTCG	(AACGT) <sub>5</sub>	~ 60
<i>Ost_chi_002</i>	GGCTGCAAAATTTGGGTCTA	CGGATGAGACCGAAATAACC	(AG) <sub>5</sub>	~100
<i>Ost_chi_002.4</i>	AGACATTACTAGAAGGTCCACAGAA	GACATTATCATTTTGAAAGTTGTCTT	(TGAT) <sub>9</sub>	~100
<i>Ost_chi_003.1</i>	AGCCTCCAAATAGGTCACCA	ATTGTGAGGGGACTCATGGA	(ATC) <sub>9</sub>	~110
<i>Ost_chi_007.2</i>	CCCACCATATCCCCATCTTA	GGTAGTACAACAAGATGACGAACAA	(TA) <sub>13</sub>	~100
<i>Ost_chi_013.1</i>	TTGGAATATTCGGCATTGTG	TCCAGAAAAGAAAGACTGAATGA	(ATCT) <sub>6</sub>	~119
<i>Ost_chi_015.2</i>	CGACGAACTCGAAATAAAGGA	TTGTCAGTCTGCCATTAGAGTCA	(AT) <sub>11</sub>	~120
<i>Ost_chi_016</i>	GGATCGCCTCCAACCTACAGA	TTCGTTTGGTTTGGTTTACG	(AT) <sub>18</sub>	~120
<i>Ost_chi_017</i>	AAAAGGAGCCTCATGGTGAA	ATAGTAACGTACACCGCCGC	(AT) <sub>20</sub>	~150
<i>Ost_chi_017.3</i>	AAACGTACGGACCAACAACC	CTCGACTTCACTGTACTGTCTCG	(AG) <sub>12</sub>	~100
<i>Ost_chi_018.4</i>	GCAAGAACTTGTAACCAATCCC	AGAGCCAATTTTCGTGTTTGA	(TTCAA) <sub>5</sub>	~140
<i>Ost_chi_019</i>	ACCTGACGTCGGTTCTCAGT	CCCGTCGCTGAGTTAGTAGG	(AC) <sub>17</sub>	~100
<i>Ost_chi_022</i>	CTCGGGCTGTTTTTCATGTTT	AACCCAATCGAAACAACACC	(GA) <sub>10</sub>	~150
<i>Ost_chi_023</i>	ACCCGTCGTAACCTCGTAAC	TCCGTACTCGTACTACCGGTCT	(GTAC) <sub>5</sub>	~150
<i>Ost_chi_024</i>	AAATAAGTTCACCCGACCCC	TTACGTTTCGGTTACGGTTACG	(AAGT) <sub>5</sub>	~110
<i>Ost_chi_026</i>	ACACCCTCGACATGAGGATT	GAAGACCACATTCGAGTCATTG	(TGA) <sub>8</sub>	~130

**Table 2.3(continued):** Primers that produced a clear band in the expected size region

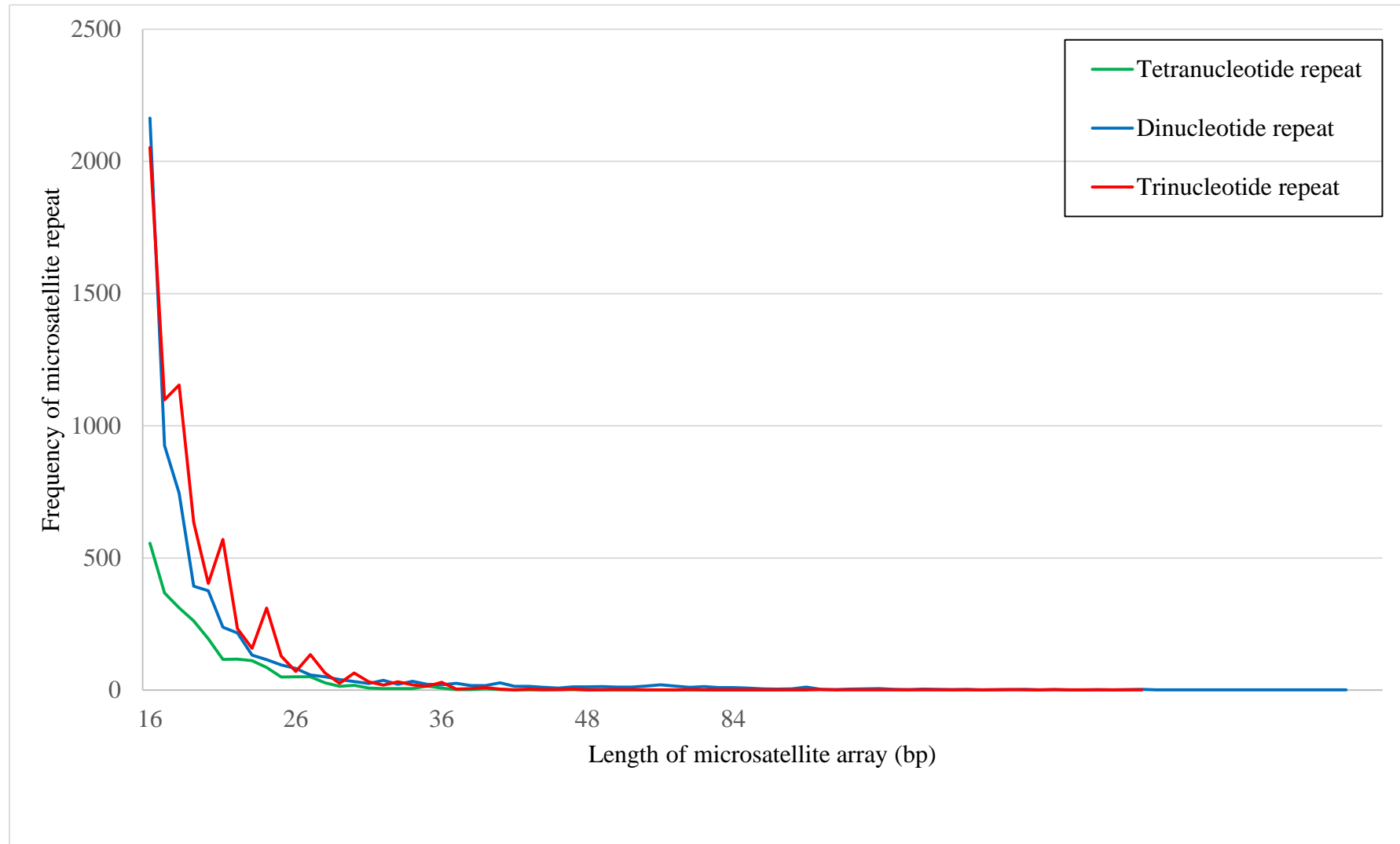
Primer Name	5' to 3' Forward Primer Sequence	5' to 3' Primer Sequence Reverse	Repeat Motif	Estimated product size (bp)
<i>Ost_chi_028</i>	GGGCCAATTGTTGATTGTTC	AGGCAAAGGTCTCAGGGATT	(GGT) <sub>6</sub>	~130
<i>Ost_chi_030.1</i>	TGAAACGACAATCTCAAAACC	CGTTCGTTAGTGTGTTGTGTTAG	(TA) <sub>24</sub>	~180
<i>Ost_chi_037</i>	GTTTGGACGAAGGAGCAATC	CCCAAAGGTTTCATTACATTTC	(AAAC) <sub>5</sub>	~190
<i>Ost_chi_043.1</i>	AGGCCAATATACGCGAAGTC	TCGTGTGCCTTGTAGAGACG	(GA) <sub>10</sub>	~150
<i>Ost_chi_047.1</i>	TTCAGTTTGCATATAAACTCGTTG	CGGTAGGTAGTAAGAACGTAGTAGTG	(TA) <sub>13</sub>	~140
<i>Ost_chi_051.1</i>	TGGAAAAGAGAAAAGGTCATTG	TTACGTTTCCTTCTTCGTACCAT	(TA) <sub>14</sub>	~140
<i>Ost_chi_053.1</i>	GCCAACAGACTAACC GGAGT	GGTAAAGTTTTAAGTTCGTGTCTGTC	(GA) <sub>23</sub>	~150
<i>Ost_chi_068.1</i>	TTTGGTACTCTCCGTCCGTC	GTATCCTCCTGTCCCAACCA	(CTGT) <sub>5</sub>	~160
<i>Ost_chi_074</i>	GCCGGTGATATCATCTTCGT	AGGGTGCAGTCAACGTATCC	(GCG) <sub>9</sub>	~200
<i>Ost_chi_075</i>	TGCGTACTCGTCTTGGATTG	CCTATTGTACGGAGTACTACTAAACCT	(AATC) <sub>9</sub>	~200
<i>Ost_chi_076.1</i>	CACTGAGCTAAGGGGTAAATCC	AGCCGAATTCACTATGTGCC	(TA) <sub>12</sub>	~200
<i>Ost_chi_084</i>	CGTGATGCGTATCGAAAGAA	TCTTGTCATGGCTATCGTGC	(AT) <sub>8</sub>	~200
<i>Ost_chi_087</i>	TCACACTTCGTCTCACCCTG	CTAATGACCCGATGGTGCTT	(AG) <sub>17</sub>	~200
<i>Ost_chi_096.1</i>	TTGCCAAAACGCTTA ACTGA	AGGACTCCACATGGAGACATAA	(TTG) <sub>10</sub>	~200
<i>Ost_chi_105</i>	AAGCCTTATTAGGTCAAAAGGTCA	CCAGAGAATCAACAACACATTTTC	(TA) <sub>14</sub>	~250
<i>Ost_chi_110</i>	AATTTAAGATGCAATCTTGTAATGTT	GTCCCCACCTTCCATTGAC	(CATG) <sub>6</sub>	~250
<i>Ost_chi_113</i>	CGACGTACGAAACGGTTAGG	TTACGGGTAAACGGTAACGGA	(TAC) <sub>7</sub>	~250
<i>Ost_chi_120</i>	CCTTGGCCAGAGGTTAATTG	TTACGTTACGTTAGTTTAGTTTCGTTT	(AG) <sub>26</sub>	~300



**Figure 2.5:** Contribution of each repeat type to genomic DNA



**Figure 2.6:** Frequency of repeat motif of each microsatellite type



**Figure 2.7:** Length distributions of the microsatellite arrays

**Table 2.4:** Characterisation of microsatellite loci

<i>Primer Pair</i>	<i>n</i>	<i>Allele size range</i>	<i>Na</i>	<i>Ho</i>	<i>He</i>	<i>Fis</i>	<i>pHWE</i>
<i>Ost_chi_013.1</i>	37	104-124	4	0.486	0.497	0.021	0
<i>Ost_chi_074</i>	15	206 - 218	3	0.188	0.588	0.681	0.005
<i>Ost_chi_076.1</i>	16	222-241	4	0.5	0.572	0.126	0.306
<i>Ost_chi_084</i>	30	230-234	3	0.6	0.455	-0.348	0.675

**Table 2.5a:** BLAST hits with published *Ostrea edulis* repetitive DNA on GenBank

<i>454 Sequence</i>	<i>e-value</i>	<i>Hit Accession</i>	<i>Sequence information</i>
GQ6D37002JSVIY	5.44E-36	AF274886	Microsatellite <i>Oe3/37</i>
GQ6D37002JR5Y7	1.40E-55	FJ236829	Microsatellite <i>Oed315</i>
GQ6D37002JIVW8	3.40E-53	AF190985	Microsatellite <i>Oe1/10</i>
GQ6D37002HJRD2	2.81E-91	AF297864	Microsatellite <i>HA21</i>
GQ6D37002FBTPU	1.95E-41	FJ236824	Microsatellite <i>Oed243</i>
GQ6D37002FHVWN	7.03E-24	AF297862	Microsatellite <i>HA7</i>
GQ6D37002FR8MK	2.75E-49	FJ236825	Microsatellite <i>Oed258</i>
GQ6D37002G2DIF	1.85E-85	AF297863	Microsatellite <i>HA10</i>
GQ6D370021C6PR	1.34E-21	FJ236819	Microsatellite <i>Oed202</i>
GQ6D37002GY4NP	1.06E-97	AF190986	Microsatellite <i>Oe1/21</i>
GQ6D3700216UIE	1.32E-50	FJ236822	Microsatellite <i>Oed234</i>
GQ6D3700212BVY	3.29E-27	FJ236816	Microsatellite <i>Oed180</i>
GQ6D37002ILUPB	1.36E-21	AF297865	Microsatellite <i>HA11a</i>
GQ6D37002IVCIS	2.79E-31	AF310010	Microsatellite <i>OeduH15</i>
GQ6D37002J2M7H	1.55E-63	AJ864934	<i>BclII</i> repeat element
GQ6D37002ILYOA	6.46E-48	AJ864931	<i>BclII</i> repeat element
GQ6D37002IHAUN	4.06E-42	AJ864933	<i>BclII</i> repeat element
GQ6D37002IB2DA	9.57E-69	AJ864927	<i>BclII</i> repeat element
GQ6D37002I4UZJ	1.32E-56	AJ864929	<i>BclII</i> repeat element
GQ6D37002FNKDL	9.57E-69	AJ864930	<i>BclII</i> repeat element
GQ6D37002FNKDL	3.22E-72	AJ864928	<i>BclII</i> repeat element
GQ6D37002F7LDL	8.44E-62	AJ864932	<i>BclII</i> repeat element

**Table 2.5b:** BLAST hits with published *Ostrea conchaphila* repetitive DNA on GenBank

454 Sequence	e-value	Hit Accession	Sequence information
GQ6D37002F8V9W	1.20E-35	EU587420	A107 microsatellite sequence
GQ6D37002FWYOP	3.33E-94	EU587488	D130 microsatellite sequence
GQ6D37002G86QI	2.42E-52	EU587462	C130 microsatellite sequence
GQ6D37002GPZXC	8.84E-41	EU587444	C102 microsatellite sequence
GQ6D37002HHCC2	1.70E-59	EU587483	D123 microsatellite sequence
GQ6D37002J08NF	1.84E-174	EU587454	C117 microsatellite sequence
GQ6D37002JCZK0	3.52E-24	EU587485	D126 microsatellite sequence
GQ6D37002JHQWV	1.85E-107	EU587410	A5 microsatellite sequence
GQ6D37002JIJKU	2.31E-106	EU587477	D115 microsatellite sequence
GQ6D37002JIJKU	2.02E-92	EU587405	D104 microsatellite sequence
GQ6D37002JUJI9	1.25E-65	EU587459	C124 microsatellite sequence
GQ6D37002JZOLO	3.65E-28	EU587478	D117 microsatellite sequence

**Table 2.5c:** BLAST hits with published *Crassostrea gigas* repetitive DNA on GenBank

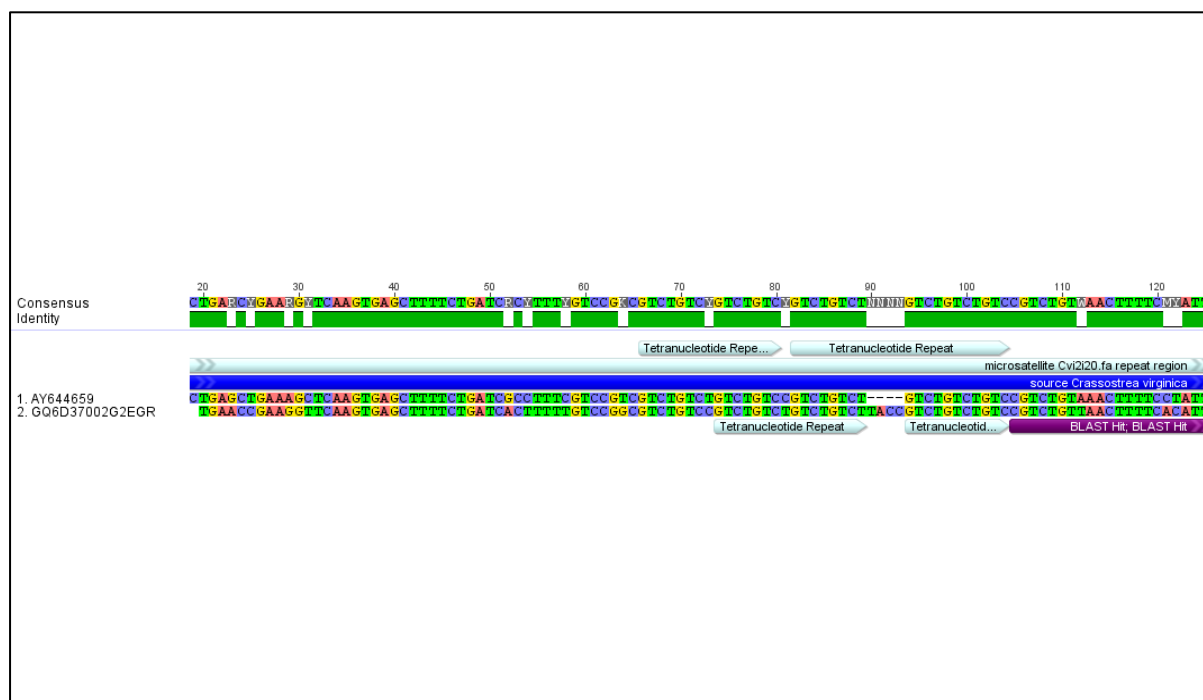
454 Sequence	e-value	Hit Accession	Sequence information
GQ6D37002FNKDL	2.61E-66	AJ864912	BclI repeat element
GQ6D37002FNUKH	7.00E-61	AJ864902	BclI repeat element
GQ6D37002IM7ST	1.50E-37	AF051185	BZ72x microsatellite
GQ6D37002JJ1BS	1.04E-45	AJ864913	BclI repeat element

**Table 2.5d:** BLAST hits with published *Crassostrea virginica* repetitive DNA on GenBank

454 Sequence	e-value	Hit Accession	Sequence information
GQ6D37002IBJDZ	5.99E-54	AY644659	Cvi2i20,fa microsatellite
GQ6D37002G2EGR	3.32E-24	AF276251	Cvi-10 microsatellite
GQ6D37002ILGW1	2.45E-58	AJ864923	BclI repeat element
GQ6D37002ILGW1	2.45E-58	AJ864924	BclI repeat element
GQ6D37002ILGW1	2.45E-58	AJ864925	BclI repeat element
GQ6D37002ILGW1	2.45E-58	AJ864926	BclI repeat element

**Table 2.5e** BLAST hits with published *Crassostrea hongkongensis* repetitive DNA on GenBank

454 Sequence	e-value	Hit Accession	Sequence information
GQ6D37002HJLEB	1.46E-71	GQ925415	CHK62 microsatellite sequence
GQ6D37002HJLEB	1.20E-68	GU952821	Ch301 microsatellite sequence
GQ6D37002HJLEB	2.29E-70	GU952836	Ch316 microsatellite sequence
GQ6D37002JCC5O	5.54E-71	GU952824	Ch304 microsatellite sequence
GQ6D37002HJLEB	1.01E-72	GQ925415	Chk62 microsatellite sequence
GQ6D37002HJLEB	2.95E-47	GU952836	Ch316 microsatellite sequence
GQ6D37002HJLEB	1.09E-72	HM461239	Ch306 microsatellite sequence
GQ6D37002HJLEB	1.77E-33	HM461241	Ch308 microsatellite sequence
GQ6D37002HJLEB	9.13E-36	HM461247	Ch314 microsatellite sequence

**Figure 2.8:** Alignment of a potentially orthologous microsatellite repeat shared between *Ostrea chilensis* (GQ6D37002G2EGR) and *Crassostrea virginica* (AY644659).

## 2. 4 Discussion

An abundance of microsatellite regions has been found in the genome of *O. chilensis*. The microsatellite loci were further examined to describe patterns, with an aim to describe mutational processes that are generating the microsatellites. Unexpectedly, tri- nucleotide repeat motifs were the most common (Figure 2.5). In coding regions tri-repeats are found to be the most common repeat type in most taxonomic levels (Toth *et al.* 2000). The reason is because an addition or subtraction of a tri-repeat would not cause a change in the reading frame- meaning it is likely that it would not be under selection (although some diseases are caused by tri-nucleotide repeat expansions that could come under selection pressure). However the 454 sequencing was obtained from genomic DNA which includes both exons and introns. It is expected that the most common repeat in genomic DNA are dinucleotide repeats. The 454 sequencing could be biased towards exons or *O. chilensis* is un-usual in its microsatellite abundance. Microsatellites are thought to be generated by slippage of the DNA polymerase enzyme during replication. It may be possible that the polymerase enzyme may slip more often on different repeat types. However, in this study there seems to be no difference between the mean lengths of each repeat type - the mean length seems to be constrained to around 12 bp to 15 bp for all repeat types (Figure 2.7). This suggests that there could be a limit to the size of the microsatellite repeat length. Moreover, there was a clear pattern in the length of microsatellites with very few microsatellite regions being any longer than 30 bp. Some studies have suggested there is a general trend of increasing expansion of the length of the microsatellite repeat (Chambers and MacAvoy, 2000). However, there is clearly a limit to the length as is demonstrated in the present study. The reason for this upper size limit is thought to be evolution of point mutations within the microsatellite region, which can result in polymerase stability around the repeat sequence area (Kruglyak *et al.* 1998) this seems to be the case in Figure 2.8 (this is discussed later in the chapter).

(AC)<sub>n</sub>, (AT)<sub>n</sub> and (AG)<sub>n</sub> are more common than (CG)<sub>n</sub> repeats (Figure 2.6). The (CG)<sub>n</sub> repeat has been found to be rare in many other studies ((Toth *et al.* 2000, Subramanian *et al.* 2002, Katti *et al.* 2001, Karaoglu *et al.* 2005, Kim *et al.* 2008, Castagnone-Sereno *et al.* 2010, Sonah *et al.* 2011 and Megl  cz *et al.* 2012). It is likely that the finding in this thesis research and other studies that (CG)<sub>n</sub> repeats are rare is a genuine occurrence. This pattern cannot be explained by CG content of the genome in *O. chilensis* (CG content of the 454 database is 35.7%). A possible reason for this is that there are three hydrogen bonds between C and G

and only two between A and T, which could increase the stability and hence reduce mutability. Furthermore, (CG)<sub>n</sub> repeats can undergo methylation of cytosine in most eukaryotic genomes, the methylated cytosine tends to mutate to thymine (Bird *et al.* 1986, Schorderet *et al.* 1992, Pelizzola *et al.* 2011). Toth *et al.* (2000) reported the patterns of repeat motif are not consistent across taxa; the most abundant di-repeat is (AC)<sub>n</sub> in mammals, whereas (AG)<sub>n</sub> is under-represented in *Caenorhabditis elegans*. Therefore, it is difficult to make further comparisons here as the finer details seem to be species specific.

The most common tri-repeats in the present study are (AAT)<sub>n</sub> and (ATC)<sub>n</sub>, interestingly here the (CCG)<sub>n</sub> repeat is one of the rarest, but in the study of Toth *et al.* (2000) they found this repeat to be the most frequent in vertebrate groups, which again demonstrates that there is no common pattern of repeat motif across taxa. By far the most common tetranucleotide repeat in the *O. chilensis* data set is (ACGT)<sub>n</sub>. Toth *et al.* (2000) found that repeats that contained <50% GC were more abundant; there was, however, no common taxonomic pattern. These species-specific patterns indicate different forces acting on evolution of the repeats.

In a recent study by Meglécz *et al.* (2012) the authors conducted a survey of 454 shotgun sequences from 154 non-model organisms. Their finding suggest that there is much heterogeneity in comparison of microsatellite sequences in evolutionary clades that are older than around 200 million years ago (MYA) – a phylogenetic signal was found in the comparisons that are more recent than this cut off point. This suggests that sampling within the oyster clade could reveal a phylogenetic signal, which could help further describe patterns and processes of evolution in this group. However, there are few comparative studies that have characterised microsatellite regions - these research questions need to be addressed as it will lead to a better understanding of the evolutionary mechanisms that are generating microsatellite loci. Of particular interest would be to compare the repeat motifs at different hierarchal levels in the Class Bivalvia to find out if there is any phylogenetic conservation of repeat motif.

One limitation of this study is that a 454 genome sequence may be biased, and may not fully represent the entire genome of the species in question – therefore any comparisons of findings may not be biologically meaningful. However, this concern has been addressed by Meglécz *et al.* (2012), the authors found that while a 454 shotgun reads are not a perfect representation of the genome – they did find a high correlation of variables generated

through 454 sequencing and their genomic counterparts. The authors conclude that inferences made using 454 shot gun reads are biologically meaningful.

PCR primers for length polymorphic microsatellite loci have been designed for *O. chilensis* in this thesis study, however, for a robust population genetic study more work is needed to confirm their reliability. Loci *Ost\_chi\_076.1* and *Ost\_chi\_084* had significant deviations from HWE; the reasons could be null alleles. Molluscs have been shown to have a high proportion of null alleles (Li *et al.* 2003; Astanei *et al.* 2005). Null alleles are thought to result from substitutions in the flanking region, which could inhibit PCR primer binding. McInerney *et al.* (2011) found the occurrence of cryptic repetitive DNA in the flanking sequence, which cause multiple PCR primer binding making it difficult to interpret.

The approach of using 454 sequencing, followed by touchdown PCR and M13 sequence tagging, has been shown to be a fast and cost effective method of generating novel microsatellite markers for a number of species. However, the process has been problematic with *O. chilensis*. A total of 125 primer pairs were tested but only four those primer pairs were shown to consistently amplify DNA of the expected size range. Other primer pairs had problems associated with: excessive stutter, the presence of monomorphic peaks, or simply did not amplify DNA. This is a very low success rate compared with other studies in our laboratory (e.g. Constable 2014; Hannan 2014; Silva and Gardner 2014). Failure rates are generally not published in the literature (presumably due to the belief that generating robust markers from genomic sequencing data is now routine), so it is difficult to assess if other researchers have had similar problems.

In a search for orthologous DNA sequences, the custom blast database was BLASTED against all repetitive elements found for oysters on GenBank. There were many highly significant ( $<e^{-20}$ ) matches between the 454 database and microsatellite sequences published for oyster species that were retrieved from GenBank (Table 2.5a to 2.5e). PCR primers already developed for these loci could be transferable to *O. chilensis*. Many of the blast hits were downstream or upstream of the microsatellite repeat region so in many cases it was impossible to compare the size or repeat unit of the microsatellite to the *O. chilensis* database sequences, but given the very low e-values overall, it is unlikely that these hits were matched by chance and they are likely to reflect conserved sequence regions which are homologous between species. It should also be noted that the ‘best’ hits were restricted to single matches, therefore, looking for more than one match with potentially lower e-values could reveal more

coverage of the microsatellite regions, and allow comparison of repeat motifs. There was, however, one exception: in one locus the query and reference sequences overlapped to show a common microsatellite loci (Figure 2.7). The figure illustrates an alignment between a sequence from the 454 data set and the microsatellite region *Cvi2i20.fa* from *Crassostrea virginica*. Here it is possible to see the BLAST hit is in the conserved flanking region, and the query and reference sequences overlap to reveal a common microsatellite motif. In *O. chilensis* the repeat appears to have become a compound repeat - the accumulation of substitutions in a microsatellite repeat have been postulated as the cause of limited expansion (death of the microsatellite). It should be noted that the BLAST hits are limited by a 'bias' on GenBank – not all oyster species have had their genomic sequences annotated for microsatellite loci – the identification of further orthologous sequences is expected if more research is conducted in non-commercial oyster species or when full genomes are annotated for microsatellite sequences.

Many of the BLAST hits were for transposable elements, and there were many sequence matches to the transposon related satellite sequence *BcII*. This has been found in a number of oyster species (López-Flores, *et al.* 2004), but it is the first time *BcII* has been identified in *O. chilensis*. Closer investigation of the association of transposons with microsatellite loci is needed as it could be the reason for low success rate in the amplification of microsatellite loci in this thesis. This will provide an avenue for further research.

Interpretation of microsatellite alleles has been problematic in some studies due to the occurrence of null alleles, homoplasy and deviations from Hardy-Weinberg expectations. Reece *et al.* (2004) found high levels of polymorphism in microsatellite flanking regions resulting in null (non-amplifying) alleles, and some markers were also found to depart from Mendelian ratios, both as heterozygote excess and deficit. Hedgecock *et al.* (2004) found 49 of 96 microsatellite markers to contain null alleles in the Pacific Oyster (*Crassostrea gigas*). Li *et al.* (2003) tested microsatellite markers for inheritance patterns in the abalone *Haliotis discus hannai*, of the seven loci they tested four contained null alleles. Not conforming to HWE could mean that there are high levels of inbreeding, or selection at a particular locus, or it could be due to the Wahlund effect (Nielsen *et al.* 2003). Further work is therefore needed to determine the reason for a deviation from HWE of the loci developed in this study. This could follow the work of Brownlow *et al.* (2008), in this study they aimed to distinguish whether deviations from HWE were due to natural processes or null alleles. To achieve this they redesigned primers for loci that showed a heterozygote deficiency – in some individuals

they were able to amplify a second allele – demonstrating that the deviation from HWE was due to unreliable primers and not natural causes. This could be the case with *Ost\_chi\_076.1* due to there being fewer heterozygotes than expected.

Another potential problem with microsatellite markers is homoplasy, which in microsatellite alleles of the same length can be a source of underestimation of allelic diversity. This may then overestimate the rate of gene flow when mutation rate is high (Epperson, 2004). There are two types of homoplasy: ‘detectable’ and ‘un-detectable’. Undetectable homoplasy could occur if the two alleles are identical by size but not by descent, detectable homoplasy occurs when there is a point mutation in the flanking region creating an allele the same size as an existing one. Adams *et al.* (2004) found homoplasy was only common for compound and interrupted repeats and that empirical estimates of detectable homoplasy reported only a slight (1-2%) underestimation of genetic differentiation. Sequencing of parental microsatellite alleles to establish actual allele sizes can reveal size homoplasy and the presence of more than one lineage. MacAvoy *et al.* (2008) found size homoplasy within most of their microsatellite loci for greenshell mussels, with one locus showing almost three times as many alleles by sequence as by size.

## 2. 5 Summary

1. This study has identified many microsatellite regions for the commercially important oyster species *Ostrea chilensis*. Four sets of PCR primers have been designed to amplify polymorphic microsatellite loci. More work is needed to develop a panel of reliable PCR primers for length polymorphic microsatellite loci. The data presented will aid that work. Due to the difficulties in developing reliable length polymorphic microsatellite primers for *O. chilensis* an alternative marker was employed for use in the population genetics study presented in the next chapter.

2. The presence of orthologous microsatellite regions in different oyster species has been identified. They could be used to study the evolution of microsatellites in oysters, and they could have potential for marker transfer to *O. chilensis* for use in population genetic studies.

3. Further study into the relationship between flanking regions of microsatellite repeats, and the association of transposons in the microsatellite containing sequences, could be an interesting area of further research, and a step towards identifying the reasons for the difficulties in this study.

## 2.6 References

- Abdelkrim J, Robertson B, Stanton J, Gemmell N. (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46: 185–192
- Astane I, Gosling E, Wilson JIM, Powell E. (2005). Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Molecular Ecology*, 14(6), 1655-1666.
- Barbara T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, 16(18), 3759-3767.
- Barbará T, Palma-Silva C, Paggi GM, Bered F. (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*.16: 3759-3767
- Bird A. (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321: 209–213
- Brownlow RJ, Burke T, Horsburgh GJ, Fish JD, Bell JJ, Dawson DA. (2008). Hidden genotype inconsistencies in invertebrates: a method for genotype validation and primer assessment in heterozygote deficient species. *BMC Genetics*, 9:55
- Castagnone-Sereno P, Danchin EGJ, Deleury E, Guillemaud T, Malausa T. (2010) Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne incognita*. *BMC Genomics* 11: 598
- Castoe T, Poole A, Gu W, De Koning A, Daza J. (2010) Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources* 10: 341–347
- Chambers GK, MacAvoy ES. (2000). Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 126(4), 455-476
- Constable, HB. (2014). Population structure, temporal stability and seascape genetics of two endemic New Zealand Pleuronectiformes, *Rhombosolea plebeia* (sand flounder) and *R. leporina* (yellowbelly flounder). Ph.D. Thesis. Victoria University of Wellington: N.Z.
- Ellegren H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews genetics*, 5(6), 435-445.

- Galbraith HS, Smith CM, Wozney KM, Zanatta DT, Wilson CC. (2011). Development and characterization of nine microsatellite loci for the endangered Kidneyshell, *Ptychobranchus fasciolaris*, and cross-amplification in closely-related lampsilines (Bivalvia: Unionoida). *Conservation Genetics Resources*, 3(3), 533-536.
- Gardner M, Fitch A, Bertozzi T, Lowe A. (2011) Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11: 1093–1101
- Glenn TC, Stephan W, Dessauer HC, Braun MJ. (1996). Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Molecular Biology and Evolution*, 13(8), 1151-1154.
- Goldstein D, Pollock D. (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. *Journal of Heredity* 88: 335–342
- Goldstein D, Schlötterer C. (1999) Microsatellites: evolution and applications. New York, USA: Oxford University Press Inc. 352 p
- Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P. (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11: 591–611
- Hannan DA. (2014). Population genetics and connectivity in *Paphies subtriangulata* and *Paphies australis* (Bivalvia: Mesodesmatidae). Ph.D. Thesis. Victoria University of Wellington: N.Z.
- Harr B, Schlötterer C. (2000). Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics*, 155(3), 1213-1220.
- Hedrick PW. (2000) Genetics of Populations 2nd Ed, Boston: Jones and Bartlett
- Kang JH, Kim YK, Park JY, Noh ES, Jeong JE, Lee YS, Choi TJ. (2013). Development of microsatellite markers for a hard-shelled mussel, *Mytilus coruscus*, and cross-species transfer. *Genetics and Molecular Research*, 12(3), 4009-4017.
- Karaoglu H, Lee CMY, Meyer W. (2005) Survey of simple sequence repeats in completed fungal genomes. *Molecular Biology and Evolution* 22: 639–649

- Katti MV, Ranjekar PK, Gupta VS. (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* 18: 1161–1167
- Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD. (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research*, 18(1), 30-38.
- Kim TM, Laird PW, Park PJ. (2013). The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell*, 155(4), 858-868.
- Kim TS, Booth JG, Gauch HG, Sun Q, Park J. (2008) Simple sequence repeats in *Neurospora crassa*: distribution, polymorphism and evolutionary inference. *BMC Genomics* 9: 31
- Kimura M, Ohta T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences*, 75(6), 2868-2872.
- Korbie DJ, Mattick JS. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3(9), 1452-1456.
- Kruglyak S, Durrett RT, Schug M, Aquadro CF. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceeding of the National Academy of Sciences U S A* 95:10774–10778
- Leese F, Held C. (2011). Analyzing intraspecific genetic variation: a practical guide using mitochondrial DNA and microsatellites. *Held, C., S. Koenemann, CD Schubart (eds): Phylogeography and Population Genetics in Crustacea, Crustacean Issues Vol 19, CRC Press*, 3-30.
- Levinson G, Gutman GA. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, 4(3), 203-221.
- Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003) Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes*; 3:228-232
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular ecology*, 11(12), 2453-2465.

López-Flores I, de la Herrán R, Garrido-Ramos MA, Boudry P, Ruiz-Rejón C, Ruiz-Rejón M. (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene*, 339, 181-188.

MacAvoy ES, Wood AR, Gardner JPA. (2008). Development and evaluation of microsatellite markers for identification of individual Greenshell™ mussels *Perna canaliculus* a selective breeding programme. *Aquaculture*, 274(1), 41-48.

Malausa T, Gilles A, Megléc E, Blanquart H, Duthoy S. (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644

Marín A, Fujimoto T, Arai K. (2012). Isolation and characterization of 12 microsatellite loci in the Peruvian scallop *Argopecten purpuratus* and cross-species amplification in other scallop species (family Pectinidae). *Conservation Genetics Resources*, 4(1), 179-182.

Martin J, Pech N, Megléc E, Ferreira S, Costedoat C. (2010) Representativeness of microsatellite distributions in genomes, as revealed by 454 GS-FLX titanium pyrosequencing. *BMC Genomics* 11: 560

McInerney CE, Allcock AL, Johnson MP, Bailie DA, Prodöhl PA. (2011). Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in molluscs. *Heredity*, 106(1), 78-87

Megléc E, Anderson SJ, Bourguet D, Butcher R, Caldas A, Cassel-Lundhagen A, d'Acier AC, Dawson DA, Faure N, Fauvelot C, Franck P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju J, Patt A, Petenian F, Silvain JF, Wilcock HR. (2007) Microsatellite flanking region similarities among different loci within insect species. *Insect Molecular Biology* 16:175–185

Megléc E, Costedoat C, Debut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26(3):403-404

Megléc E, Nève G, Biffin E, Gardner MG. (2012). Breakdown of phylogenetic signal: a survey of microsatellite densities in 454 shotgun sequences from 154 non model eukaryote species. *PloS one*, 7(7), e18229

- Megléc E, Pech N, Gilles A, Martin JF, Gardner MG. (2012). A shot in the genome: how accurately do shotgun 454 sequences represent a genome? *BMC research notes*, 5(1), 259
- Megléc E, Petenian F, Danchin E, D'Acier AC, Rasplus JY, Faure E. (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology*, 13(6), 1693-1700.
- Messier W, Li SH, Stewart CB. (1996). The birth of microsatellites. *Nature*, 381, 483.
- Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Grønkjær P. (2003). Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology*, 12(6), 1497-1508.
- Noor MA, Feder JL. (2006). Speciation genetics: evolving approaches. *Nature Reviews Genetics*, 7(11), 851-861.
- Ohta T, Kimura M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical research*, 22(02), 201-204.
- Peakall R, Smouse PE. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6, 288-295.
- Peakall R, Smouse PE. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537-2539.
- Pelizzola M, Ecker JR. (2011). The DNA methylome. *FEBS letters*, 585(13)
- Reece KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity*, 95(4), 346-352.
- Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* 46: 217–223
- Schorderet D, Gartler S. (1992) Analysis of CpG suppression in methylated and non-methylated species. *Proceedings of the National Academy of Sciences U SA* 89: 957–961
- Schuelke M. (2000). An economic method for the fluorescent labelling of PCR fragments. *Nature biotechnology*, 18(2), 233-234.

- Silva CN, Gardner JPA. (2014). Development and characterisation of 12 microsatellite markers for the New Zealand endemic scallop *Pecten novaezelandiae*. *Conservation Genetics Resources*, 6(2), 327-328.
- Slatkin M. (1995). Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution*, 12(3), 473-480.
- Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK. (2011) Genome-wide distribution and organization of microsatellites in plants: An insight into marker development in *Brachypodium*. *PLoS One* 6, e21298.
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15: 199–203
- Taylor JS, Durkin JM, Breden F. (1999). The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Molecular Biology and Evolution*, 16(4), 567-572.
- Tóth G, Gáspári Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research*, 10(7), 967-981.
- Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), 535-538.
- Zhang DX, Hewitt GM. (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular ecology*, 12(3), 563-584.

## 2.7 Appendix

### *Appendix 2.1:* Hexanucleotide repeats

Repeat Motif	Frequency
AACGTT	655
AAACGT	346
ACACAT	254
AACGGT	239
AACCGT	214
AACCCT	142
AAAAAC	110
AAAAAT	110
AACTAC	97
AAACTT	95
AAAGTT	81
AAAAAG	78
AAAACT	68
ACTAGT	65
AAAATT	61
ACCCGT	61
AATAGT	59
AAGGTC	56
ACCCCC	52
AATACT	50
AAACCT	48
ACACGT	47
ACGTAT	47
ACATAT	42
AAATAT	40
AACCTT	40
ACATGT	39
AATTAC	36
AAAAGT	33
ACACAG	33
ACGACT	32
AAAGGT	30
ACCGGT	30
AAAATG	28
ACGAGT	26
AAATAC	22
AAATCT	22

AAATTC	22
AAATTT	21
AAACGG	20
AAACAG	19
ACAGGT	19
ACCACT	19
AGATAT	19
ACAGAG	18
AAAATC	17
ACATAG	17
AAATGT	16
AAGAGT	16
AAGTAG	16
ACGTAG	16
AGGGGG	15
AAAGAG	14
AAAGAT	14
AATTAG	14
AAACAC	13
AAGTAC	13
AAAACC	12
AACAAT	12
AGAGCG	12
ATCCCG	12
AAATAG	11
AATGAT	11
ACAGAT	11
AACTAT	10
AAGGGT	10
ACCCCT	10
AAAAGG	9
AAACAT	9
AAATTG	9
AACAGC	9
AATTAT	9
AAAGTG	8
AACATT	8
AAGAGG	8
AAGTCT	8
AAGTGT	8
AAACTC	7
AACACT	7
AACGAC	7
AACTTG	7
AAGTAT	7
AATCAT	7

AGAGAT	7
ATCCCC	7
AAAACG	6
AAAGTC	6
AACACC	6
ACACTC	6
ACTCCT	6
AAACCC	5
AAATGG	5
AACAAG	5
AACTGT	5
AAGAAT	5
AAGGAC	5
AAGGAT	5
AATGTC	5
ACCTAT	5
ACTCTC	5
AAACCG	4
AACAGT	4
AACATC	4
AACTAG	4
AACTCT	4
AACTTC	4
AAGGGG	4
AATATC	4
AATGGT	4
ACACCC	4
AGATCG	4
AAACTG	3
AAAGAC	3
AACCCC	3
AACCCG	3
AACGAG	3
AAGATC	3
AATATG	3
AATATT	3
AATCCC	3
AATGAC	3
AATGTG	3
AATTCC	3
ACCCTC	3
ACGGAT	3
AGGGAT	3
CCCCCG	3
AAAAGC	2
AAATCG	2

AACACG	2
AACATG	2
AACCAC	2
AACCAG	2
AACCAT	2
AACCGG	2
AACGGG	2
AACTGG	2
AAGACG	2
AAGACT	2
AAGCAT	2
AAGCGG	2
AAGGAG	2
AAGTCC	2
AATACC	2
AATACG	2
AATCAG	2
AATCCG	2
AATCTC	2
AATCTG	2
AATTCTG	2
ACAGGG	2
ACAGTC	2
ACATCT	2
ACATGG	2
ACCACG	2
ACCAGT	2
ACCCCG	2
ACCCTG	2
ACCGAT	2
ACCTAG	2
ACCTCT	2
ACTAGG	2
ACTATC	2
ACTCAT	2
AGAGCT	2
AGAGGG	2
AGATGG	2
AAAGCC	1
AAAGCT	1
AAAGGC	1
AAAGGG	1
AACAGG	1
AACGAT	1
AACGTC	1
AACGTG	1

AACTCG	1
AACTGC	1
AAGATG	1
AAGATT	1
AAGCAC	1
AAGCGT	1
AAGGGC	1
AAGTCG	1
AATAGC	1
AATAGG	1
AATCAC	1
AATCCT	1
AATCGG	1
AATCGT	1
AATGAG	1
ACACCT	1
ACACTG	1
ACAGCC	1
ACAGCT	1
ACAGTG	1
ACATGC	1
ACCATG	1
ACCCAG	1
ACCGAG	1
ACCGCG	1
ACCGGG	1
ACCTCC	1
ACCTGC	1
ACGAGG	1
ACGATC	1
ACGATG	1
ACGCAT	1
ACGCGC	1
ACGGGG	1
ACGTCC	1
ACGTCG	1
ACTATG	1
ACTGAT	1
AGATCT	1
AGCCGG	1
AGGATC	1
AGGATG	1
AGGCCG	1
AGGCCT	1
AGGGCG	1
ATATGC	1

ATCATG	1
ATCGCC	1

**Appendix 2.2:** Pentanucleotide repeats

<b>Motif</b>	<b>Frequency</b>
AACGT	3433
ACCGT	581
ACTAT	246
AACTT	232
ACACT	222
AAACT	199
AACCT	114
AAAGT	110
ACTCT	109
AAACG	87
AAAAT	82
AAAAC	67
AAATC	67
AAGAG	62
AAGGT	60
AAATT	44
ACCCT	44
AAAAG	41
AATAT	41
AACAC	31
AATTC	31
AATAC	30
ACCCC	30
ACGAG	30
AAATG	21
ACACG	16
AATGT	15
AACGG	13
AACAG	11
AACCG	11
AAGAT	11
AGAGG	11
AACTG	10
ACAGT	9
ACTAG	8
AACAT	7
ACGCG	7
AGGCC	7
AGGGG	7

ACACC	6
AAACC	5
AACCC	5
ACAGG	5
ACTCC	5
AAAGG	4
AATAG	4
AATCT	4
ACCGG	4
ATATC	4
AATCG	3
ACCAG	3
ACCAT	3
ACTGC	3
AAAGC	2
AAGTC	2
ACCCG	2
ACGGG	2
AGATG	2
AGCCG	2
AGCGG	2
AGGAT	2
CCGCG	2
AAGAC	1
AAGCG	1
AAGGG	1
AAGTG	1
AATGC	1
AATGG	1
ACATC	1
ACCTG	1
ACGAT	1
AGCCC	1
AGCCT	1
ATCCC	1

# Chapter Three: A Population Genetics Analysis of *Ostrea chilensis* Using Randomly Amplified Polymorphic DNA

---

*"Nothing in Biology makes sense except in the light of evolution" – Theodosius Dobzhansky*

*"Nothing in evolution makes sense except in the light of population genetics." – Michael Lynch*

## 3.0 Aims

This is the first study into the population genetic structure of the commercially important oyster, *Ostrea chilensis*, using samples from New Zealand and Chilean sites. The aims of the study are to: (1) develop genetic markers for the species, (2) investigate patterns of genetic structuring in this species, (3) test for the presence of boundaries to gene flow in New Zealand, and (4) determining the level of gene flow between New Zealand and Chile.

## 3.1 Introduction

Understanding a species population genetic structure is important for a number of reasons: it provides a base line measure of genetic diversity, this can be used to monitor subsequent genetic changes; it also identifies genetic resources that can potentially be used for selective breeding and aquaculture. At a broader scale, it can provide an insight into the larger scale patterns and processes that generate biodiversity. The evolutionary potential of a species and its adaptability is assumed to be higher with increased genetic variation, therefore, it should be monitored over time, and used to inform management, and conservation plans.

### 3.1.1 Spatial genetics in the marine environment

There are three spatial patterns that are used to describe the population genetic structure of a species: (1) Panmixia, the complete mixing of alleles among individuals that make up a single population; (2) Distinct populations, whereby individuals will breed mainly with individuals from their local area creating genetically distinct groups; (3) isolation-by-distance, based on an island model of dispersal, populations that are adjacent to each other

are more closely related than geographically distant populations (Wright, 1943); and (4) Chaotic patchiness, where the patterns of genetic structure do not reflect geographic proximity – this fourth spatial pattern is thought to be the result of stochastic processes.

Despite the lack of any obvious barriers to gene flow, and the vast dispersal capabilities of many marine organisms, it is not uncommon for populations to display significant levels of genetic differentiation (Neilsen *et al.* 2009a). The processes that generate this differentiation involve a complex interaction of the organisms' biology with environmental and physical variables, which may generate barriers to gene flow and/or create selection gradients.

The biology and behaviour of an organism are a major influence to the population genetic structure of the species. Shanks and Brink, (2005) found that larvae of the *Tellina* spp. and *Mulinia lateralis* were able to avoid being swept off shore and on shore by adjusting their swimming behaviours to reflect times of upwelling and down welling. In marine organisms, and particularly marine invertebrates, pelagic larval duration (PLD) is thought to be an important factor that determines the levels of genetic differentiation. Previous studies have suggested that population subdivision is higher in species with direct development rather than planktonic development; and that increasing dispersal duration should facilitate increased gene flow among populations, this assumption is generally true. The Brittle Star *Amphipholis squamata*, is a brooding species, which is thought to limit dispersal. In New Zealand it has been shown to have distinct population subdivision (Roy and Sponer, 2002). Despite having the longest known PLD of any known species (>2 years), the New Zealand rock lobster (*Jasus edwardsii*) has three distinct populations within New Zealand (Thomas and Bell, 2013). Recently, in a large-scale analysis of the effects of geography and life history traits in benthic marine fishes, PLD was found to be unable to account for genetic variation (Riginos *et al.* 2011). Also, Weersing and Toonen (2009) analysed 300 peer reviewed studies, they did not find a relationship with PLD and genetic structure after direct developers were removed from the study. These results suggested that other factors such as larval behaviour or oceanographic features are more influential in creating genetic structure. However, Faurby and Barber (2012) in a theoretical analysis of the correlation between PLD and estimates of population genetic structure found that a weak correlation between PLD and genetic structure could be the result of variation and uncertainty in the terms associated with inferring genetic structure from  $F_{st}$  estimates. Two dispersal mechanisms that have received less attention are: sperm dispersal and post-larval dispersal by drifting or rafting. Sperm dispersal has been identified as the main strategy of connectivity between fragmented populations of the sponge

*Scopalina lophyropoda* (Blanquer and Uriz, 2010). Sperm dispersal could be an important mechanism of gene flow in brooding organisms such as *O. chilensis*. Although empirical studies of the colonial hydroid *Hydractinia echinata* have shown sperm dispersal only contributes to gene flow over small spatial scales (3 metres) (Yund, 1990). Dispersal by rafting invertebrates has achieved much more attention (Grantam *et al.* 2003; Highsmith, 1985). Dispersal by rafting on kelp has been suggested for the occurrence in Chile of *O. chilensis*, which is thought to have originated in New Zealand (Foighil *et al.* 1999).

The sea appears to be an open environment without barriers to gene flow, however, a number of studies have reported strong barriers to gene flow in the ocean. Thornhill *et al.* (2008) found a barrier to intercontinental movements at the Arctic Polar Front, for the nemertean worm, *Parborlasia corrugatus*. In New Zealand, upwelling in the Cook Strait has been posited as a barrier to gene flow in a number of studies of different species: in the echinoderm *Patiriella regularis* (Waters and Roy, 2004); the New Zealand green shell mussel *Perna canaliculus* (Apte and Gardener, 2001); the brittle star *Ampholis squamata* (Sponer, 2002); and in three limpet species *Cellana ornata*, *Cellana radians* and *Cellana flava* (Goldstein *et al.* 2006). Although, the mechanistic cause is unknown anthropogenic influences have been implicated as barriers to gene flow. Puritz and Toonen (2011) in a study of the Bat Star (*Patiria miniata*) found the coastal populations of the Southern Californian Bight land runoff and pollution are not only increasing population genetic structure, but also decreasing genetic diversity in populations closest to the pollution.

Studies that tested for correlations between environmental factors and genetic structure in the marine environment (termed seascape genetics), have identified selection gradients that may shape population genetic structure despite high levels of migration (Selkoe *et al.* 2010). Persistent environmental gradients or complex mosaics of environmental variables may impose divergent selection on populations. This may result in individuals becoming locally adapted. The selection pressures may overwhelm the effects of gene flow from other non-adapted, distant populations - especially if they are large populations. It is important to appreciate that molecular subdivision is not needed for local adaptation. In populations with high gene flow there can be homogeneity of neutral allele frequencies among populations, but not all loci are necessarily homogeneous if they are under selection. Therefore, a different population structure could be found when comparing neutral loci with those under selection. There has been a growing awareness that there are fine scale environmental gradients in the sea; recent studies have shown persistent environmental variables that can generate fine scale

differences in genetic variation. Clines of environmental variation in the sea may promote local adaptation – although this is not an accepted fact. For example, the patterns of variation at the rhodopsin gene (RH1), which is crucial for dim light vision in vertebrates was found to match the characteristics of specific light environment suggesting local adaptation acting on the RH1 gene (Larmuseau *et al.* 2010).

Often the pattern of genetic structure in marine species fits a model that has been described as ‘chaotic genetic patchiness’. Although marine populations have significant levels of pairwise genetic differentiation they are often not correlated with geography, and/or the genetic structure can display temporal instability (Johnson and Black 1982; Hedgecock 1994; Selkoe *et al.* 2006; Arnaud-Haond *et al.* 2008). An unresolved question is whether this apparent ‘chaos’ could be explained with better models.

### 3.1.2 Previous work on the study species

Oysters are an interesting group for studying the comparative effects of parental care. Oysters show a graduation of parental care from broadcast spawners to brooding species that release their young either as early planktotrophic larval stages, or as advanced larvae capable of immediate metamorphosis (Buroker, 1985). All oysters in the genus *Ostrea* brood their oysters in the pallial cavity (Buroker, 1985). *O. chilensis* has the longest brooding period lasting 8 weeks (Toro, 1990). The fully developed larvae are able to settle within minutes of release - although a portion of them may remain in the plankton for up to 20 days (Cranfield *et al.* 1968; DiSalvo *et al.* 1983; Cranfield and Michael 1989). This brooding strategy is also thought to prevent high larval dispersal (Toro and Gonzalez, 1990). An average of 70,000 larvae can be expected per brooding female; this is low fecundity in comparison with other *Ostrea* species (Winter *et al.* 1984). The closely related flat oyster *Ostrea edulis* is a highly fecund species, with females spawning as many as 1-2 million eggs per brood (Cole, 1941). Fecundity and brooding strategy are important drivers of genetic structure and have achieved much attention in the literature.

Previous work on the population genetics of *O. chilensis* has been carried out in Chile. In Chile the distribution of *O. chilensis* is restricted to the Isle of Chiloé and the Islands of Guaitecas. Toro and Aguilla (1996) conducted a study using allozymes as genetic markers in populations of *O. chilensis* that covered the whole distribution of this species in Chile. They found significant deficiencies of heterozygotes at some loci, as has been observed in a

number of bivalve studies (eg. Gaffney *et al* 1990). The work of Toro and Vergara (1996) suggests that post-settlement selection against heterozygotes may be a cause of these deficiencies. The mean value of Nei's genetic distance ( $I$ ) (Nei, 1972) was 0.846 that suggests differentiation between populations. The authors proposed that this is due to the shortened larval life stage and hence limited dispersal resulting in population structuring.

More recently, Toro and Gonzalez (2009) investigated the genetic variation in natural populations in Chile. They used Randomly Amplified Polymorphic DNA (RAPDs) to analyse the levels of population differentiation. The authors sampled individuals from the same five wild populations along the inland coast of Chiloé and Guaitecas Islands as the previous study by Toro and Vergara (1995). Two of these populations have not been restocked, the other three populations had been stocked by the transfer of juveniles. Values of Nei's genetic distance did not reveal significant differences among populations with an average of 0.113. The range of genetic differences is 0.0713 (between Caulin and Quempillen populations) to 0.1549 (between Rilán and Melinka populations). They found levels of gene flow to be high ( $N_m = 1.587$ ), this is high enough to break down genetic structuring caused by drift, given that genetic drift will result in substantial local differentiation if  $N_m < 1$  but not if  $N_m$  is  $> 1$  (Slatkin, 1987). However, they did find a significant correlation between geographical distance and genetic distance supporting an isolation-by-distance model of population structure. The authors suggest the high levels of gene flow could either be a consequence of anthropogenic movement of juvenile (seed) for stocking or the influence of oceanographic processes.

### **3.1.3 Genetic Markers**

To elucidate the pattern of genetic structuring in *O. chilensis* samples in Chile and New Zealand, randomly amplified polymorphic DNA (RAPD) markers were used in this thesis research. Although RAPDs can have the problem of difficult interpretation, and concerns of limited reproducibility; they are simple to use, because no prior genomic information is needed, they only require a small amount of development effort, and the alleles can be quickly identified using gel electrophoresis. Also, a large number of markers that can enable high coverage of the genome can be quickly generated. Despite their limitations, RAPDs have been successfully used to study a range of marine species for example, the endemic New Zealand mussel *Perna canaliculus* (Star *et al.* 2003), two species of New Zealand

endemic limpets (*Scutellastra kermadecensis* and *Siphonaria raoulensis*) (Wood and Gardner, 2007) and in Chilean populations of *Ostrea chilensis* (Toro and Gonzalez, 2007).

### 3.1.4 Hypothesis

#### **Hypothesis One: genetic structure**

A very short pelagic larval duration results in limited dispersal potential, this results in genetic difference among populations, populations have limited exchange resulting in genetic differentiation ( $F_{st} > 0$ , or  $\Phi_{PT} > 0$ ).

#### **Hypothesis Two: Isolation by distance**

When a Mantel test is conducted between geographical distance and genetic distance, there is a positive relationship ( $R_{xy} > 0$ ). This would suggest that the population genetic structure around New Zealand can be explained by an isolation-by-distance model.

#### **Hypothesis Three: Barriers to gene flow**

Using the AWclust method to analyse the number of possible groups within the multilocus genotype data set for all individuals in all sampling sites, the most likely grouping will divide sampling sites into the *a priori* grouping of above and below 42°S as found in a number of other New Zealand studies of population genetics (e.g. Wei *et al.* 2013b).

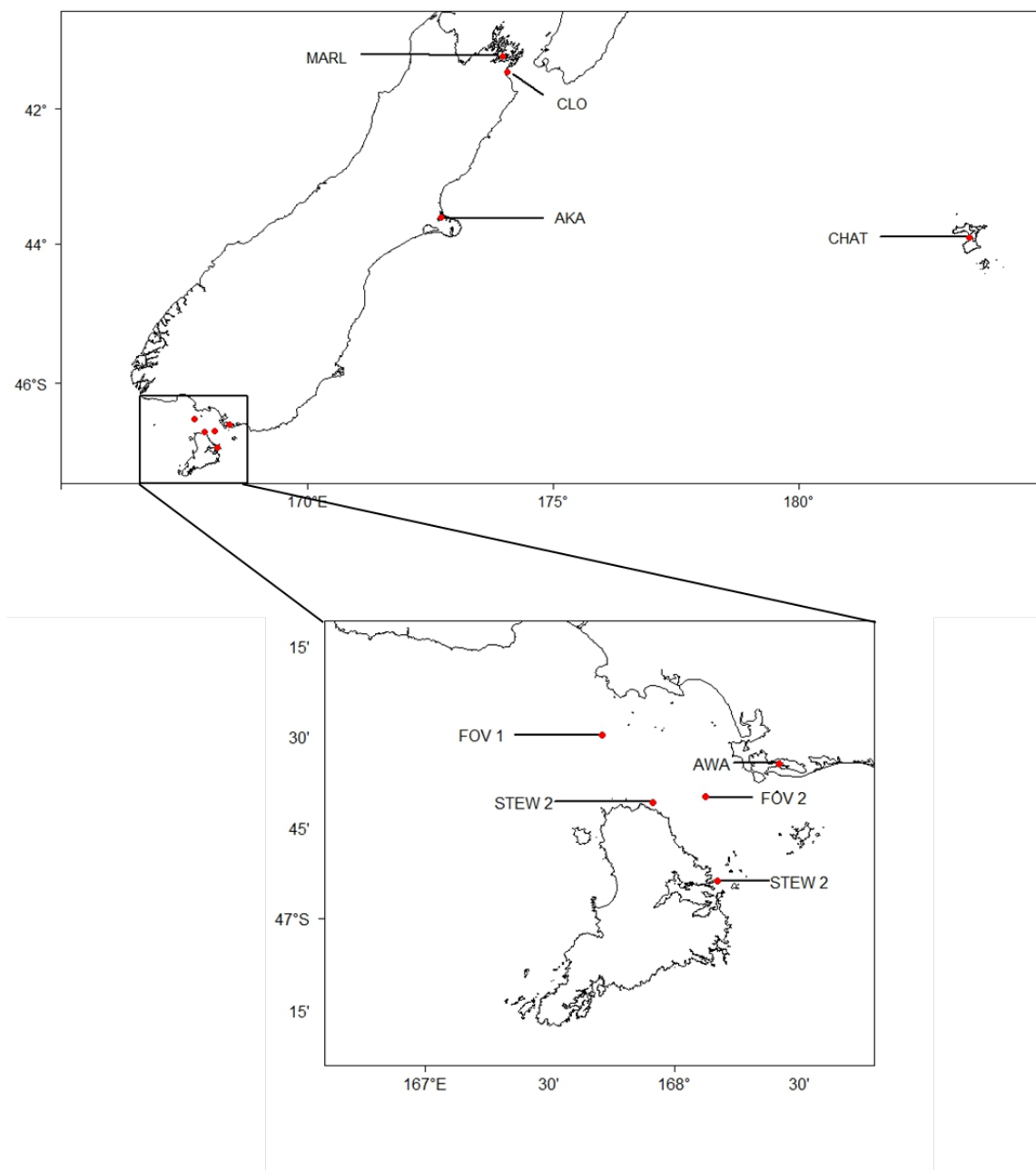
## 3.2 Methods

### 3.2.1 Sample collection

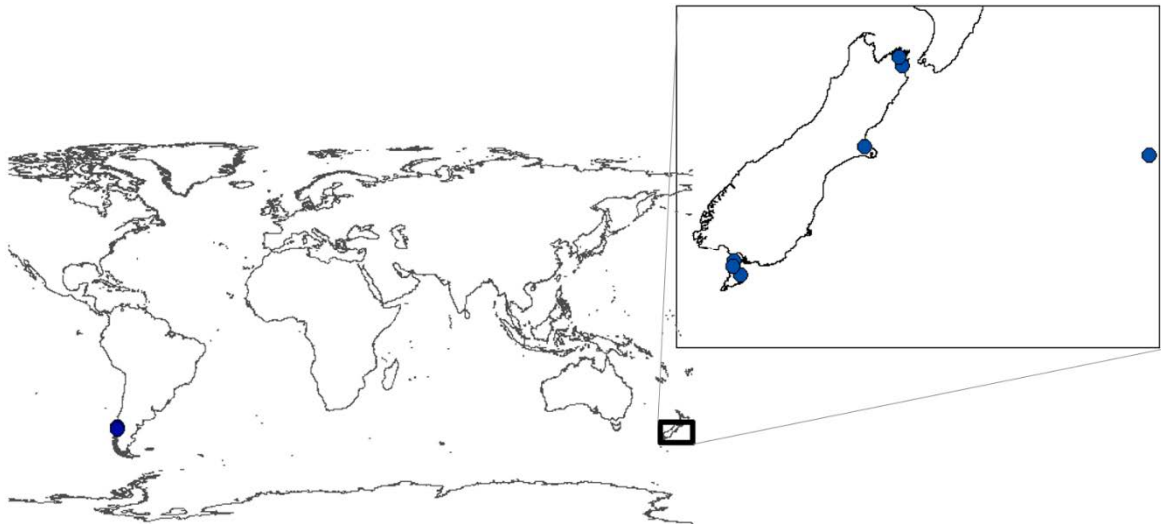
Frozen and live samples were collected from locations around New Zealand and Chile (Table 3.1) and sent to the laboratory. Approximately 60 – 100 mg of mantle tissue from each individual oyster was dissected, and then transferred to and stored in sterile 1.5 ml cryovials containing 70% ethanol and stored 4°C for later processing. Figures 3.1 and 3.2 show the locations of sampling sites in New Zealand and Chile.

**Table 3.1:** Sampling size and localities

<i>Sampling site</i>	<i>n</i>	<i>Decimal degrees</i>	
		<i>Latitude</i>	<i>Longitude</i>
<b><i>Northern group</i></b>			
Cloudy Bay (CLO)	24	-412620	1740407
Marlborough Sounds (MARL)	37	-411143	1735806
<b><i>Eastern group</i></b>			
Chatham Island (CHAT)	35	-435443	-1763235
Akaroa peninsula (AKA)	22	-433652	1724304
<b><i>Chilean group</i></b>			
Chiloé Island, Chile (CHI_1)	23	-422248	-732510
Chiloé Island, Chile (CHI_2)	22	-430140	-732930
<b><i>Southern group</i></b>			
Foveaux Strait (FOV_1)	23	-465551	1677118
Foveaux Strait (FOV_2)	39	-466646	1681403
Stewart Island (STEW_1)	34	-465349	1681024
Stewart Island (STEW_2)	33	-464053	1675448
Awarua Estuary (AWA)	33	-463428	1682459



**Figure 3.1:** Sampling locations around New Zealand



**Figure 3.2:** Location of Chilean samples relative to New Zealand

### 3.2.2 DNA Extraction

The DNA extraction procedure was as follows: 600  $\mu$ l of cell lysis buffer (100 mM Tris-HCl, pH 8.0, 50 mM EDTA, pH 8.0, 1% SDS (Sodium Dodecyl Sulfate)) was added to 1.5 ml tube. Approximately 15 mg of mantle tissue was added to the tube containing the lysis buffer and 6.0  $\mu$ l of proteinase K (20 mg/ml) was added to the tube; this was incubated at 65°C until the tissue was thoroughly dissolved. The tubes were then cooled to 4°C. Then 200  $\mu$ l of Ammonium acetate was added to the tube, mixed thoroughly, cooled to 4°C for 10 min; the tubes were then spun in a centrifuge at 12,000 rpm for 10 min at 4°C. The supernatant was transferred to a new tube, and then centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was transferred to a new tube, and an equal volume of isopropanol (~600  $\mu$ l) was added, mixed and left to rest of 2 minutes. It was then spun in the centrifuge at 12,000rpm for 10 minutes. The supernatant was removed and discarded with care to keep DNA at the bottom of the tube; 1 ml 70% ethanol was then added and inverted a number of times to wash the pellet. This was then spun in centrifuge at 12,000 rpm for 10 min. The ethanol was then discarded, the pellet was air dried for 10-20 min by leaving the cap of the tube open. The pellet was then resuspended in 100  $\mu$ l water. The solution was then stored at -20°C until further analysis.

### **3.2.3 Primer Screening**

Initial screening was performed using three individuals from disparate geographic locations and 30 decamer RAPD primers (Appendix 3.1) from Operon Technologies and 4 RAPD primers that were designed by Toro and Gonzalez (2007). Each individual was tested twice with the same primer. Primers that were retained for further analysis were those that showed reproducible, polymorphic and clear bands.

### **3.2.4 PCR protocol**

For all PCRs the 12.5µl reaction mixture contained 6µl of MyTaQ red mix (Bioline), 3µl of a single decamer primer, and 1µl of DNA (around 10ng/µl). All amplifications were performed in a thermocycler (Labnet model: multigene gradient) The thermocycler was preheated to 94°C and programmed for 40°C cycles: 4 initial cycles of 3 minutes at 94°C, 3 minutes at 36°C and 3 minutes at 72°C; followed by 36 cycles of 30 seconds at 94°C, 1 minute at 36°C, 2 minutes at 72°C; and then a final extension period of 8 minutes at 72°C. This PCR product was then held at 4°C until further analysis.

### **3.2.5 Analysis of RAPD-PCR products**

PCR products were separated by electrophoresis in agarose gels. 10 µl of each PCR reaction was loaded to a 2% agarose in 0.5 TBE (0.045 M Tris, 0.045 M boric acid, 0.001 M EDTA) containing 0.5 µg/ml ethidium bromide; Hyperladder II (Bioline) was added to multiple lanes as a size standard. Electrophoresis was performed at 70 V for 3 hours. The DNA of the PCR product was visualised using UV illumination and the gels were photographed using a gel docking system (Uvitech Cambridge). The gels were inverted to a negative using built in software, and images were then analysed using GelAnalyzer ([www.gelalyzer.com](http://www.gelalyzer.com)), which estimates the size of each band by producing a calibration curve using the ladder of known fragments. This programme automates the process of genotyping making the determination of loci less subjective.

Loci were named by the RAPD primer used and the estimated size. For example OPC\_09\_1600 is the loci/band amplified by the primer OCP 9 and 1600 is the estimated band size (in base pairs) (see Appendix 3.1 for a list of the loci). The bands were assessed for presence or absence in each DNA sample.

### **3.3 Data Analysis**

The following assumptions have been made about the data: (1) all of the loci have two alleles (presence of band and absence of band); (2) the bands do not co-migrate to the same position on the gel; (3) the populations are in Hardy-Weinberg Equilibrium (HWE) (4) the loci are from neutral parts of the genome (5) the loci are independent (there is no linkage disequilibrium). Where individual loci give very different values (2 standard deviations from the mean  $F_{st}$ ) they have been excluded and the data set re-tested to examine their effects. Since it is difficult to test for HWE with dominant markers some of the later analysis is frequency based which does not assume HWE.

A data set of presence or absence of homologous bands was created using all individuals. To test the reproducibility of the bands and the scoring error, a second data set was generated using 24 DNA samples taken from a range of sampling sites, and scored using the same protocols. The DNA samples were selected at random from unknown sampling sites and unknown oysters by another person. The samples were assayed for RAPD band diversity using all the primers as described above. The identity of the RAPD bands was compared to that of the original RAPD bands, and a scoring error rate of 2.5% was obtained, this was attributed to 5 loci that were scored differently. Once these loci were removed from the second data set an error rate of 0.01% was obtained. These loci were removed from the original data set before further analysis as they were determined to not give reproducible results.

#### **3.3.1 Genetic diversity**

Each RAPD band was treated as an individual locus with two alleles (absence and presence of band). Allele frequencies were estimated using the Bayesian method of Zhivotovsky (Zhivotovsky, 1999), with a non-uniform prior distribution implemented in the programme AFLP\_SURV v1.0 (Vekemans *et al.* 2002). This method gives an unbiased estimate of gene diversity and population genetic structure from dominant markers (Zhivotovsky, 1999; Krauss, 2000). The number of polymorphic markers at the 5% level, i.e. with allele frequencies within the range of 0.05 to 0.95, and Nei's gene diversity (H) were also computed in AFLP\_SUR.

### 3.3.2 Outlier loci

$F_{st}$  was calculated using the allele frequencies generated from the AFLP\_SUR analysis. Loci that had values that were two standard deviations away from the mean  $F_{st}$  over all loci were determined to be outliers. This analysis was used to identify single presumed neutral loci that maybe subject to diversifying selection, systematic scoring errors, or other factors that could affect their usefulness as markers of gene flow and drift. A global analysis was carried out using all loci and then for comparison the analysis was carried out with outlier loci removed. This analysis is basically an exploratory data analysis, rather than a method to identify loci under selection, which would require different assumptions.

### 3.3.3 Spatial genetic analysis

Wright's fixation index,  $F_{st}$  and Nei's genetic distance was computed in AFLP\_SURV and genetic differentiation among populations was tested using a permutation test of 5000 pseudoreplicates.  $P$ -values were corrected using the false discovery rate correction (FDR) (Verhoeven *et al.* 2005).

AMOVA (Excoffier *et al.* 1992) calculates an  $F_{st}$  analogue ( $\Phi PT$ ) which estimates variation within populations, among populations, and among regions. This analysis was implemented in GenAlEx v6.5 (Peakall and Smouse 2006, 2012). A matrix of squared Euclidean distances was computed from individual phenotypes (presence or absence of a band, rather than allele frequency). AMOVA gives estimates of population genetic structure from dominant markers concordant with those estimated from co-dominant markers (Isabel *et al.* 1999). Since it is a frequency based analysis it also has the advantage of not assuming HWE.

Pairwise population  $\Phi PT$  values were calculated using GenAlEx. Estimates of gene flow ( $N_m$ ) were also calculated using the formula  $0.25((1/\Phi PT)-1)$ . Measurement of  $N_m$  is an estimate of the gene flow over many generations. Indirect estimation of  $N_m$  from  $\Phi PT$  involves numerous assumptions: constant population size; random migration; no selection; mutation or spatial structure, these are often violated in natural populations (Whitlock and McCauley, 1999). These indirect estimates of  $N_m$  must be interpreted with caution - but they can still provide useful information about the approximate magnitude of gene flow (Neigel, 2002).

### 3.3.4 Ordination

Principal Coordinate analysis is a way to describe patterns in data of high dimensions. It has the benefit over other methods which visualise the relationship in data sets – such as tree building methods which assume a hierarchical genetic structure; while this is a reasonable assumption at taxonomic levels above species, it is not always true at the level of population (Smouse, 1998). The PCoA was carried out using a distance matrix of pairwise Nei's genetic distance; this was implemented in the GenAlEx package.

### 3.3.5 Test of isolation by distance using the partial Mantel test

The association between genetic distance and geographic distance was analysed at different geographic scales, firstly over a large scale including sampling sites from Chile and New Zealand, and then at smaller scales first just including sampling sites around New Zealand and finally those samples found in the Southern grouping. Spatial genetic distance was assessed at the population level, by testing the association between pairwise estimates of population genetic differentiation and the natural logarithm of distances (km) between sampling sites. This was conducted using a partial Mantel test of matrix correspondence, using the procedure of Smouse *et al.* (1986). The measure of genetic differentiation ( $F_{st}$ ) was linearized ( $F_{st}/(1-F_{st})$ ) following Rousset (1997). Geographical distance was measured using the minimum coastline distance using Google Earth. The statistical significance was obtained via 999 random permutations of the genetic distance matrix. The partial Mantel test was carried out using GenAlEx.

### 3.3.6 Assignment test

To provide a direct measurement of contemporary dispersal the probability of identity was estimated. This method assumes HWE and no linkage disequilibrium between loci. The test gives estimates of dispersal that are comparable to mark-recapture data (Berry *et al.* 2004). To assess probability of identity the assignment calculator DOH was used (<http://www2.biology.ualberta.ca/jbrzusto/Doh.php>). This test takes genotypes of individuals from several populations and determines from which population each individual is most likely to derive from (Paetkau *et al.* 1997). The individual is assigned to the population in which the individual genotype has the highest expected frequency based on the observed

distributions of alleles (Patkeau *et al.* 1995, 1997). When analysed using all individuals, a probability of correct assignment for each nominal population can be calculated.

### 3.3.7 Population genetic structure

The software package AWclust (Gao and Starmer, 2008) was used to examine the multilocus similarity among individuals and to estimate the probable number of genetically differentiated groups within the data set. This non-parametric method does not make assumptions about HWE or linkage disequilibrium (LD). This approach is useful for molluscs because there have been a number of reports that have found large heterozygote deficiencies in molluscs. AWclust is specifically designed for the analysis of SNPs, but since they are usually bi-allelic they are essentially genotyped in the same way as RAPDs, meaning the RAPD data set can easily be analysed using this software.

The data was first assessed using the MDS facility to confirm there were no outliers. The gap statistic function was then used to test for the presence of groups from 2 to 8 within the data set. The individual-specific assignment information was used to examine the assignment success of this test and to quantify the likely existence of  $n$  groups based on the analysis. Here the *a priori* hypothesis that there is a barrier to gene flow around the Cook Strait around 42°S which would result in a population division of populations north and south of this divide. The data was divided into three groups: 1) All New Zealand populations excluding the Chatham Islands; 2) all New Zealand populations excluding the Chatham Islands and excluding the outlier loci; and 3) all populations.

### 3.4 Results

#### 3.4.1 Genetic diversity

Total number of loci retained for analysis for each primer were *OPC\_09* = 13 bands, *OST\_02* = 9 bands, *OPE\_05* = 10 bands, and *OPE\_02* = 9 bands. Private bands were not detected for any sampling sites. For clarity when referring to the individual phenotype: the PCR products as visualised on the electrophoresis gel are termed as bands, when the bands are interpreted as genotypes they are termed loci each with two alleles ‘dominant’ being the presence of a band and ‘recessive’ being the absence of a band.

A total of 41 loci were scored for 360 individuals across 11 populations (Table 3.2). The average number of bands per individual scored was  $13.3 \pm 4.2$  (mean  $\pm$  standard deviation). The number of polymorphic loci at the 5% level ranged from 92.7% in the Foveaux Strait 2 population to 43.9% in the Stewart Island population. Average Nei’s gene diversity ( $H_j$ ) across all populations was  $0.26 \pm 0.09$ .

#### 3.4.2 Outlier loci

Two loci were shown to have large  $F_{st}$  values that were more than two standard deviations away from the mean. The two loci were *OPE\_02\_100* and *OPE\_05\_2000*. These loci were included in a global analysis; they were then removed in separate analysis to explore their effects.

#### 3.4.3 Spatial genetic analysis

Significant genetic differentiation among populations was revealed ( $F_{st} = 0.194$ ,  $p < 0.0001$ ;  $\Phi_{PT} = 0.243$   $p < 0.0001$ ). The highest pairwise  $\Phi_{PT}$  value of 0.401 (Table 3.3) was between MARL and AKA, the next highest was between MARL and CHI\_2. The lowest values of 0.015 are between FOV\_1 and FOV\_2, the values of between FOV\_1 and AWA and FOV\_2 and AWA are also comparatively low at 0.060 and 0.075 respectively. Pairwise population levels of gene flow ranged from comparatively low 0.373 migrants per generation between AKA and MARL up to comparatively high levels between FOV\_1 and FOV\_2 at 16.406 migrants per generation.

Hierarchical AMOVA analysis was used to assess the level of genetic variation between the *a priori* groupings of northern, eastern, southern and Chilean sampling sites (Table 3.5). The AMOVA analysis revealed most of the variation (76%) was found to be within sampling localities, 22% was found among populations, and the final 2% of variation was revealed to be among the regions.

#### 3.4.4 Partial Mantel test

Results of the partial Mantel test are shown in Figures 3.3a to 3.3h. Tests for associations between genetic divergence ( $(F_{st}/(1-F_{st}))$ ) and the natural log of geographic distance at the population level were carried out over a range of geographical divisions.

Using all geographical locations revealed a statistically non-significant association between geographical distance and genetic distance ( $r_{xy} = 0.265$ ,  $p = 0.120$ ,  $r_2 = 0.0704$ ) (Figure 3.3a). The analysis over the New Zealand scale again revealed an insignificant association ( $r_{xy} = 0.002$ ,  $p = 0.464$ ,  $r_2 = 0.006$ ) (Figure 3.3b). The Chatham Islands sampling site was removed from the analysis and an insignificant association was revealed ( $r_{xy} = 0.064$ ,  $p = 0.3511$ ,  $r_2 = 0.0041$ ) (Figure 3.3c). The small-scale analysis of the southern sample sites revealed an insignificant association with genetic distance and geographic distance ( $r_{xy} = -0.174$ ,  $p = 0.421$ ,  $r_2 = 0.0304$ ) (Figure 3.3d). Outlier loci were removed and the mantel test was conducted using the new pairwise  $F_{st}$  values. Again, there was an insignificant association when all sampling sites were compared ( $r_{xy} = 0.230$ ,  $p = 0.142$ ,  $r_2 = 0.0531$ ) (Figure 3.3e). However, using only New Zealand populations a significant isolation by distance profile was found ( $r_{xy} = 0.324$ ,  $p = 0.030$ ,  $r_2 = 0.1052$ ) (Figure 3.3f). With the Chatham Islands removed there was still a significant isolation by distance profile ( $r_{xy} = 0.341$ ,  $p = 0.043$ ,  $r_2 = 0.1166$ ) (Figure 3.3g). In the southern sampling site the association between geographical and genetic distance was still insignificant ( $r_{xy} = -0.173$ ,  $p = 0.4288$ ,  $r_2 = 0.03$ ). However, the low  $r_2$  values showed that only a small portion of the data (0.06% to 11.6 %) was explained by the regression line.

#### 3.4.5 Population assignment test

Using all loci, the results of the assignment test suggest there is population structuring, given that most of the individuals were assigned to their correct populations (Table 3.6a). 100 % of the Chilean (Chi\_1) individuals were assigned to the correct population. Individuals from Foveux\_1, Foveaux\_2 and Awarua were assigned to the correct population (42%, 41% and

56% respectively) these lower success rates of assigning samples suggest that they should be treated as one population. A very similar pattern was found when the outlier loci were removed, the values of correct assignment remained the same except for Marlborough, which went from 97% to 95% correct assignment, and Stew\_2 which went from 90 to 92% correct assignment (Table 3.6b).

Overall the assignment success was very high: 78.9% for both data sets of all loci and with the outlier loci removed.

### **3.4.6 Ordination**

The first two co-ordinates of the PCoA of Nei's genetic distance explained 57.61% of the variation in the data set (Figure 3.4). Overall the PCoA has not separated the populations into groups that reflect geographical location, that is, populations that have a close geographical affinity are not grouped together in the analysis. The Foveaux Strait and Stewart Island populations, and those that have a large geographical distance between them seem to have clustered together on the PCoA, such as Chile and the Akaroa peninsula. However, the two Foveaux Strait populations (Fov\_1 and Fov\_2) have clustered together with a close association with the Awarua Estuary population, suggesting that they could be grouped together as one population. Also, the Cloudy Bay (CLO) sampling site and Chatham Island sampling site have a close association to the exclusion on the Marlborough Sounds (MARL) sampling site. The same relationship among sampling sites was found when the outlier loci were removed from the analysis so the relationship is not shown.

**Table 3.2:** Summary of randomly amplified polymorphic DNA marker (RAPD) loci and Nei's gene diversity for each sampling site.

<i>Sampling site</i>	<i>Population size</i>	<i>Number of polymorphic loci</i>	<i>Number of bands scored per individual (mean <math>\pm</math> SD)</i>	<i>H (Nei's gene diversity mean <math>\pm</math> SD)</i>
<b><i>Northern group</i></b>				
Cloudy	24	33	11.931 $\pm$ 3.316	0.255 $\pm$ 0.024
Marlborough	37	27	14.205 $\pm$ 2.419	0.243 $\pm$ 0.029
<b><i>Eastern group</i></b>				
Chathams	35	33	14.051 $\pm$ 4.279	0.269 $\pm$ 0.028
Akaroa	22	30	13.667 $\pm$ 3.017	0.270 $\pm$ 0.028
<b><i>Chilean group</i></b>				
Chile_1	23	21	10.542 $\pm$ 1.668	0.182 $\pm$ 0.026
Chile_2	22	30	13.958 $\pm$ 2.596	0.255 $\pm$ 0.027
<b><i>Southern group</i></b>				
Foveaux_1	23	35	17.750 $\pm$ 2.608	0.312 $\pm$ 0.026
Foveaux_2	39	38	17.667 $\pm$ 4.453	0.398 $\pm$ 0.020
Stewart_1	34	18	7.386 $\pm$ 2.255	0.171 $\pm$ 0.028
Stewart_2	33	33	11.795 $\pm$ 3.533	0.252 $\pm$ 0.025
Awarua	35	33	14.077 $\pm$ 3.770	0.283 $\pm$ 0.028

**Table 3.3a:** Global data set,  $\Phi_{PT}$  values below diagonal based on 9999 permutations is shown below the diagonal. Number of migrants ( $N_m$ ) is shown above the diagonal.  $\Phi_{PT}$  values given in bold were not statistically significant after false discovery correction for multiple testing.

	<b>CLO</b>	<b>MARL</b>	<b>CHAT</b>	<b>CHI_1</b>	<b>AKA</b>	<b>CHI_2</b>	<b>FOV_1</b>	<b>FOV_2</b>	<b>STEW_1</b>	<b>STEW_2</b>	<b>AWA</b>
<b>CLO</b>	-	0.676	1.232	0.904	0.632	0.577	1.193	1.083	0.684	1.131	2.764
<b>MARL</b>	0.27	-	0.567	0.443	0.373	0.385	0.545	0.446	0.437	0.967	0.769
<b>CHAT</b>	0.169	0.306	-	1.014	0.535	0.567	1.179	0.937	0.764	1.243	1.754
<b>CHI_1</b>	0.217	0.361	0.198	-	0.758	0.986	0.987	0.97	0.851	0.94	1.131
<b>AKA</b>	0.284	0.401	0.319	0.248	-	0.581	0.579	0.485	0.517	0.721	0.69
<b>CHI_2</b>	0.302	0.394	0.306	0.202	0.301	-	0.814	0.825	0.819	0.822	0.825
<b>FOV_1</b>	0.173	0.315	0.175	0.202	0.302	0.235	-	16.406	0.517	0.927	3.903
<b>FOV_2</b>	0.188	0.359	0.211	0.205	0.34	0.233	<b>0.015</b>	-	0.516	0.814	3.066
<b>STEW_1</b>	0.268	0.364	0.246	0.227	0.326	0.234	0.326	0.326	-	0.834	0.797
<b>STEW_2</b>	0.181	0.205	0.167	0.21	0.258	0.233	0.212	0.235	0.231	-	1.347
<b>AWA</b>	0.083	0.245	0.125	0.181	0.266	0.233	0.06	0.075	0.239	0.156	-

**Table 3.3b:** Outlier loci removed,  $\Phi$ PT values below diagonal based on 9999 permutations is shown below the diagonal. Number of migrants ( $N_m$ ) is shown above the diagonal.  $\Phi$ PT values given in bold were not statistically significant after false discovery correction for multiple testing.

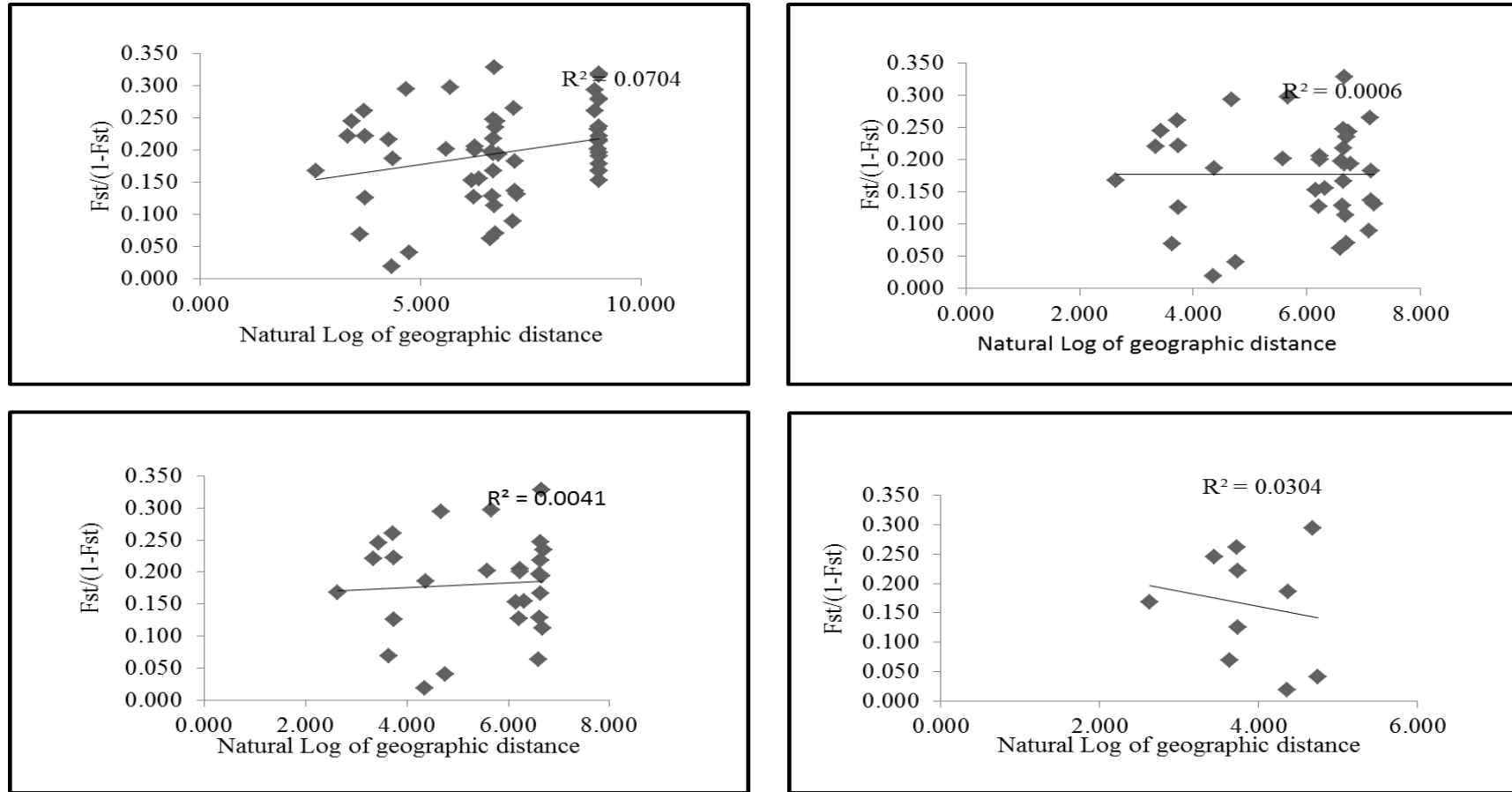
	CLO	MARL	CHAT	CHI_1	AKA	CHI_2	FOV_1	FOV_2	STEW_1	STEW_2	AWA
CLO		0.748	1.279	0.651	0.904	0.578	1.267	1.147	0.698	1.114	2.919
MARL	0.251		0.615	0.479	0.49	0.428	0.639	0.511	0.494	1.031	0.905
CHAT	0.164	0.289		0.604	0.996	0.559	1.172	0.93	0.757	1.219	1.734
CHI_1	0.278	0.343	0.293		0.925	0.691	0.67	0.563	0.594	0.814	0.787
AKA	0.217	0.338	0.201	0.213		0.973	0.975	0.963	0.841	0.928	1.099
CHI_2	0.302	0.369	0.309	0.266	0.204		0.807	0.822	0.813	0.819	0.81
FOV_1	0.165	0.281	0.176	0.272	0.204	0.236		16.522	0.513	0.93	3.921
FOV_2	0.179	0.329	0.212	0.308	0.206	0.233	<b>0.015</b>		0.514	0.815	3.092
STEW_1	0.264	0.336	0.248	0.296	0.229	0.235	0.328	0.327		0.837	0.785
STEW_2	0.183	0.195	0.17	0.235	0.212	0.234	0.212	0.235	0.23		0.837
AWA	0.079	0.216	0.126	0.241	0.185	0.236	0.06	0.075	0.242	0.158	

**Table 3.4:** Pairwise  $F_{st}$  between sampling sites using global data set below the diagonal and with outlier loci removed above the diagonal.

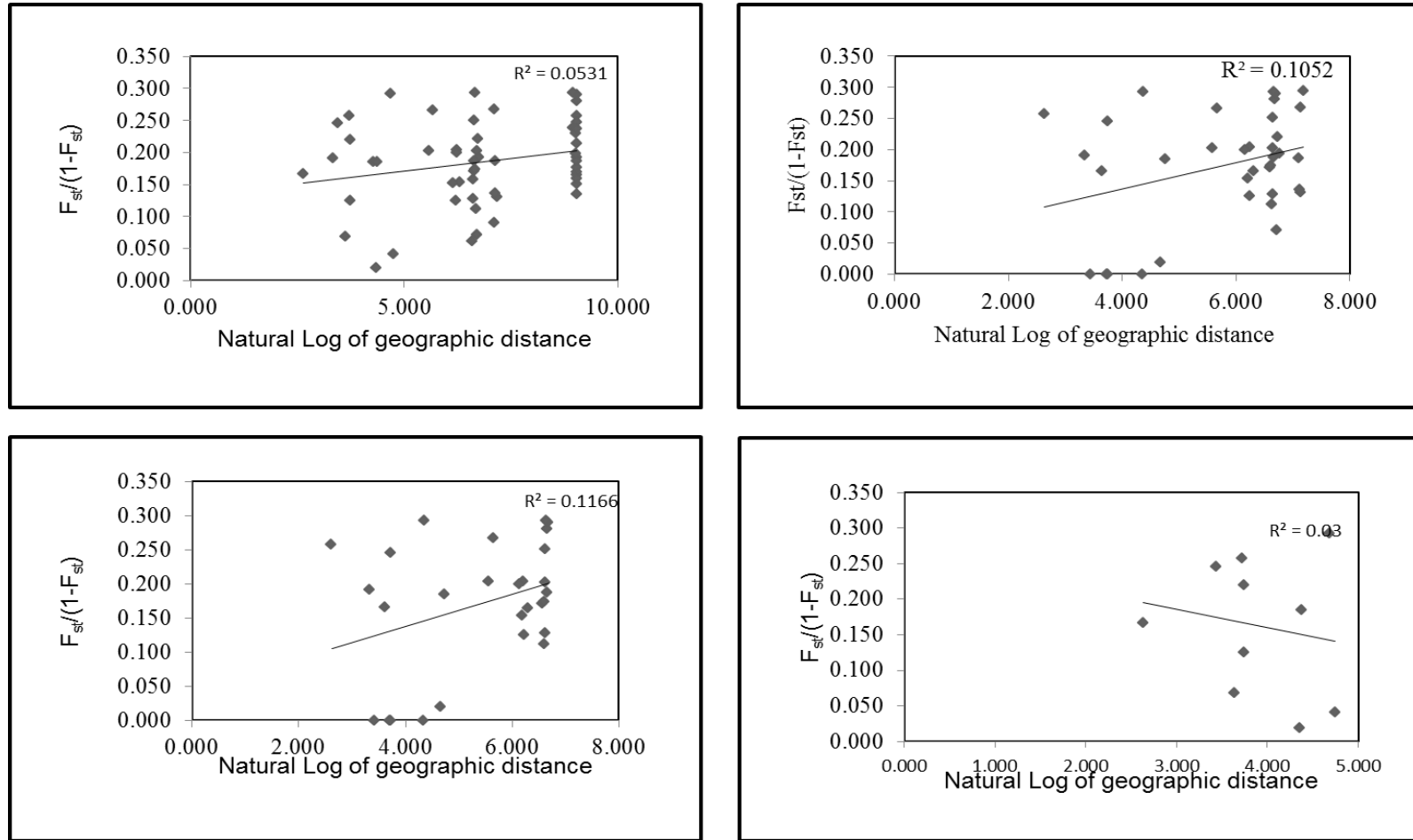
	CLO	MARL	CHAT	CHI_1	AKA	CHI_2	FOV_1	FOV_2	STEW_1	STEW_2	AWA
CLO		0.192	0.071	0.215	0.203	0.281	0.112	0.128	0.251	0.171	0.062
MARL	0.221		0.221	0.257	0.267	0.29	0.202	0.188	0.293	0.175	0.157
CHAT	0.071	0.244		0.239	0.193	0.294	0.131	0.136	0.267	0.187	0.09
CHI_1	0.222	0.316	0.261		0.136	0.186	0.169	0.193	0.248	0.187	0.16
AKA	0.202	0.297	0.193	0.168		0.165	0.153	0.125	0.204	0.2	0.152
CHI_2	0.278	0.319	0.293	0.216	0.167		0.176	0.151	0.231	0.237	0.199
FOV_1	0.113	0.235	0.131	0.196	0.155	0.178		0.02	0.292	0.185	0.041
FOV_2	0.128	0.218	0.136	0.214	0.127	0.153	0.019		0.258	0.166	0.068
STEW_1	0.247	0.328	0.265	0.279	0.205	0.232	0.294	0.261		0.246	0.221
STEW_2	0.167	0.194	0.183	0.216	0.2	0.237	0.186	0.168	0.245		0.125
AWA	0.063	0.197	0.09	0.19	0.153	0.201	0.041	0.069	0.222	0.126	

**Table 3.5:** AMOVA using binary distance matrix using 4 regions (Northern, Eastern, Southern and Chilean)

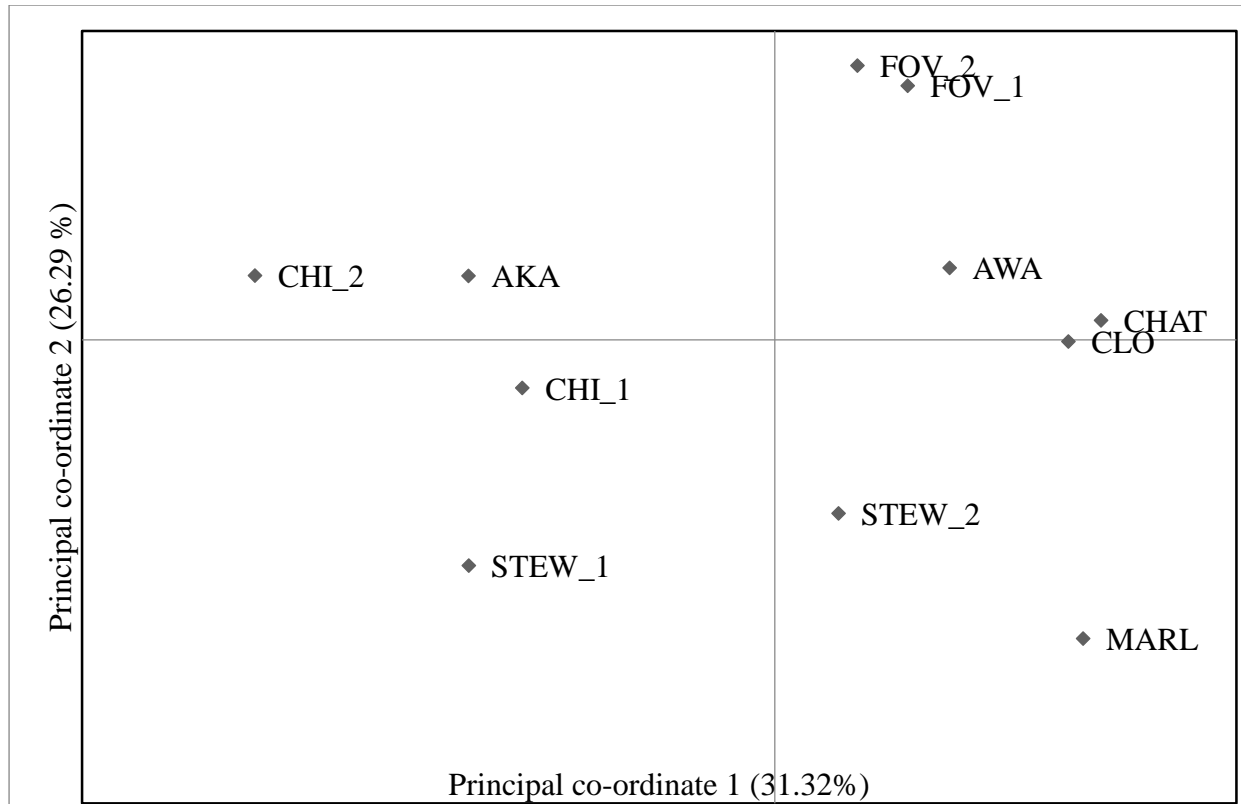
Source of variation	df	Sums of Squares	Mean Variance	Estimated Variance	% Variation	$\Phi$ statistics	P*	
Among Regions	3	286.884	95.628	0.264	3%	$\Phi_{RT}$	0.027	0.001
Among Pops	7	540.787	77.255	2.125	22%	$\Phi_{PR}$	0.222	0.001
Within Pops	348	2587.11	7.434	7.434	76%	$\Phi_{PT}$	0.243	0.001
Total	358	3414.79		9.824	100%			



Figures show Mantel test of Geographic distance vs ( $F_{st}/(1-F_{st})$ ) genetic distance **Figure 3.3a (top left):** All samples **Figure 3.3b (top right):** New Zealand only samples. **Figure 3.3c (bottom left):** New Zealand only samples not including Chatham Islands samples. **Figure 3.3d (bottom right):** Southern populations.



Figures show Mantel test of Geographic distance vs  $((F_{st}/(1-F_{st})))$  genetic distance **Figure 3.3e (top left):** All sampling sites with outlier loci removed. **Figure 3.3f (top right):** New Zealand sampling sites with outlier loci removed. **Figure 3.3g (bottom left):** New Zealand minus the Chatham Islands sampling sites with outlier loci removed. **Figure 3.3h (bottom right):** Southern sampling sites with outlier loci removed.



**Figure 3.4:** Global data set. Nei's genetic distance Principal Co-ordinate analysis

**Table 3.6a:** Global data set. Assignment test results for individuals from each population. Values are percentage of from each collection site

	CLO	MARL	CHAT	CHI_1	AKA	CHI_2	FOV_1	FOV_2	STEW_1	STEW_2	AWA
CLO	<b>76</b>	0	3	0	3	0	7	0	0	3	7
MARL	0	<b>97</b>	3	0	0	0	0	0	0	0	0
CHAT	10	0	<b>79</b>	0	3	0	0	0	0	5	3
CHI_1	0	0	0	<b>100</b>	0	0	0	0	0	0	0
AKA	0	0	0	0	<b>96</b>	0	0	4	0	0	0
CHI_2	0	0	0	0	0	<b>96</b>	0	4	0	0	0
FOV_1	4	0	0	0	0	0	<b>42</b>	33	0	0	21
FOV_2	0	0	0	0	0	0	46	<b>41</b>	0	0	13
STEW_1	3	0	0	0	0	3	0	0	<b>95</b>	0	0
STEW_2	0	3	3	0	0	0	3	0	0	<b>90</b>	3
AWA	15	0	8	0	0	0	15	3	0	3	<b>56</b>

**Table 3.6b:** Outlier loci removed. Assignment test results for individuals from each population. Values are percentage of from each collection site. Correct assignments are in bold.

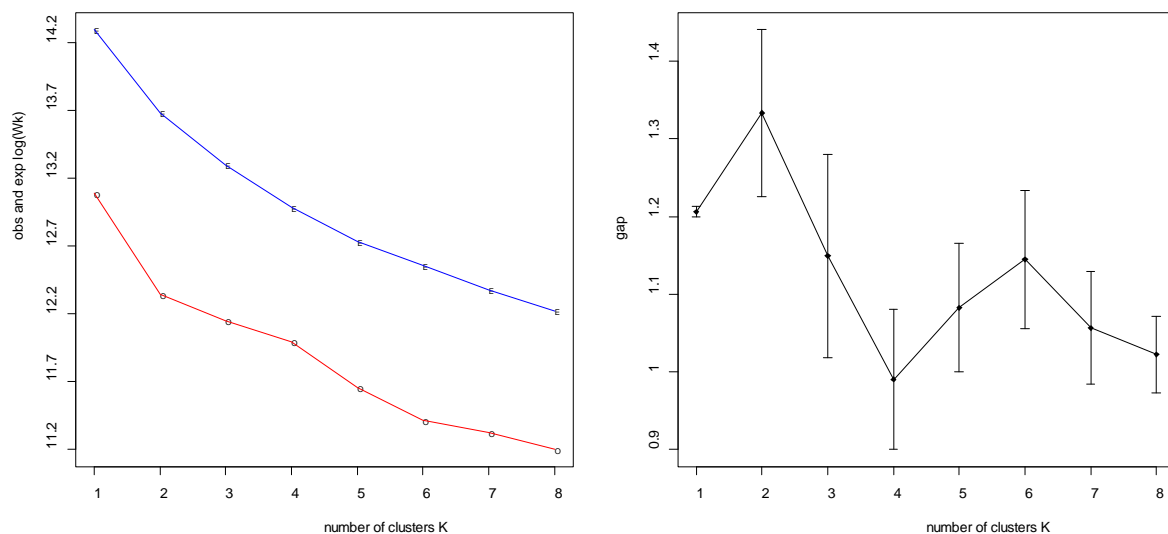
	CLO	MARL	CHAT	CHI_1	AKA	CHI_2	FOV_1	FOV_2	STEW_1	STEW_2	AWA
CLO	<b>76</b>	0	7	0	3	0	7	0	0	3	3
MARL	3	<b>95</b>	0	0	0	0	0	3	0	0	0
CHAT	8	0	<b>79</b>	0	3	0	0	0	0	5	5
CHI_1	0	0	0	<b>100</b>	0	0	0	0	0	0	0
AKA	0	0	0	0	<b>96</b>	0	0	4	0	0	0
CHI_2	0	0	0	0	0	<b>96</b>	0	4	0	0	0
FOV_1	4	0	0	0	0	0	<b>42</b>	33	0	0	21
FOV_2	0	0	0	0	0	0	46	<b>41</b>	0	0	13
STEW_1	3	0	0	0	0	3	0	0	<b>95</b>	0	0
STEW_2	0	3	3	0	0	0	0	0	0	<b>92</b>	3
AWA	15	0	8	0	0	0	15	3	0	3	<b>56</b>

### 3.4.7 Population Structure

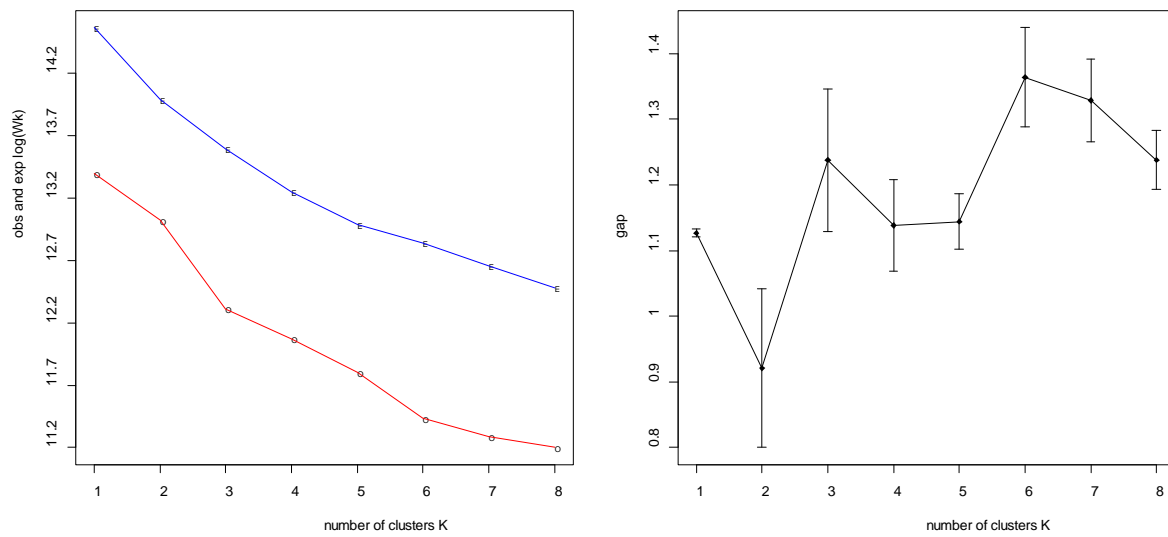
The AWclust nonparametric analysis of possible groups within the multilocus genotype data set for all individuals in all sampling sites, excluding the Chilean and Chatham Island sampling sites, revealed the highest gap value (1.4) at  $K = 2$  (Figure 3.5a). This analysis indicates the presence of two groups. However, the results do not support the *a posteriori* hypothesis of a barrier to gene flow in the Cook Strait region around  $42^{\circ}\text{S}$ . Although most of the populations do split into a grouping that suggests there is a barrier as shown in Table 3.10 the two Stewart Island populations cluster with the two populations above the  $42^{\circ}\text{S}$  divide.

With the outlier loci removed the number of clusters with the highest gap statistic was 6 (Figure 3.5b).

Using the whole data set revealed the highest gap value (1.3) at  $K = 2$  (Figure 3.7). This division is largely attributed to a Chilean grouping and a New Zealand grouping. (Table 3.11)



**Figure 3.5a:** Number of populations assigned by the AWclust analysis using global data set



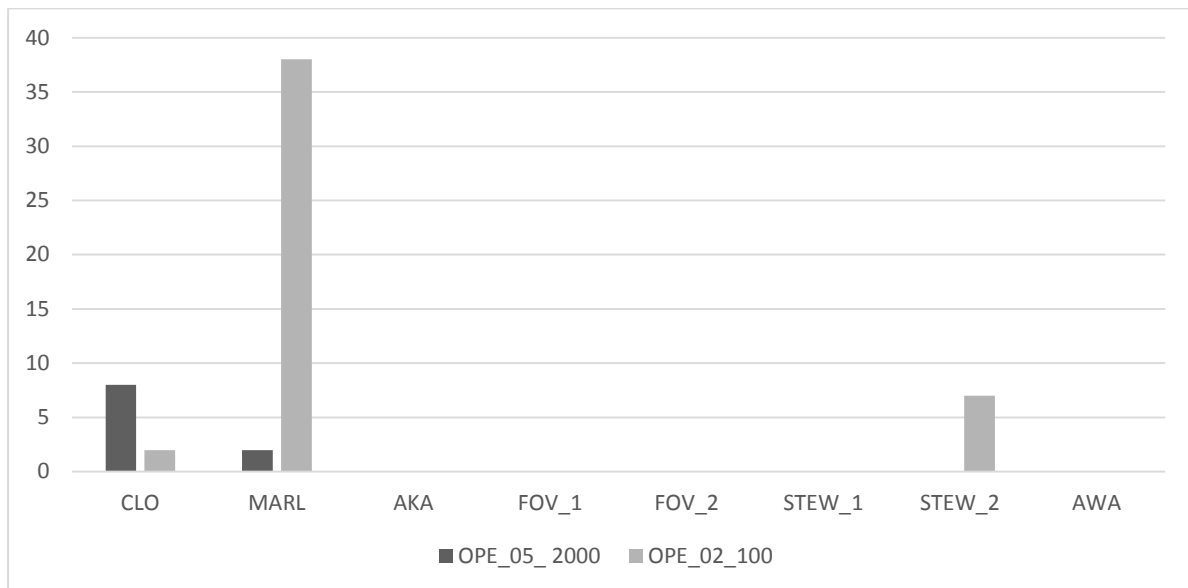
**Figure 3.5b:** Number of populations assigned by the AWclust analysis when outlier loci are removed

**Table 3.10:** Percentage of individuals assigned to each cluster. Correct a priori assignment in bold

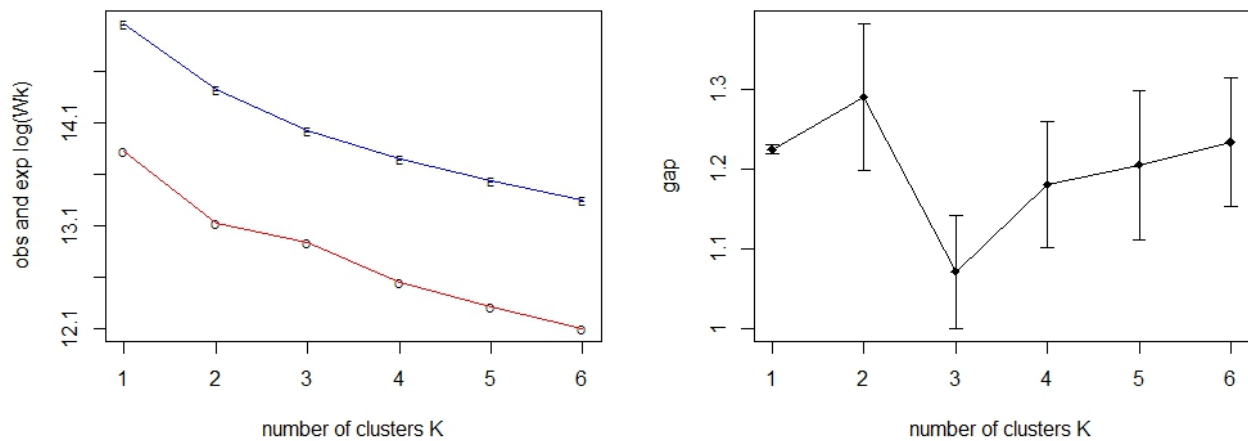
Population	Assigned to Population 1 (%)	Assigned to Population 2 (%)
<i>Above 42°S</i>	<b>72.4</b>	27.6
Cloudy Bay		
Marlborough Sounds	<b>92.6</b>	7.4
<i>Below 42°S</i>		
Akaroa peninsula	0	<b>100</b>
Foveaux Strait 1	0	<b>100</b>
Foveaux Strait 2	2.6	<b>97.4</b>
Stewart Island 1	100	<b>0</b>
Stewart Island 2	92	<b>8</b>
Awarua Estuary	0	<b>96.7</b>

**Table 3.11:** Assignment to the most parsimonious cluster of k=3 produced without the outlier loci included

Population	Assigned to pop 1 (%)	Assigned to Pop 2 (%)	Assigned to Pop 3(%)
<i>Above 42°S</i>			
Cloudy Bay	34	52	14
Marlborough Sounds	87	5	8
<i>Below 42°S</i>			
Akaroa peninsula	100	0	0
Foveaux Strait 1	100	0	0
Foveaux Strait 2	0	0	100
Stewart Island 1	0	0	0
Stewart Island 2	0	100	0
Awarua Estuary	0	0	100



**Figure 3.6:** Frequency of band presence of outlier loci by population



**Figure 3.7:** Number of populations assigned by the AWclust using the whole data set

**Table 3.12:** Assignment to the most parsimonious cluster of k=2 using the whole data set

Population	Assigned to Population 1 (%)	Assigned to Population 2 (%)
Cloudy Bay	79	21
Marlborough Sounds	97	3
Chatham Island	31	69
Akaroa peninsula	0	100
Chile_1	25	75
Chile_2	0	100
Foveaux Strait 1	96	4
Foveaux Strait 2	97	3
Stewart Island 1	100	0
Stewart Island 2	97	3
Awarua Estuary	72	28

### 3.5 Discussion

In any population genetic investigation whether on land or in the sea, there are three models that are used explain the apparent structure: (1) Panmixia; (2) isolation-by-distance, and (3) population divergence within or among populations. The outcome of the present analysis largely fits with the third model, since there is evidence of significant differences between populations, and only a weak isolation by distance profile was found among the New Zealand sampling sites when the outlier loci were removed. There is significant and comparatively high levels of population genetic structuring ( $F_{st} = 0.1941$ ,  $p < 0.0001$ ;  $\Phi_{PT} = 0.243$   $p < 0.0001$ ). In a similar study using RAPDs to investigate the population genetics of *O. chilensis* in Chile, Toro and Gonzalez (2007) found the inter-population genetic diversity of *O. chilensis* was  $F_{st} = 0.2846$ . In both the present study and that of Toro and Gonzalez, (2007) the inter-population genetic diversity is approximately one order of magnitude larger than has been reported for the sister taxa *Ostrea edulis* (the European oyster), which has been shown to be around  $F_{st} = 0.019$ , (Launey, 2001). This difference has been attributed to the greatly reduced PLD in *O. chilensis*, which could cause reduced gene flow and increased inbreeding leading to genetically distinct populations. With the outlier loci removed the population differentiation was still comparatively high ( $F_{st} = 0.1834$ ,  $p < 0.0001$ ;  $\Phi_{PT} = 0.238$ ,  $p < 0.0001$ ).

Gene flow between the Foveaux Strait sampling sites and the geographically close Awarua Estuary sampling sites was high (Table 3.3a) (Between  $N_m = 16.4$  and  $N_m = 3.06$ ), this indicates that gene flow is reducing the effect of genetic drift, preventing differentiation among populations (Slatkin, 1987). Between most of the sampling sites the  $N_m < 1$ , therefore, the effects of genetic drift and mutation are not balanced out by gene flow - over generations there would be an accumulation of genetic differentiation. From a genetic standpoint, the concept of stock is defined as a population of individuals that are part of the same reproductive process, and genetically distinct from other populations of the same species (Shaklee and Bentzen, 1998). Although there are only two sampling sites in the Foveaux Strait the finding of high gene flow suggests they could be treated as the same genetic stock. In other sampling sites around New Zealand gene flow is much lower, suggesting the presence of genetically distinct stocks.

The results of the Mantel test found no association between genetic difference and geographic distance across most of the divisions. When the outlier loci were removed an IBD

profile was found in the New Zealand only samples (Figure 3.3f), this was also found when the Chatham Islands samples were removed (Figure 3.3g). However, the  $R^2$  values are very low around 0.06 to 0.02, only a very small proportion of the variation in the data set can be explained by the regression line. Perhaps this could be resolved using greater sampling in new locations, particularly in regions between the sample sites to improve the geographic spread of sites - this would better fit the stepping stone model which is a requirement for an IBD profile. Historically, *O. chilensis* spat has been moved around New Zealand this could have disrupted the historical patterns of differentiation. Loss of genetic signal due to historical movements of Chilean populations of *O. chilensis* was attributed to the lack of geographical structure found by Toro and Gonzalez, (2007).

The PCoA was used to visualise the relationship between pairwise genetic distances of sampling locations. Some of the sampling locations cluster in a pattern that is not congruent with geographical location, for example the Akaroa peninsula has clustered with the Chilean samples, and the two Stewart Island populations are far away from their geographic neighbours in the Foveaux Strait, also the Cloudy Bay sampling site and the Marlborough Sounds sampling site have not clustered close to each other. However, there is a pattern that correlates with geographic location of some of the populations. The two Foveaux Strait populations and the geographically close population the Awarua Estuary have clustered together. Between the two Foveaux Strait populations the estimated  $N_m$  is three orders of magnitude higher than observed for other pairwise comparisons; also the assignment test did not perform well between these two populations - suggesting that they should be treated as one population. An explanation for the apparent incongruence between genetic structure and geographic location is the phenomenon of chaotic genetic patchiness. Further study is needed here to assess genetic structure on a temporal scale or to compare the population genetic structure to environmental variables in a seascape genetics framework – if a different population structure were to be found it is likely to be result of stochastic processes.

The results of the AWClust analysis (Figure 3.5a) revealed two distinct groups. Table 3.10 presents the division of the populations into these two groups, this grouping is attributed to a north/south divide of the South Island, with the exception of the Stewart Island populations that cluster with the northern Samples - despite being the southernmost New Zealand populations. This is not consistent with a larger scale but similar study by Wei *et al.* (2013a), a north/south divide among populations of *Perna canaliculus* was found, here the genetic subdivision was found to occur at South of the Cook Strait around 42°S. Many population

genetic studies in New Zealand have documented the existence of this particular discontinuity (as reviewed by Ross *et al.* 2009; Gardner *et al.* 2010). This study has not identified the barrier to gene flow. The possible reasons for this are that the barrier to gene flow is due to historical influences and *O. chilensis* populations have reached equilibrium; or the barrier to gene flow is still present but cannot be detected due to human mediated stock movements. When the outlier loci were removed from the analysis (OPE\_02\_100 and OPE\_05\_2000), a different geographical pattern was found (Table 3.11 and Figure 3.5b). The frequency of band presence (that is amplified product) for the outlier loci is plotted in Figure 3.6, the division into two groups is largely due to the presence of these particular bands. Interestingly, Wei *et al.* (2013a) found that the genetic discontinuity was present only when a locus that was thought to be under selection was included, suggesting that the genetic discontinuity was due to a selection gradient rather than a physical barrier to gene flow. Using all populations (Figure 3.7) (the inclusion of the Chilean populations and the Chatham Islands) the AWClust analysis revealed two groups: most of the individuals appeared in group one, while 100% of the Chilean samples appeared in group two – demonstrating that the Chilean populations are divergent from the New Zealand populations. Although the barrier to gene flow has not been found in this study – a question that still needs to be addressed is why does it occur in some species and not others? A future research focus could seek to determine whether the multi-species pattern of genetic discontinuity is due to the same selection pressures, if any. Reciprocal transplant of organisms found above and below the divide could be a way to test for local adaptation. Addressing this question is important in order to increase understanding of community ecology, the impacts of climate change, and the spatial management of marine ecosystems (Sanford and Kelly, 2011). A further hypothesis to explain this division observed in multiple species, is that the discontinuity is due to secondary contact of glacial refugia, if this is the case then it would be likely to see a similar pattern in other organisms occurring in this locality.

The analyses showed that there was significant population structure with a high level of correctly assigned individuals to populations. This population genetic structuring in some cases reflects the geographical distribution of the species, but in other regions the results were unexpected. One possible explanation for this is that oyster spat has been moved around New Zealand to form harvestable populations, which would explain the lack of a geographic pattern. The assignment tests tend to reflect contemporary patterns of gene flow; a high level of correct assignment would suggest that there is reduced contemporary gene flow. Low

levels of gene flow could be the result of the very short PLD. This apparent genetic structure seems to be largely the result of self-recruitment in populations, when the individual genotypes are randomised into 3, 6 and 8 populations panmixia was found (Appendix 3.4 to 3.9) - supporting the hypothesis of population structure.

The Stewart Island populations do not seem to cluster with geographical neighbouring populations (Figure 3.4). This could be the result of stochastic processes or human mediated influences. The impact of stochastic influences of the apparent genetic structure should in future studies, be tested by conducting temporal sampling. A potential problem that may have affected some sampling sites is the presence of cryptic speciation; this could be an issue in the Stewart Island population that were collected in the shallow intertidal region, this is an area that is not a commercial fishery and perhaps could be a sympatric cryptic species. It is not uncommon for cryptic mollusc species to be found living in sympatry. For example Baker *et al.* (2003) found cryptic speciation had occurred in sympatrically occurring freshwater molluscs. Cryptic speciation would add to the difficulty of identifying discrete stocks for management, this has been a problem in well-studied commercially important species for example, the squids of the genus *Doryteuthis* (Sales *et al.* 2013). This presence of cryptic speciation in Stewart Island could easily be tested by sequencing mitochondrial genes and testing them in a phylogenetic context with other oyster populations.

The development of genetic markers such as randomly amplified polymorphic DNA (RAPD), short tandem repeats (STR or microsatellites) and single nucleotide polymorphisms (SNP) among others, are increasingly being applied to aquaculture research for not only measuring levels of genetic variability between populations, but also to assess levels of inbreeding, parentage assignment, species and strain identification. In the present study, the objective was to assess the levels of genetic variation among natural populations of *O. chilensis*, it should be taken as a starting point for future genetic monitoring in the same localities these findings will assist effective management programmes of this species.

### 3.6 Summary

1. The brooding life strategy of *Ostrea chilensis* appears to have resulted in reduced gene flow between populations compared to the sister taxa *Ostrea edulis*. This was first documented by Toro and Gonzalez (2007) in Chilean populations, and is supported here. The brooding strategy and limited gene flow could explain the limited distribution of *Ostrea chilensis* in New Zealand.
2. The presence of a divide around 42°S in the Tasman Strait has been documented in a number of New Zealand studies. Its presence has been not been found in the present study. This could be due to a number of reasons: the barrier is due to historical influences and *O. chilensis* has reached equilibrium frequencies, or the apparent genetic structure is due to human mediated stock movements.
3. More work is needed to investigate the potential of chaotic genetic patchiness, or to compare the genetic structure to environmental variables in a seascape genetics framework. The seascape genetics will be addressed in the preceding chapter.

### 3.7 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(9), 1617-1628.
- Apte S, Star B, Gardner JPA. (2003). A comparison of genetic diversity between cultured and wild populations, and a test for genetic introgression in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791). *Aquaculture*, 219(1), 193-220.
- Arnaud-Haond S, Vonau V, Rouxel C, Bonhomme F, Prou J, Goyard E, Boudry P. (2008). Genetic structure at different spatial scales in the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesian lagoons: beware of sampling strategy and genetic patchiness. *Marine Biology*, 155(2), 147-157.
- Baker AM, Bartlett C, Bunn SE, Goudkamp K, Sheldon F, Hughes JM. (2003). Cryptic species and morphological plasticity in long-lived bivalves (Unionoida: Hyriidae) from inland Australia. *Molecular Ecology*, 12(10), 2707-2717.
- Berry O, Tocher MD, Sarre SD. (2004). Can assignment tests measure dispersal? *Molecular Ecology*, 13(3), 551-561.
- Blanquer A, Uriz M J. (2010). Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC evolutionary biology*, 10(1), 13.
- Buroker NE. (1985). Evolutionary patterns in the family Ostreidae: larviparity vs. oviparity. *Journal of experimental marine biology and ecology*, 90(3), 233-247.
- Chaparro OR. (1990). Effect of temperature and feeding on conditioning of *Ostrea chilensis* Philippi, 1845 reproducers. *Aquaculture Research*, 21(4), 399-405.
- Cole HA. (1941). The fecundity of *Ostrea edulis*. *Journal of the Marine Biological Association of the United Kingdom*, 25(02), 243-260.
- Cowen RK, Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, 1, 443-466.
- Cranfield HJ, Michael KP. (1989). Larvae of the incubatory oyster *Tiostrea chilensis* (Bivalvia: Ostreidae) in the plankton of central and southern New Zealand. *New Zealand journal of marine and freshwater research*, 23(1), 51-60.

- Cranfield HJ. (1968). An unexploited population of oysters, *Ostrea lutaria* Hutton, from Foveaux Strait: Part I. Adult stocks and spatfall distribution. *New Zealand journal of marine and freshwater research*, 2(1), 3-22.
- DiSalvo LH, Alarcon E, Martinez E. (1983). Induced spat production from *Ostrea chilensis* Philippi 1845 in mid-winter. *Aquaculture*, 30(1), 357-362
- Excoffier L, Smouse PE, Quattro JM. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491.
- Faurby S, Barber PH. (2012). Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology*, 21(14), 3419-3432.
- Gaffney PM, Scott TM, Koehn RK, Diehl WJ. (1990). Interrelationships of heterozygosity, growth rate and heterozygote deficiencies in the coot clam, *Mulinia lateralis*. *Genetics*, 124(3), 687-699.
- Gao X, Starmer JD. (2008). AWclust: point-and-click software for non-parametric population structure analysis. *BMC bioinformatics*, 9(1), 77.
- Gardner JPA, Bell JJ, Constable HB, Hannan D, Ritchie PA, Zuccarello GC. (2010). Multi-species coastal marine connectivity: a literature review with recommendations for further research: Victoria University of Wellington report (ZBD2009-10) to the New Zealand Ministry of Fisheries. Wellington: New Zealand Ministry of Fisheries. p.50.
- Goldstein SJ, Gemmell NJ, Schiel D R. (2006). Molecular phylogenetics and biogeography of the nacellid limpets of New Zealand (Mollusca: Patellogastropoda). *Molecular phylogenetics and evolution*, 38(1), 261-265
- Grantham BA, Eckert GL, Shanks AL. (2003). Dispersal potential of marine invertebrates in diverse habitats. *Ecological Applications*, 13(sp1), 108-116.
- Hedgecock, D. (1994). Temporal and spatial genetic structure of marine animal populations in the California Current. *California Cooperative Oceanic Fisheries Investigations Reports*, 35, 73-81.
- Hellberg ME, Burton RS, Neigel JE, Palumbi SR. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, 70(Supplement 1), 273-290.

- Highsmith RC. (1985). Floating and algal rafting as potential dispersal mechanisms in brooding invertebrates. *Marine ecology progress series*. 25(2), 169-179.
- Isabel N, Beaulieu J, Thériault P, Bousquet J. (1999). Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. *Molecular Ecology*, 8(3), 477-483
- Johnson MS, Black R. (1984). Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution*, 1371-1383.
- Krauss SL. (2000). Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology*, 9(9), 1241-1245.
- Larmuseau MH, Vancampenhout KIM, Raeymaekers JA, Van Houdt JK, Volckaert FA. (2010). Differential modes of selection on the rhodopsin gene in coastal Baltic and North Sea populations of the sand goby, *Pomatoschistus minutus*. *Molecular ecology*, 19(11), 2256-2268.
- Launey S, Barre M, Gerard A, Naciri-Graven Y. (2001). Population bottleneck and effective size in *Bonamia ostreae*-resistant populations of *Ostrea edulis* as inferred by microsatellite markers. *Genetical research*, 78(03), 259-270.
- Levin LA. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, 46(3), 282-297.
- Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Carvalho, GR. (2009). Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, 9(1), 276.
- Ó Foighil D, Marshall BA, Hilbish TJ, Pino MA. (1999). Trans-Pacific range extension by rafting is inferred for the flat oyster *Ostrea chilensis*. *Biological Bulletin*, 122-126.
- Paetkau D, Strobeck C. (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular ecology*, 4(4), 519-520.
- Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck, C. (1997). An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, 147(4), 1943-1957.

- Palumbi SR. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological applications*, 13(sp1), 146-158.
- Peakall R, Smouse PE. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6, 288-295.
- Peakall R, Smouse PE. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537-2539.
- Puritz JB, Toonen RJ. (2011). Coastal pollution limits pelagic larval dispersal. *Nature communications*, 2, 226.
- Riginos C, Douglas KE, Jin Y, Shanahan DF, Trembl EA. (2011). Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, 34(4), 566-575.
- 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(5), 1009–1027.
- Rousset F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145(4), 1219-1228.
- Sales JBDL, Shaw PW, Haimovici M, Markaida U, Cunha DB, Ready J, Sampaio I. (2013). New molecular phylogeny of the squids of the family Loliginidae with emphasis on the genus *Doryteuthis* Naef, 1912: Mitochondrial and nuclear sequences indicate the presence of cryptic species in the southern Atlantic Ocean. *Molecular phylogenetics and evolution*, 68(2), 293-299.
- Sanford E, Kelly MW. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3, 509-535.
- Selkoe KA, Gaines SD, Caselle JE, Warner RR. (2006). Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, 87(12), 3082-3094.
- Selkoe KA, Watson JR, White C, Horin TB, Iacchei M, Mitarai S, Toonen RJ. (2010). Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, 19(17), 3708-3726.

- Shaklee JB, Bentzen, P. (1998). Genetic identification of stocks of marine fish and shellfish. *Bulletin of Marine Science*, 62(2), 589-621.
- 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.
- Slatkin M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787-792.
- Smouse PE, Long JC, Sokal RR. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic zoology*, 627-632.
- Smouse PE. (1998). To tree or not to tree? *Molecular Ecology*, 7(4), 399-412.
- Sponer R, Roy MS. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*, 56(10), 1954-1967.
- Sponer R, Roy MS. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*, 56(10), 1954-1967.
- Thomas L, Bell JJ. (2013). Testing the consistency of connectivity patterns for a widely dispersing marine species. *Heredity*, 111(4), 345-354.
- Thornhill DJ, Mahon AR, Norenburg JL, Halanych KM. (2008). Open-ocean barriers to dispersal: a test case with the Antarctic Polar Front and the ribbon worm *Parborlasia corrugatus* (Nemertea: Lineidae). *Molecular ecology*, 17(23), 5104-5117.
- Toro JE, Aguila PR. (1996). Genetic differentiation of populations of the oyster *Ostrea chilensis* in southern Chile. *Aquatic Living Resources*, 9(01), 75-78.
- Toro JE, González CP. (2009). The genetic structure of the Chilean oyster (*Ostrea chilensis* Philippi, 1845) in natural populations of southern Chile based on RAPDs analysis. *Revista De Biología Marina Y Oceanografía*, 44(2), 467-476.
- Toro JE, Vergara AM. (1996). Multiple-locus heterozygosity, physiology and growth at two different stages in the life cycle of the Chilean oyster *Ostrea chilensis*. *Marine Ecology-Progress Series*, 134, 151-158.

- Vekemans X, Beauwens T, M. Lemair T, Roldan-Ruiz I, 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, 11, 139-151.
- Verhoeven KJ, Simonsen KL, McIntyre LM. (2005). Implementing false discovery rate control: increasing your power. *Oikos*, 108(3), 643-647.
- Waters JM, Roy MS. (2004). Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*, 13(9), 2797-2806.
- Weersing K, Toonen R J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, 393(1), 12.
- Wei K, Wood AR, Gardner JPA (2013a). Population genetic variation in the New Zealand greenshell mussel: locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. *Marine biology*, 160(4), 931-949.
- Whitlock MC, McCauley DE. (1999). Indirect measures of gene flow and migration:  $F_{st} \neq 1/(4Nm + 1)$ . *Heredity*, 82(2), 117-125
- Winter JE, Toro JE, Navarro JM, Valenzuela GS, Chaparro OR. (1984). Recent developments, status, and prospects of molluscan aquaculture on the Pacific coast of South America. *Aquaculture*, 39(1), 95-134.
- Wood AR, Gardner JPA (2007). Small spatial scale population genetic structure in two limpet species endemic to the Kermadec Islands, New Zealand. *Marine Ecology Progress Series*, 349, 159-170.
- Wright S. (1943). Isolation by distance. *Genetics*, 28(2), 114.
- Yund PO. (1990). An in situ measurement of sperm dispersal in a colonial marine hydroid. *Journal of Experimental Zoology*, 253(1), 102-106.
- Zhivotovsky LA. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, 8(6), 907-913.

### 3.8 Appendix

**Appendix 3.1:** RAPD primers tested those in bold were kept for further analysis

RAPD Primer	Sequence	Band Pattern
OST 1	GGGCGCGAGT	monomorphic bands
<b>OST 2</b>	<b>GCGGCTGGAG</b>	<b>polymorphic bands</b>
OST 3	TAGCCCGCTT	monomorphic bands
OST 4	AACACACGAG	monomorphic bands
OP C-01	TTCGAGCCAG	weak amplification
OP C-02	GTGAGGCGTC	weak amplification
OP C-03	GGGGGTCTTT	bands not clear
OP C-05	GATGACCGCC	bands not clear
OP C-06	GAACGGACTC	bands not clear
OP C-07	GTCCCGACGA	polymorphic bands
OP C-08	TGGACCGGTG	bands not clear
<b>OP C-09</b>	<b>CTCACCGTCC</b>	<b>polymorphic bands</b>
OP C-11	AAAGCTGCGG	bands not clear
OP C-12	TGTCATCCCC	bands not clear
OP C-13	AAGCCTCGTC	bands not clear
OP C-14	TGCGTGCTTG	bands not clear
OP C-15	GACGGATCAG	bands not clear
OP C-16	CACACTCCAG	bands not clear
OP C-17	TTCCCCCAG	bands not clear
OP C-18	TGAGTGGGTG	bands not clear
OP C-19	GTTGCCAGCC	bands not clear
<b>OP E-02</b>	<b>GGTGCGGGAA</b>	<b>polymorphic bands</b>
OP E-03	CCAGATGCAC	polymorphic bands
OP E-04	GTGACATGCC	polymorphic bands
<b>OP E-05</b>	<b>TCAGGGAGGT</b>	<b>polymorphic bands</b>
OP E-06	AAGACCCCTC	bands not clear
OP E-07	AGATGCAGCC	polymorphic bands
OP E-08	TCACCACGGT	monomorphic bands
OPE_09	CTTCACCCGA	monomorphic bands

OPE_10	CACCAGGTGA	monomorphic bands
OPE_11	GAGTCTCAGG	monomorphic bands
OPE_12	TTATCGCCCC	monomorphic bands
OPE_13	CCCGATTTCGG	band not clear
OPE_14	TGCGGCTGAG	monomorphic bands
OPE_15	ACGCACAACC	monomorphic bands
OPE_16	GGTGACTGTG	monomorphic bands
OPE_17	CTACTGCCGT	monomorphic bands
OPE_18	GGACTGCAGA	monomorphic bands
OPI_01	ACCTGGACAC	monomorphic bands
OPI_02	GGAGGAGAGG	monomorphic bands
OPI_03	CAGAAGCCCA	monomorphic bands
OPI_04	CCGCCTAGTC	monomorphic bands
OPI_05	TGTTCCACGG	monomorphic bands
OPI_06	AAGGCGGCAG	monomorphic bands
OPI_07	CAGCGACAAG	monomorphic bands
OPI_08	TTTGCCCGGT	monomorphic bands
OPI_09	TGGAGAGCAG	monomorphic bands
OPI_10	ACAACGCGAG	monomorphic bands
OPI_11	ACATGCCGTG	monomorphic bands
OPI_12	AGAGGGCACA	monomorphic bands
OPI_13	CTGGGGCTGA	bands not clear
OPI_14	TGACGGCGGT	monomorphic bands
OPI_15	TCATCCGAGG	monomorphic bands
OPI_16	TCTCCGCCCT	monomorphic bands
OPI_17	GGTGGTGATG	monomorphic bands
OPI_18	TGCCCAGCCT	bands not clear
OPI_19	AATGCGGGAG	monomorphic bands
OPI_20	AAAGTGCGGG	monomorphic bands
OPG_06	GTGCCTAACC	monomorphic bands
OPG_08	TCACGTCCAC	monomorphic bands

**Appendix 3.2:** Allele frequency estimates using the Bayesian method of Zhivotovsky (1999). ‘a’ represents the ‘dominant’ allele (presence of a band) and ‘b’ represents the ‘recessive’ allele (absence of a band).

Allele		CLO	MAR L	CHAT	CHI_1	AKA	CHI_2	FOV_ 1	FOV_ 2	STEW _1	STEW _2	AWA
OPC_ 09_16 00	a	0.2	0.005	0.138	0.184	0.032	0.01	0.142	0.256	0.051	0.069	0.009
	b	0.8	0.995	0.862	0.816	0.968	0.99	0.858	0.744	0.949	0.931	0.991
OPC_ 09_14 00	a	0.045	0.005	0.07	0.005	0.306	0.032	0.075	0.256	0.005	0.009	0.022
	b	0.955	0.995	0.93	0.995	0.694	0.968	0.925	0.744	0.995	0.991	0.978
OPC_ 09_ _1200	a	0.453	0.044	0.156	0.046	0.01	0.078	0.49	0.418	0.307	0.069	0.35
	b	0.547	0.956	0.844	0.954	0.99	0.922	0.51	0.582	0.693	0.931	0.65
OPC_ 09_ _1100	a	0.244	0.018	0.07	0.184	0.306	0.291	0.119	0.359	0.02	0.054	0.093
	b	0.756	0.982	0.93	0.816	0.694	0.709	0.881	0.641	0.98	0.946	0.907
OPC_ 09_ _900	a	0.138	0.058	0.07	0.005	0.739	0.178	0.267	0.483	0.035	0.117	0.138
	b	0.862	0.942	0.93	0.995	0.261	0.822	0.733	0.517	0.965	0.883	0.862
OPC_ 09_85 0	a	0.027	0.018	0.336	0.068	0.291	0.232	0.097	0.24	0.005	0.259	0.078
	b	0.973	0.982	0.664	0.932	0.709	0.768	0.903	0.76	0.995	0.741	0.922
OPC_ 09_ _730	a	0.29	0.031	0.12	0.068	0.032	0.466	0.49	0.483	0.035	0.203	0.255
	b	0.71	0.969	0.88	0.932	0.968	0.534	0.51	0.517	0.965	0.797	0.746
OPC_ 09_ _670	a	0.063	0.626	0.138	0.09	0.01	0.261	0.097	0.071	0.005	0.203	0.154
	b	0.937	0.374	0.862	0.91	0.99	0.739	0.903	0.929	0.995	0.797	0.846
OPC_ 09_ _570	a	0.158	0.072	0.251	0.136	0.402	0.261	0.267	0.398	0.425	0.101	0.291
	b	0.842	0.928	0.749	0.864	0.598	0.739	0.733	0.602	0.575	0.899	0.709
OPC_ 09_ _440	a	0.081	0.144	0.409	0.113	0.17	0.291	0.49	0.531	0.134	0.32	0.462
	b	0.919	0.856	0.591	0.888	0.83	0.709	0.51	0.469	0.866	0.68	0.538
OPC_ 09_ _350	a	0.556	0.563	0.555	0.685	0.276	0.427	0.749	0.722	0.005	0.101	0.543
	b	0.444	0.437	0.445	0.315	0.724	0.573	0.251	0.278	0.995	0.899	0.457
OPC_ 09_ _200	a	0.2	0.311	0.555	0.489	0.562	0.427	0.033	0.021	0.698	0.15	0.064
	b	0.8	0.689	0.445	0.511	0.438	0.573	0.967	0.979	0.302	0.85	0.936
OST_ a	a	0.186	0.018	0.186	0.005	0.594	0.078	0.19	0.289	0.253	0.009	0.217

*Chapter Three: A Population Genetics Analysis of Ostrea chilensis Using Randomly Amplified Polymorphic DNA*

02_12 00												
	b	0.814	0.982	0.814	0.995	0.406	0.922	0.81	0.711	0.747	0.991	0.783
OST_ 02_10 00	a	0.065	0.031	0.068	0.005	0.107	0.607	0.267	0.359	0.365	0.023	0.336
	b	0.935	0.969	0.932	0.995	0.893	0.393	0.733	0.641	0.635	0.977	0.664
OST_ 02_85 0	a	0.383	0.1	0.47	0.209	0.215	0.01	0.417	0.398	0.213	0.098	0.415
	b	0.617	0.9	0.53	0.791	0.785	0.99	0.583	0.602	0.787	0.902	0.585
OST_ 02_70 0	a	0.103	0.223	0.1	0.005	0.307	0.39	0.119	0.256	0.233	0.27	0.086
	b	0.897	0.777	0.9	0.995	0.693	0.61	0.881	0.744	0.767	0.73	0.914
OST_ 02_68 0	a	0.065	0.257	0.068	0.184	0.133	0.232	0.119	0.256	0.103	0.762	0.099
	b	0.935	0.743	0.932	0.816	0.867	0.768	0.881	0.744	0.897	0.238	0.901
OST_ 02_50 0	a	0.443	0.507	0.499	0.005	0.16	0.078	0.452	0.289	0.175	0.146	0.375
	b	0.557	0.493	0.501	0.995	0.84	0.922	0.548	0.711	0.825	0.854	0.625
OST_ 02_43 0	a	0.589	0.311	0.529	0.005	0.058	0.032	0.417	0.439	0.296	0.373	0.415
	b	0.411	0.689	0.471	0.995	0.942	0.968	0.583	0.561	0.704	0.627	0.585
OST_ 02_30 0	a	0.748	0.626	0.726	0.159	0.307	0.152	0.624	0.682	0.021	0.42	0.589
	b	0.252	0.374	0.274	0.841	0.693	0.848	0.376	0.318	0.979	0.58	0.411
OST_ 02_20 0	a	0.748	0.39	0.869	0.025	0.244	0.078	0.353	0.359	0.037	0.162	0.355
	b	0.252	0.61	0.131	0.975	0.756	0.922	0.647	0.641	0.963	0.838	0.645
OST_ 02_10 0	a	0.065	0.311	0.1	0.005	0.011	0.01	0.054	0.24	0.005	0.33	0.059
	b	0.935	0.689	0.9	0.995	0.989	0.99	0.946	0.76	0.995	0.67	0.941
OPE_ 05_ 2000	a	0.195	0.005	0.019	0.685	0.01	0.009	0.015	0.021	0.004	0.01	0.009
	b	0.805	0.995	0.981	0.315	0.99	0.991	0.985	0.979	0.996	0.99	0.991
OPE_ 05_ 1600	a	0.032	0.005	0.071	0.005	0.053	0.139	0.038	0.18	0.042	0.026	0.009
	b	0.968	0.995	0.929	0.995	0.947	0.861	0.962	0.82	0.958	0.974	0.991
OPE_ 05_ 1400	a	0.032	0.091	0.006	0.005	0.053	0.264	0.248	0.306	0.004	0.076	0.066
	b	0.968	0.909	0.994	0.995	0.947	0.736	0.752	0.694	0.996	0.924	0.934
OPE_ 05_ 1200	a	0.032	0.185	0.006	0.209	0.336	0.116	0.219	0.272	0.004	0.129	0.037
	b	0.968	0.815	0.994	0.791	0.664	0.884	0.781	0.728	0.996	0.871	0.963
OPE_ 05_10 00	a	0.032	0.047	0.217	0.046	0.098	0.35	0.416	0.378	0.004	0.129	0.111
	b	0.968	0.953	0.783	0.954	0.902	0.65	0.584	0.622	0.996	0.871	0.889
OPE_ 05_10 00	a	0.054	0.047	0.566	0.025	0.195	0.162	0.31	0.418	0.029	0.076	0.244

*Chapter Three: A Population Genetics Analysis of Ostrea chilensis Using Randomly Amplified Polymorphic DNA*

05_880												
	b	0.946	0.953	0.434	0.975	0.805	0.838	0.69	0.582	0.971	0.924	0.756
OPE_05_680	a	0.121	0.137	0.126	0.025	0.145	0.05	0.038	0.209	0.123	0.01	0.066
	b	0.879	0.863	0.874	0.975	0.855	0.95	0.962	0.791	0.877	0.99	0.934
OPE_05_450	a	0.304	0.494	0.338	0.757	0.739	0.847	0.599	0.531	0.37	0.692	0.636
	b	0.696	0.506	0.662	0.243	0.261	0.153	0.401	0.469	0.63	0.308	0.364
OPE_05_250	a	0.817	0.616	0.747	0.685	0.517	0.116	0.835	0.585	0.004	0.82	0.846
	b	0.183	0.384	0.253	0.315	0.483	0.884	0.165	0.415	0.996	0.18	0.154
OPE_05_160	a	0.145	0.088	0.006	0.046	0.01	0.75	0.015	0.209	0.244	0.166	0.051
	b	0.855	0.912	0.994	0.954	0.99	0.25	0.985	0.791	0.756	0.834	0.949
OPE_02_1400	a	0.015	0.005	0.019	0.048	0.133	0.009	0.054	0.194	0.004	0.021	0.024
	b	0.985	0.995	0.981	0.952	0.867	0.991	0.946	0.806	0.996	0.979	0.976
OPE_02_1200	a	0.133	0.81	0.019	0.305	0.375	0.674	0.68	0.615	0.219	0.021	0.312
	b	0.867	0.19	0.981	0.695	0.625	0.326	0.32	0.385	0.781	0.979	0.688
OPE_02_800	a	0.165	0.03	0.377	0.752	0.727	0.615	0.749	0.682	0.202	0.206	0.291
	b	0.835	0.97	0.623	0.248	0.273	0.385	0.251	0.318	0.798	0.794	0.709
OPE_02_600	a	0.438	0.057	0.418	0.567	0.594	0.844	0.749	0.585	0.03	0.427	0.548
	b	0.562	0.944	0.582	0.433	0.406	0.156	0.251	0.415	0.97	0.573	0.452
OPE_02_400	a	0.308	0.301	0.628	0.094	0.275	0.145	0.384	0.418	0.362	0.473	0.458
	b	0.692	0.699	0.372	0.906	0.725	0.855	0.616	0.582	0.638	0.527	0.542
OPE_02_360	a	0.015	0.631	0.019	0.005	0.011	0.009	0.142	0.289	0.004	0.115	0.334
	b	0.985	0.369	0.981	0.995	0.989	0.991	0.858	0.711	0.996	0.885	0.666
OPE_02_260	a	0.102	0.111	0.019	0.048	0.034	0.031	0.142	0.24	0.004	0.206	0.25
	b	0.898	0.889	0.981	0.952	0.966	0.969	0.858	0.76	0.996	0.794	0.75
OPE_02_220	a	0.102	0.667	0.17	0.142	0.011	0.031	0.19	0.289	0.004	0.523	0.04
	b	0.898	0.333	0.83	0.858	0.989	0.969	0.81	0.711	0.996	0.477	0.96
OPE_02_100	a	0.072	0.81	0.112	0.005	0.011	0.009	0.013	0.021	0.004	0.206	0.01
	b	0.928	0.19	0.888	0.995	0.989	0.991	0.987	0.979	0.996	0.794	0.99

**Appendix 3.3:** Individual locus  $F_{st}$ . Loci highlighted in bold are outlier loci

<b>RAPD Locus</b>	<b><math>F_{st}</math></b>
OPC_09_1600	0.07925
OPC_09_1400	0.14442
OPC_09	0.18134
OPC_09	0.10226
OPC_09	0.27666
OPC_09_850	0.10673
OPC_09	0.18097
OPC_09	0.20879
OPC_09	0.07398
OPC_09	0.12376
OPC_09	0.222
OPC_09_200	0.24129
OST_02_1200	0.17379
OST_02_1000	0.21545
OST_02_850	0.11549
OST_02_700	0.07859
OST_02_680	0.21565
OST_02_500	0.14512
OST_02_430	0.16972
OST_02_300	0.24818
OST_02_200	0.3083
OST_02_100	0.14644
<b>OPE_05_2000</b>	<b>0.47099</b>
OPE_05_1600	0.056
OPE_05_1400	0.12278
OPE_05_1400	0.09661
OPE_05_1000	0.14808
OPE_05_880	0.18426
OPE_05_680	0.03997
OPE_05_450	0.12346
OPE_05_250	0.31703
OPE_05_160	0.31194
OPE_02_1400	0.0743
OPE_02_1200	0.29907
OPE_02_800	0.27574
OPE_02_600	0.23014
OPE_02_400	0.09088
OPE_02_360	0.29793
OPE_02_260	0.07751
OPE_02_220	0.27125
<b>OPE_02_100</b>	<b>0.50614</b>

**Appendix 3.4** Pairwise  $\Phi_{PT}$  values for randomised individuals in four populations based on 9999 permutations is shown below the diagonal

Pop1	Pop2	Pop3	Pop4	
0.001	-			Pop1
0.001	0.001	-		Pop2
0.001	0.001	0.001	-	Pop3
0.001	0.001	0.001	0.001	Pop4

**Appendix 3.5** AMOVA table for the four random populations

Source	df	SS	MS	Est. Var.	%
Among Pops	3	19.207	6.402	0.000	0%
Within Pops	355	2818.403	7.939	7.939	100%
Total	358	2837.610		7.939	100%

**Appendix 3.6** Pairwise  $\Phi_{PT}$  values for randomised individuals in six populations based on 9999 permutations is shown below the diagonal

Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6	
0.001						Pop 1
0.008	0.001					Pop 2
0.001	0.006	0.001				Pop 3
0.001	0.001	0.001	0.001			Pop 4
0.001	0.001	0.001	0.001	0.001		Pop 5
0.001	0.009	0.001	0.001	0.001	0.001	Pop 6

**Appendix 3.7** AMOVA table for the six random populations

Source	df	SS	MS	Est. Var.	%
Among Pops	5	35.700	7.140	0.000	0%
Within Pops	353	2801.910	7.937	7.937	100%
Total	358	2837.610		7.937	100%

**Appendix 3.8** AMOVA table for the eight random populations

Source	df	SS	MS	Est. Var.	%
Among Pops	7	45.933	6.562	0.000	0%
Within Pops	351	2791.677	7.953	7.953	100%
Total	358	2837.610		7.953	100%

**Appendix 3.9** Pairwise  $\Phi_{PT}$  values for randomised individuals in eight populations based on 9999 permutations is shown below the diagonal

Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	
0.001								Pop1
0.004	0.001							Pop2
0.001	0.001	0.001						Pop3
0.001	0.001	0.001	0.001					Pop4
0.001	0.001	0.001	0.001	0.001				Pop5
0.001	0.001	0.001	0.001	0.001	0.001			Pop6
0.001	0.001	0.001	0.001	0.001	0.004	0.001		Pop7
0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	Pop8

# Chapter Four: A Seascape Genetics Analysis of New Zealand Populations of *Ostrea chilensis*

---

## 4.0 Aims

This chapter made use of the Marine Environmental Classification system (Snelder *et al.* 2005). This system was developed for the entire New Zealand Exclusive Economic Zone (EEZ). Using this data set, the aims of this chapter were to examine the genetic structuring of *Ostrea chilensis* - as described in Chapter Three - in the context of near shore coastal oceanographic and geospatial variables. The influence of environmental variability on macrogeographic genetic variation was assessed. This study will complement a number of studies that are beginning to elucidate the effect of environmental and geospatial variables on the genetic structuring of a range of New Zealand marine taxa.

## 4.1 Introduction

Two fundamental goals of a population genetic analysis should be to seek explanations for apparent genetic structure and to identify candidate genes that are under selection. In order to reach these goals the population genetics data set should be analysed within a framework that considers environmental and geophysical variables. In marine biology, this interdisciplinary approach has been termed ‘seascape genetics’ (Galindo *et al.* 2006; Selkoe *et al.* 2008). Selkoe *et al.* (2008) stated that one of the major limitations in the study of marine population genetics is the poor fit with the three traditional models of (1) panmixia, (2) population subdivision, and (3) an isolation-by-distance model. It is because of the lack of a good fit to these models that marine population genetic structure is often explained by a fourth model: chaotic patchiness - which is thought to be the result of different selection pressures in each population, or the result of stochastic influences such as weather patterns influencing larval dispersal, or variance in reproductive success (sweepstakes hypothesis) (Hedgecock, 2011).

A seascape analysis attempts to explain apparent genetic structure through correlation with environmental and geophysical variables. This addresses the problem where there seems to be incongruence between genetic structure and geographical location. Furthermore, it seeks

to elucidate the environmental variables that are having the most influence on genetic structure, this could lead to the identification of loci under selection.

Euclidean distance between two sites in some cases is not helpful in estimating realised pelagic larval dispersal - since oceanographic features such as fronts, eddies and currents can prevent dispersal. Seemingly close populations are actually very distant in terms of connectivity (Weersing and Toonen, 2009). Particle studies have shown that even along straight coast lines oceanographic processes can generate high patchiness in larval dispersal (Johnson and Black, 2006). These processes can lead to patterns that cannot be explained by an IBD model (using Euclidean distance), or generating a population structure that appears to be incongruent with geographical location (such as distantly situated populations being more genetically similar than closely situated populations) (White *et al.* 2010). A recent meta-analysis by Shafer and Wolf (2013) found the widespread presence of what they term isolation by ecology (IBE)). An IBE was recently found in the marine environment by Nanninga *et al.* (2014) that explains the correlation of mean chlorophyll-*a* concentration (used here as a proxy for an environmental gradient) and genetic structure of the anemone fish (*Amphiprion bicinctus*). Selkoe *et al.* (2010) found kelp cover to be a strong predictor of genetic structure across three species (kelp bass (*Paralabrax clathratus*), Kellet's whelk (*Kelletia kelletii*), and the California spiny lobster (*Panulirus interruptus*)).

Specific environmental variables may produce selection gradients that could prevent settlement or survival. There is a growing body of research that has found evidence of local adaptation in marine invertebrates (Sanford and Kelly, 2011). Environmental heterogeneity between populations is expected to reduce the fitness of dispersers thereby leading to divergent selection (Marshall and Morgan, 2011; Bonte, 2012), generating seemingly chaotic patchiness. There is much evidence for changes in allele frequencies that correlate with environmental gradients in the marine systems (reviewed by Schmidt *et al.* 2008) this is indicative of the influence of selection pressures under an IBE model (Riginos and Liggins, 2013).

#### **4.1.1 Recent studies in New Zealand**

New Zealand has a diverse range of environments with heterogeneous habitats interacting with complex geophysical factors. This provides a range of variables to test the correlation of genetic structure with the environment. The mainland of New Zealand spans 13 degrees of latitude ranging from sub-tropical in the far north to sub-Antarctic in far south. The oceanography consists of a complex current system characterized by a predominantly westward flowing Tasman Current, with the separation of the North and South islands by the D'Urville Current that flows through the Cook Strait (Wallis and Trewick 2009). Strong tidal flows, turbulent mixing of water masses and upwelling occurring in the Cook Strait is thought to cause biogeographic discontinuities between northern groups and southern groups (Apte and Gardner 2002, Waters and Roy 2004).

There are currently only three seascape analyses that have been conducted in New Zealand waters, these studies form a useful comparison for the influences on apparent genetic structure on *O. chilensis*. Wei *et al.* (2013b) examined the correlation between genetic structure and environmental variables in the endemic New Zealand greenshell mussel (*Perna canaliculus*). In previous studies they found pronounced genetic structure north and south of 42°S in the Cook Strait Wei *et al.* (2013a). Wei *et al.* (2013b) hypothesised that this genetic discontinuity was due to global fluctuations in sea level approximately 1 MYA and the complex geological processes that created the Cook Strait. This is supported by estimated divergence times in the northern and southern populations of *P. canaliculus* being around 1.3 MYA. In their study Wei *et al.* (2013b), did not find the discontinuity in the Cook Strait to be correlated with environmental variables, which supports the hypothesis that this is due to historical factors. However, they found that the overall genetic structuring around New Zealand is largely associated with various metrics of Sea Surface Temperature (SST). This suggests that genetic structure could be influenced by a number of factors both historical and contemporary in this species. Another complicating factor is that the genetic discontinuity was best supported with loci that were potentially under selection – this could indicate selection gradients rather than historical influences (Wei *et al.* 2013a).

In a recent Seascape study by Constable (2014), of two endemic, sympatrically occurring flounder species in New Zealand. The environmental variables that best predicted the genetic structure of the Yellowbelly Flounder (*Rhombosolea leporina*) were a combination of latitude, sediment, mean orbital velocity and sea surface temperature. The genetic structure of

the Sand Flounder (*Rhombosolea plebeia*) was best explained by a combination of latitude, longitude, sea surface temperature, mean orbital velocity, the width of the estuary mouth and the sediment type. Although the species are closely related and existing sympatrically the seascape analysis suggested different environmental variables that could indicate previously undiscovered niche partitioning.

Finally, Hannan (2013) found weak but significant population genetic structure in two mollusc species *Paphies subtriangulata* and *Paphies australis* this genetic structure was not congruent with expectations based on geographical location or life history traits of the species. There were differences in the apparent genetic population structure of these two species. Using a range of different spatial scales and measures of genetic diversity a common set of environmental variables were shown to be significant in shaping genetic structure: geographic distance, sea surface temperature, tidal currents and bed slope. These results suggest that local adaptation and restricted dispersal could be the forces that result in the apparent population genetic structure.

An understanding of the environmental patterns and processes that have created the genetic structure of marine organisms is important area of research that could greater inform disciplines such as community ecology and climate change research, as well as informing managers of marine ecosystems.

#### **4.1.2 Hypothesis**

##### **Hypothesis One: Environmental variables that influence genetic structure**

The genetic structure found in *Ostrea chilensis* is correlated with environmental variables. It is expected that the genetic split described in the Chapter Three will be explained by an environmental variable or a combination of environmental variables.

##### **Hypothesis Two: Explanatory power of outlier loci**

Separate seascape analysis with the *a priori* identified outlier loci identified in Chapter Three when removed from the data set will result in a similar pattern found in the all loci included data set (global data set). This will give some suggestion as to whether a selection gradient is affecting the outlier loci.

## **4.2 Methods**

### **4.2.1 Genetic data**

Sample collection, DNA extraction, PCR protocols, genotyping and data analysis are described in Chapter Three of this thesis. Population pairwise  $F_{st}$  and  $\Phi_{st}$  values were averaged across sampling locations to produce a single descriptive value for each of the sampling locations for the Generalised Linear Model (GLZ) following Wei *et al.* (2013b). The BEST analysis made use of the allelic frequency data (presented in Appendix 3.2).

### **4.2.2 Environmental and Geospatial data**

There are 15 environmental variables provided by the New Zealand Environmental classification (NZMEC) scheme (Appendix 4.1). Site specific data were obtained for nine of those environmental variables (as data were not available for all sampling sites): (1) Annual amplitude of sea surface temperature ( $^{\circ}\text{C}$ ), (2) Annual mean solar radiation ( $\text{W m}^{-2}$ ), (3) Depth (m), (4) Extreme Orbital velocity (m/s), (5) Freshwater fraction (FW), (6) Mean orbital velocity (m/s), (7) Spatial gradient annual mean sea surface temperature (SSTGrad)( $^{\circ}\text{C km}^{-1}$ ), (8) Summertime sea surface temperature anomaly ( $^{\circ}\text{C}$ ), (9) Winter solar radiation (Wm). Each of the variables has a spatial resolution of around 1km, variables representing long term averages of different durations and from different periods, but are typically drawn from multiple years between 1983 and 2000, or derived from multiple years between 1983 and 2000, or derived from NZ-specific models (New Zealand Ministry for the Environment 2005). Two geospatial variables were obtained in Google Earth: (11) Latitude and, (12) Longitude (both in decimal degrees).

### **4.2.3 Test of independence among variables**

A correlation analysis and a principal component analysis were carried out in order to find variables that were auto-correlated. Variables that were spatially auto-correlated at the 0.05 level were removed from the analysis. Analysis was carried out in STATISTICA v10 (Stat Soft Inc. 2011).

### **4.2.4 BEST (Biological Environmental Stepwise)**

The routine BEST from the software package PRIMER v6.0 (Clarke and Gorley 2005) was used to explore associations between population-specific genetic variation (allele frequencies

per locus) and site specific environmental/geospatial variation. The BEST analysis was performed using two data sets: (1) the global data set that included all loci; (2) the outlier loci removed (those that were shown to be more than two standard deviations from the mean  $F_{st}$  in Chapter Three).

A Bray-Curtis resemblance matrix was calculated for the environmental/geospatial data. The BIOENV subroutine was used to test for an association between the genetic and environmental matrixes using Spearman correlation coefficient ( $R_s$ ). All possible models were tested and then ranked according to which subset of environmental/geospatial variables best explained variation in the genetic data set (i.e. the model with the largest  $R_s$ ). The genetic data used in this analysis is allele frequency data (calculation of which is described in Chapter Three).

#### **4.2.5 Generalised linear model**

A generalised linear model (GLZ) was calculated using the GLZ routine in STASTICA v10 (Stat Soft Inc. 2011). A mean multilocus  $F_{st}$  or  $\Phi_{st}$  values for each population was calculated and used as a dependent variable (e.g. Wei *et al.* 2013, Foll and Gaggiotti 2006, Selkoe *et al.*, 2010) this was used to test the effects of 11 independent geospatial variables that were recorded for all populations. GLM analysis was performed using the Akaike information criterion (AIC) to rank all possible models (combinations of independent variables). The AIC identifies the best sub set of variables to include in the model: low AIC values reflect a better fit of the variables in terms of explaining variation in the model. The best-fit models (lowest AIC values) are presented as are the results of a test all effects.

### **4.3 Results**

#### **4.3.1 BEST analysis**

In an analysis of the global data set the top model of the BEST analysis identified sediment (SED) and spatial gradient annual mean sea surface temperature (SSTGrad) as the environmental variables that were most likely to explain the apparent genetic structure ( $R_s = 0.263$ ) (Table 4.1a). Overall, SED is found in nine of the top ten models and SSTGrad is found in six of the top ten models. When additional variables were added to the model the correlation ( $R_s$ ) is reduced, and the model with just one of these variables has a lower correlation. It is important to note that the model is best explained using a combination of

environmental variables, it is shown in two models that when each variable is considered alone there is a reduction in the correlation compared to when they are considered together (SED alone =  $R_s$  0.179, SSTGrad alone =  $R_s$  0.182)

With the outlier loci were removed, SED and SSTGrad were present in the models with the highest correlation ( $R_s = 0.181$ ) (Table 4.1b), although the overall correlation was reduced compared the global data set. This reduction could indicate that the two loci are important in describing the relationship between environmental and genetic variation. Table 4.1c shows the effects of removing two random loci – this was done in order to compare the effect of removing the outlier loci from the analysis. A similar pattern with a lower level of correlation ( $R_s = 0.167$ ) was found.

**Table 4.1a:** 10 best fitting models using the BEST analysis using the global data set

<i>Number of Variables</i>	<i>Correlation (<math>R_s</math>)</i>	<i>Variables</i>
2	0.263	Sed + SSTGrad
3	0.214	Sed + SSTGrad + Rad_Wint
2	0.201	Sed + Rad_wint
1	0.182	SSTGrad
1	0.179	Sed
3	0.168	Lat+ Sed + Rad_wint
4	0.168	Lat + Sed + SSTGrad + Rad_wint
3	0.164	Sed + SSTAnnAmp + SSTWint
4	0.164	Sed + SSTAnnAmp + SSTGrad + SSTWint
3	0.164	Lat + Sed + SSTGrad

**Table 4.1b:** 10 best fitting models using the BEST analysis with outlier loci removed

<i>Number of variables</i>	<i>Correlation (<math>R_s</math>)</i>	<i>Selections</i>
2	0.181	Sed + SSTGrad
2	0.175	Sed
2	0.175	Sed + SSTGrad + Latitude
4	0.169	Sed + MOrbVel +Rad_wint + Tidal
5	0.169	Sed + MOrbVel + SSTGrad + Rad_wint + Tidal
3	0.168	Sed + MOrbVel + Rad_wint
4	0.168	Sed + MOrbVel +SSTGrad +Rad_wint
5	0.168	Sed + MOrbVel +SSTAnnAmp +STTWint +Rad_wint
5	0.168	Sed+ MOrbVel + SSTAnnAmp +Orb_v_95+Tidal
6	0.168	Sed + MOrbVel + SSTAnnAMp+ STTWint +Rad_wint

**Table 4.1c:** 10 best fitting models using the BEST analysis two random loci removed

<i>Number of Variables</i>	<i>Correlation (<math>R_s</math>)</i>	<i>Variables</i>
2	0.167	Sed + SSTGrad
3	0.165	Sed + SSTGrad + Rad_Wint
2	0.165	Sed + Rad_wint
1	0.161	SSTGrad
1	0.159	Sed
4	0.159	SSTGrad + Lat+ Sed + Rad_wint
4	0.158	Lat + Sed + SSTGrad + Rad_wint
3	0.158	Sed + SSTGrad + SSTWint
4	0.158	Sed + SSTAnnAmp + SSTGrad + SSTWint
3	0.158	Lat + Sed + SSTGrad

#### **4.3.2 GLZ analysis**

In the comparison of environmental variables with the mean genetic distance ( $F_{st}$ ), using all loci in the data set, the best fit model included: Lat, Long, MeanOrbVel, STTAnAmp, SSTWint, Orb\_v\_95, Depth and Tidal, with a p value of  $<0.0001$  (Table 4.2a). With the outlier loci removed from the analysis Lat was no longer in the top model and SSTGrad was included, with a p value of  $<0.0001$  (Table 4.2b). Overall, comparing the results of the two analyses the outcome is similar apart from the inclusion of Rad\_wint in four of the top models when the outlier loci were excluded (Table 4.2b).

The GLZ was repeated with another measure of genetic distance: mean  $\Phi_{st}$  for comparison. The best fit model was the same as the analysis using  $F_{st}$  except for the exclusion of MeanOrbVel and the inclusion of SSTGrad (Table 4.3a). With the outlier loci removed the best fit model was the same as the global data set (Table 4.3b). Again, when the outlier loci were removed Rad\_wint featured in four of the best fit models (Table 4.3b).

**Table 4.2a:** Results of the GLZ analysis using the global data set using  $F_{st}$

<i>Rank</i>	<i>AIC</i>	<i>p-value</i>	<i>Lat</i>	<i>Long</i>	<i>Sed</i>	<i>MeanOrbVel</i>	<i>SSTAnAmp</i>	<i>SSTGrad</i>	<i>SSTWint</i>	<i>Depth</i>	<i>Tidal</i>
1	-1.06E+14	0.000000	x		x	x	x	x	x	x	x
2	-2.20E+13	0.000000	x	x	x		x	x	x	x	x
3	-3.44E+11	0.000000	x	x	x	x		x	x	x	x
4	-1.89E+11	0.000000	x	x	x	x	x		x	x	x
5	-7E+09	0.000000	x	x	x	x	x	x	x		x
6	-3.2E+09	0.000000	x	x		x	x	x	x	x	x
7	-3.2E+09	0.000000	x	x	x	x	x	x		x	x
8	-4E+08	0.000000	x	x	x	x	x	x	x	x	
9	-4E+08	0.000000	x	x	x	x	x	x	x	x	x
10	-524270	0.000000		x	x	x	x	x	x	x	x
		<b>p-value</b>	<b>0.314207</b>	<b>0.838306</b>	<b>0.792872</b>	<b>0.792872</b>	0.015834	<b>0.517168</b>	0.00193	0.00019	0.000000

**Table 4.2b:** Results of the GLZ analysis using the outlier removed data set using  $F_{st}$

<i>Rank</i>	<i>AIC</i>	<i>p-value</i>	<i>Lat</i>	<i>Long</i>	<i>Sed</i>	<i>MeanOrbVel</i>	<i>SSTAnAmp</i>	<i>SSTGrad</i>	<i>SSTWint</i>	<i>Depth</i>	<i>Tidal</i>
1	-7.04E+13	0.000000	x	x	x	x	x	x	x	x	
2	-7.04E+13	0.000000	x	x	x	x	x	x	x	x	x
3	-4.40E+12	0.000000	x	x	x	x	x		x	x	x
4	-1.12E+11	0.000000	x	x	x	x	x	x		x	x
5	-1073741806	0.000000	x		x	x	x	x	x	x	x
6	-402653166	0.000000	x	x	x		x	x	x	x	x
7	-125829102	0.000000		x	x	x	x	x	x	x	x
8	-16777198	0.000000	x	x		x	x	x	x	x	x
9	-104.2215477	0.000000	x	x	x		x	x	x		x
10	-90.82262246	0.000000	x	x	x	x	x		x	x	
		<i>p-value</i>	<b>0.20746</b>	<b>0.97118</b>	<b>0.673939</b>	0.044699	<b>0.358235</b>	0.009607	0.000000	0.000000	0.000000

**Table 4.3a:** Results of the GLZ analysis using  $\Phi_{st}$  and the global data set

<i>Rank</i>	<i>AIC</i>	<i>p-value</i>	<i>Lat</i>	<i>Long</i>	<i>Sed</i>	<i>MeanOrbVel</i>	<i>SSTAnAmp</i>	<i>SSTGrad</i>	<i>SSTWint</i>	<i>Depth</i>	<i>Tidal</i>
1	-1.41E+14	0.000000	x	x	x	x	x	x	x	x	
2	-1.41E+14	0.000000		x	x	x	x	x	x	x	x
3	-1.41E+14	0.000000	x	x	x	x	x	x	x	x	x
4	-3.52E+13	0.000000	x	x	x	x	x		x	x	x
5	-1.76E+13	0.000000	x	x		x	x	x	x	x	x
6	-9.90E+12	0.000000	x	x	x	x	x	x	x		x
7	-1.03E+11	0.000000	x	x	x	x	x	x		x	x
8	-8.59E+09	0.000000	x		x	x	x	x	x	x	x
9	-1.34E+09	0.000000	x	x	x	x		x	x	x	x
10	-1.68E+08	0.000000	x	x	x		x	x	x	x	x
		<i>p-value</i>	<b>0.111924</b>	<b>0.677079</b>	<b>0.133868</b>	<b>0.00054</b>	<b>0.203846</b>	<b>0.042563</b>	<b>0.000972</b>	0.000000	0.000000

**Table 4.3b:** Results of the GLZ analysis using  $\Phi_{st}$  and the global data set using phist

<i>Rank</i>	<i>AIC</i>	<i>p-value</i>	<i>Lat</i>	<i>Long</i>	<i>Sed</i>	<i>MeanOrbVel</i>	<i>SSTAnAmp</i>	<i>SSTGrad</i>	<i>SSTWint</i>	<i>Depth</i>	<i>Tidal</i>
1	-4.93E+14	0.000000	x		x	x	x	x	x	x	x
2	-2.11E+14	0.000000	x	x	x		x	x	x	x	x
3	-6.16E+13	0.000000	x	x	x	x		x	x	x	x
4	-1.76E+13	0.000000	x	x	x	x	x		x	x	x
5	-4.40E+12	0.000000	x	x	x	x	x	x	x		x
6	-3.30E+12	0.000000	x	x		x	x	x	x	x	x
7	-8.93E+11	0.000000	x	x	x	x	x	x		x	x
8	-1.07E+09	0.000000	x	x	x	x	x	x	x	x	
9	-1.07E+09	0.000000	x	x	x	x	x	x	x	x	x
10	-2.10E+07	0.000000		x	x	x	x	x	x	x	x
		<b>p-value</b>	<b>0.32431</b>	<b>0.6991</b>	<b>0.78284</b>	0.002186	<b>0.606051</b>	0.014028	0.000623	0.000000	0.000000

## 4.4 Discussion

In Chapter Three population genetic structuring was identified in *Ostrea chilensis*. The genetic structure was not explained by an IBD model and the structure seemed to be incongruent with geographical location. This is often attributed to chaotic genetic patchiness, but it is likely this structure can be better explained by correlation with environmental variables. In the present study, I used geospatial modelling to seek the environmental or physical variables that were correlated with this apparent genetic structure.

The results of the BEST analysis using the global data set suggest that a combination of two environmental variables: sediment (SED) and spatial gradient annual mean sea surface temperature (SSTGrad), SSTGrad are correlated with the apparent genetic structure as these were present in the top model ( $R_s = 0.263$ ) (Table 4.1a). When the outlier loci were removed the two top models had the same environmental variables present, albeit with a lower correlation ( $R_s = 0.181$ ) (Table 4.1b). Removing random variables resulted in a similar pattern with a lower correlation (Table 4.1c). I would expect that if the loci are under selection then a completely different set of variables would explain the genetic structure. However, since their removal causes a reduction in correlation it suggests these loci could be influential in generating overall genetic structure and they cannot be ruled out as not being under selection or at least linked to genes under selection. But it is more likely the correlation of genetic structure and environmental variables is due to a genome wide effect rather than specific loci as a similar pattern is found when removing random variables.

The analysis presented in this chapter suggests that apparent genetic structure is correlated with combinations of environmental variables, and possibly due to interactions between these variables. Therefore, no one variable is responsible for the genetic structure. The environmental variables used in the Marine Environmental Classification system (Snelder *et al.* 2005) included four variables to capture sea surface temperature (SST). These variables capture specific oceanographic processes, both physical and chemical, that affect biological patterns (Snelder *et al.* 2001). These included winter time SST (WintSST), annual amplitude of SST (SSTanamp), summer time sea surface temperature anomaly (SSTanom) and spatial gradient of annual mean SST (GradSST). SSTGrad appeared in the top models of the BEST analysis (Table 4.1a/b/c) this variable is used to recognise fronts in oceanic water masses that are expected to correlate with variation in primary productivity. The two populations in the north of the South island have the highest values for this variable (Cloudy bay: 0.0358 °C

km<sup>-1</sup> Marlborough Sounds: 0.09942 °C km<sup>-1</sup>) with the populations below 42°S in the Tasman Strait (and the Chatham Islands) having lower values.

The sediment type is a factor that determines the composition of benthic communities.

Variation in sediment type was derived from the New Zealand Region Sediments chart (Mitchel *et al.* 1989). There are 23 categories that are based on the dominant and subdominant sediment type. The sediment types were converted into effective particle size then averaged and ranked to give a continuous variable rank of sediment size, a variable that is suitable for correlation analysis. The sediment variable was not included in the final EEZ classification scheme as it was found to dominate classification schemes and was out of proportion to sediment's actual value as a predictor as shown by validation analyses, furthermore it had a low resolution. However, the sediment variable was included in this analysis as it showed some relationships with previous biological data sets (see Image *et al.* 2003). A closer look at the sediment types around New Zealand shows that the sediments are very similar around the Foveaux Strait and Stewart Island that are dominated by deep sea clay. The Akaroa population was quite different, which was dominated by Calcium sand, with mud being a sub-dominant type. Cloudy Bay and Marlborough sounds populations had similar sediments with clay being the dominate type, and with gravel sand and mud being sub-dominant types. Finally, the Chatham Islands samples had a dominant sediment type of clay and a sub- dominant type of sand. Given that there are very similar sediment types in the areas around the Foveaux Strait and a contrasting type of sediments in the Cloudy Bay and Marlborough Sounds populations it is not a surprise that sediment is a variable found in the top BEST models (Tables 4.1a/b/c). These populations have shown strong sub-division in the AWClust analysis of the previous chapter (Table 3.10 and 3.11). Unfortunately, I cannot compare my results to that of Wei *et al.* (2013b) as the sediment variable was not included in their analysis. However, many bivalves display dispersal patterns that are associated with sediment. For example, habitat selection experiments have shown that larvae of some species can choose to settle in favourable habitats, for example the work of Snelgrove *et al.* (1998) demonstrated that the bivalve *Spisula solidissima* was able to preferentially settle in sandy sediments. In the field Huxam and Richards, (2002) found that the bivalves *Cerastoderma edule* and *Macoma balthica* showed a strong association with sediment type. In this study sediment type was shown to be a strong factor in explaining the genetic structure of *Ostrea chilensis*: it remains for further investigation to see whether *O. chilensis* can select sediment type or whether there is any local adaptation to sediment type. This would also be important

knowledge for management of the species, because oyster dredging is known to affect the distribution of sediment (Cranfield *et al.* 2003).

SSTgrad and other metrics of temperature were correlated with the apparent genetic structure. However, it is impossible to explain the mechanistic causes since the temperature variables interact to create complex biological and physical interactions, although some generalisations can be made. In *O. chilensis* spawning times, larval settlement and larval survival have been shown to be strongly influenced by temperature (Chaparro, 1990). In another study of *O. chilensis* Jeffs *et al.* (2002), found that warmer temperatures could be involved in synchronizing and initiating oogenesis. These studies support the finding that SSTgrad could be correlated with genetic variation. If this is the case then differences in temperature between sampling sites could result in spawning occurring at different times, potentially reducing gene flow between populations, increasing local recruitment and genetic difference between populations via genetic drift.

The results of the GLZ indicated a more complicated correlation of genetic structure and environmental variables. Most of the environmental variables were present in the top ten models. Latitude was found in most of the top models; this may be a proxy variable for any number of environmental or oceanographic variables that change with latitude, including temperature, which is also present in the top models. This contrasting picture indicates that there was no single environmental variable contributing to the genetic structure. This could indicate that the apparent genetic structure is largely the result of neutral evolutionary forces such as drift and gene flow. However, the results of this analysis could be confounded by a loss of resolution in genetic information as the genetic diversity is described using a mean. A limitation with the use of RAPD loci for a seascape analysis is that there are only two alleles. The use of microsatellite loci for this type of analysis since they are generally multi-allelic, which gives a higher resolution, using markers such as these with a higher information content leads towards the identification of loci under selection and the mechanisms of natural selection on these loci. This lower level of resolution of RAPDs means that the results of the study should be treated as exploratory. In the study of Wei *et al.* (2013a), they identified a microsatellite locus that contributed substantially to the correlation between genetic structure and environmental variation. In the search for candidate genes for selection multi-allelic markers or markers with high genome coverage such as SNPs (single nucleotide polymorphisms) need to be developed for *O. chilensis*, this would also lead to a greater resolution in which to test the effects of environmental variables on genetic structure. Another limitation of this study is that environmental and geospatial variables that are being

tested here have no *a priori* expectations of associations with the genetic variable (e.g. Wei *et al* 2013b). It is possible that the actual causative variables of genetic diversity have not been included in the analysis, only those variables that are easily accessible have been included: those that are available from the NZMEC and those that are easy to obtain such as latitude and longitude. It is recognised here that there could be environmental variables that have more influence on the apparent genetic structure that have not been included.

Since this is a relatively new technique, there is a lack of New Zealand seascape genetics studies. In future more work needs to be done to investigate the effect of environmental variation on the genetic structure across multiple taxa. In this study environmental variables that contribute to genetic variation have been highlighted, they may not be the causal factors and, if they are, at present the mechanistic basis of the relationship cannot currently be explained. It should also be noted that impact of environmental variables on the apparent genetic structure cannot easily be separated from the effects of historical factors - such as sea level changes reducing gene flow between sub-populations. These results, however, provide the initial step to elucidating the relationship of functional genes and their relationship with environmental variables.

#### **4.5 Summary**

1. This exploratory analysis revealed that genetic structuring of *Ostrea chilensis* is correlated with sediment type and spatial gradient annual mean sea surface temperature. Sediment type is very different around sampling locations. Although this is an environmental variable with low resolution and has some limitations, it could explain the apparent genetic structure of *O. chilensis*.
2. An environmental variable linked with sea surface temperature has been shown to correlate with genetic structure. SST has been implicated with explaining the genetic structure in the study of Wei *et al.* (2013a). This study provides further evidence that variables influencing SST are important in explaining genetic structure.
3. More genetic markers with higher information content and greater genome coverage need to be developed in order to find loci that are under selection and better describe the influence of environmental variables on genetic structure.

## 4.6 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(9), 1617-1628.
- Bonte B, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M. (2012) Costs of dispersal *Biological Reviews* 87 (2), 290-312
- Chaparro OR. (1990). Effect of temperature and feeding on conditioning of *Ostrea chilensis* Philippi, 1845. *Aquaculture Research*, 21(4), 399-405
- Clarke KR, Gorley RN. (2005). PRIMER: Getting started with v6. *PRIMER-E Ltd: Plymouth, UK*.
- Constable HB. (2014). Population structure, temporal stability and seascape genetics of two endemic New Zealand Pleuronectiformes, *Rhombosolea plebeia* (sand flounder) and *R. leporina* (yellowbelly flounder). Ph. D Thesis. Victoria University of Wellington. NZ
- Cranfield HJ, Manighetti B, Michael KP, Hill A. (2003). Effects of oyster dredging on the distribution of bryozoan biogenic reefs and associated sediments in Foveaux Strait, southern New Zealand. *Continental Shelf Research*, 23(14), 1337-1357.
- Foll M, Gaggiotti O (2006). Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, 174(2), 875-891.
- Galindo HM, Olson DB, Palumbi SR (2006). Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology*, 16(16), 1622-1626.
- Hannan, D. A. (2014). Population genetics and connectivity in *Paphies subtriangulata* and *Paphies australis* (Bivalvia: Mesodesmatidae). Ph. D Thesis. Victoria University of Wellington. NZ
- Hedgecock D, Pudovkin AI. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, 87(4), 971-1002.

- Huxham M, Richards M. (2003). Can postlarval bivalves select sediment type during settlement? A field test with *Macoma balthica* (L.) and *Cerastoderma edule* (L.). *Journal of experimental marine biology and ecology*, 288(2), 279-293.
- Image K, Richardson K, Rowden A, Francis M, Snelder T. (2003). Validation of environmental variables used for New Zealand Marine Environment Classification (EEZ scale). *NIWA Report CHC2003-029*. 46 p.
- Jeffs AG, Dunphy BJ, Wells RMG. (2002). Experimental effects of water temperature on the gametogenic development of broodstock in the oyster, *Ostrea chilensis*. *Journal of Shellfish Research*, 21(2), 743-747
- Johnson M, Black R. (2006). Islands increase genetic subdivision and disrupt patterns of connectivity of intertidal snails in a complex archipelago. *Evolution*, 60 (12), 2498-2506.
- Marshall D J, Morgan SG. (2011). Ecological and evolutionary consequences of linked life-history stages in the sea. *Current Biology*, 21(18), R718-R725.
- Nanninga GB, Saenz-Agudelo P, Manica A, Berumen ML. (2014). Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea. *Molecular ecology*.
- Riginos C, Liggins L. (2013). Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass*, 7(3), 197-216.
- Sanford E, Kelly MW. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3, 509-535.
- Schmidt PS, Serrao EA, Pearson GA (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, 89, 91–107.
- Selkoe KA, Henzler CM, Gaines SD. (2008). Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, 9(4), 363–377
- Selkoe KA, Watson JR, White C, Ben Horin T, Iacchei M, Mitarai S.(2010). Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology* 19: 3708–3726
- Shafer A, Wolf JB. (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology letters*, 16(7), 940-950.

Snelder T, Leathwick J, Day K, Weatherhead M, Fenwick G, Francis M, Gorman R, Grieve J, Hadfield M, Hewitt J, Hume T, Richardson K, Rowden A, Uddstrom M, Wild M, Zeldis J. (2005). *The New Zealand Marine Environment Classification* (p. 80). Wellington, New Zealand.

Snelgrove PV, Grassle JP, Butman CA. (1998). Sediment choice by settling larvae of the bivalve, *Spisula solidissima* in flow and still water. *Journal of experimental marine biology and ecology*, 231(2), 171-190.

Wallis GP, Trewick SA. (2009). New Zealand phylogeography: evolution on a small continent. *Molecular Ecology*, 18(17), 3548-3580.

Waters JM, Roy MS. (2004). Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*, 13(9), 2797-2806.

Wei K, Wood AR, Gardner JPA (2013a). Population genetic variation in the New Zealand greenshell mussel: locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. *Marine Biology*, 160(4), 931-949.

Wei K, Wood AR, Gardner JPA (2013b). Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. *Marine Ecology Progress Series*, 477, 107-121.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*, 277(1688), 1685-1694.

## 4.7 Appendix

**Appendix 4.1:** Environmental variables from the Marine Environment Classification (MEC). Variables highlighted in bold were not independent

Name	Abbreviation	Units	Description
Tidal current	Tidal	m/s	Depth averaged maximum tidal current
Annual amplitude of sea surface temperature	SSTAnnAmp	°C	Smoothed annual amplitude of SST
<b>Annual mean solar radiation</b>	<b>RadMean</b>	<b>W m<sup>-2</sup></b>	<b>Mean extra atmospheric solar radiation modified by mean annual cloud cover</b>
Depth	Depth	m	Bathymetry grid (1 km resolution)
Extreme Orbital velocity	Orb_v_95	m/s	Orbital velocity at the bed for the 95th percentile significant wave height calculated from a 20-year wave hindcast
Freshwater fraction	FW	proportion	Proportion of fresh water based on river inputs
Mean orbital velocity	MOrbVel	m/s	Orbital velocity at the bed for the mean significant wave height calculated from a 20 year wave hindcast
Seabed curvature	Bed_curv	0.01m-1	Curvature of the surface surrounding each grid cell

Seabed platform curvature	Bed_plan	0.01m-1	Curvature of the surface perpendicular to the slope direction
Seabed rate of change of slope (profile)	Bed_prof	0.01m-1	The rate of change of slope for each cell
Sediment	Sed	n/a	Sediment type as a categorical variable
Spatial gradient annual mean sea surface temperature	SSTGrad	°C km <sup>-1</sup>	Smoothed magnitude of the spatial gradient of annual mean SST
<b>Summertime sea surface temperature anomaly</b>	<b>SSTAnom</b>	<b>°C</b>	<b>Spatial anomalies with scales between 20 and 450 km in late February when SST is typically highest</b>
<b>Winter solar radiation</b>	<b>Rad_wint</b>	<b>Wm</b>	<b>Extra atmospheric solar radiation in June, modified by mean annual cloud cover</b>
Wintertime sea surface temperature	STTWint	°C	Mean of daily data from early September when SST is typically lowest

# Chapter Five: Partial Assembly of the *Ostrea chilensis* Mitogenome and its Subsequent use in Phylogenetic Reconstruction

---

## 5.0 Aims

The aims of this chapter were, first, to identify mitochondrial DNA sequences within the 454 DNA sequence reads, and assemble them into a genome sequence; and secondly, to use the mitochondrial genome data to examine the phylogenetic position of *Ostrea chilensis*

## 5.1 Introduction

Small PGM/454 DNA sequencing runs are often used in the field of population genetics to identify microsatellite DNA loci. However, these data sets also contain other DNA sequences that can be used to more thoroughly to understand genome structure and arrangements. For example, Rasmussen and Noor (2009) used 454 data to develop useful biological information and genomic resources for a non-model organism: the Scuttle Fly (*Megaselia scalaris*).

Although the sequence data only represented 0.1 % of the genome they were able to: assemble the full mitogenome, identify repetitive elements (such as transposons and microsatellites), and identify gene homologs. Similar work was carried out by Leese *et al.* (2012), they found a positive correlation between the number of mitochondrial fragments and the total number of fragments within 454 data sets from a range of taxa.

### 5.1.1 Mitochondrial DNA

The mitogenome is the only extra-nuclear genome in animals (Boore, 1999). Typically the mitogenome is compact, within Ostreidae the size ranges from 16.2 kbp to 22.5kbp (Table 5.1). The mitogenome usually codes for 13 proteins, 22 transfer RNAs and two ribosomal RNAs (Danic-Tchaleu *et al.* 2011). Within Bivalvia the gene order seems to be highly variable, and this heterogeneity can be phylogenetically informative (Yuan *et al.* 2012). However, in oysters the gene order seems to be conserved (Danic-Tchaleu *et al.* 2011, Ren *et*

al. 2010). There are often few intergenic regions except for a single large non-coding region that is thought to contain elements that control the initiation of replication and transcription (Shadel and Clayton, 1997). Size variation in the mitogenome is usually due to the different lengths of the non-coding region (Danic-Tchaleu *et al.* 2011).

During the last few years, there has been a significant increase in the number of complete mitogenomes available for a range of species. Among the bivalves, there are currently 293 mitogenomes available on GenBank for a range of individuals and species. Within Ostreidae there are eight *Crassostrea* species; three *Ostrea* species and one species from the genus *Saccostrea* which is *Saccostrea mordax*. The number of mitogenomes available on GenBank is expected to rapidly increase as researchers are beginning to use genomic data sets to assemble full mitogenomes from their sequencing data (e.g. Zeng *et al.* 2014).

### 5.1.2 Oyster mitogenomes and phylogenetics

Recently, the phylogenetic relationships of *Ostrea* and *Crassostrea* species have been studied using whole mitochondrial genomes sequences. Danic-Tchaleu *et al.* (2011) sequenced the complete mitogenome of the European flat oyster (*Ostrea edulis*) they then used all of the protein coding genes for all of the published oyster mitogenomes to reconstruct a phylogeny.

The mitogenome of *O. edulis* was amplified in four overlapping fragments using species-specific primers. These primers bound to particular gene regions and amplified a product of around 4193-5383bp. The purified products were then directly used as templates in cycle sequencing reactions with dye-labelled terminators. Specific primers were designed and used for primer-walking-sequencing, which was performed for both strands of each sample using an ABI genetic analyser. The sequences were then assembled into a contig. Subsequently, protein coding and ribosomal RNA genes were identified using BLAST searches of GenBank, and also by alignment with previously published mollusc mitogenomes. The amino acid sequences of protein coding genes were inferred with ORF Finder (Rombel *et al.* 2002). Transfer RNAs were identified using DOGMA (Wyman *et al.* 2004) and tRNAscan-SE (Lowe *et al.* 1997).

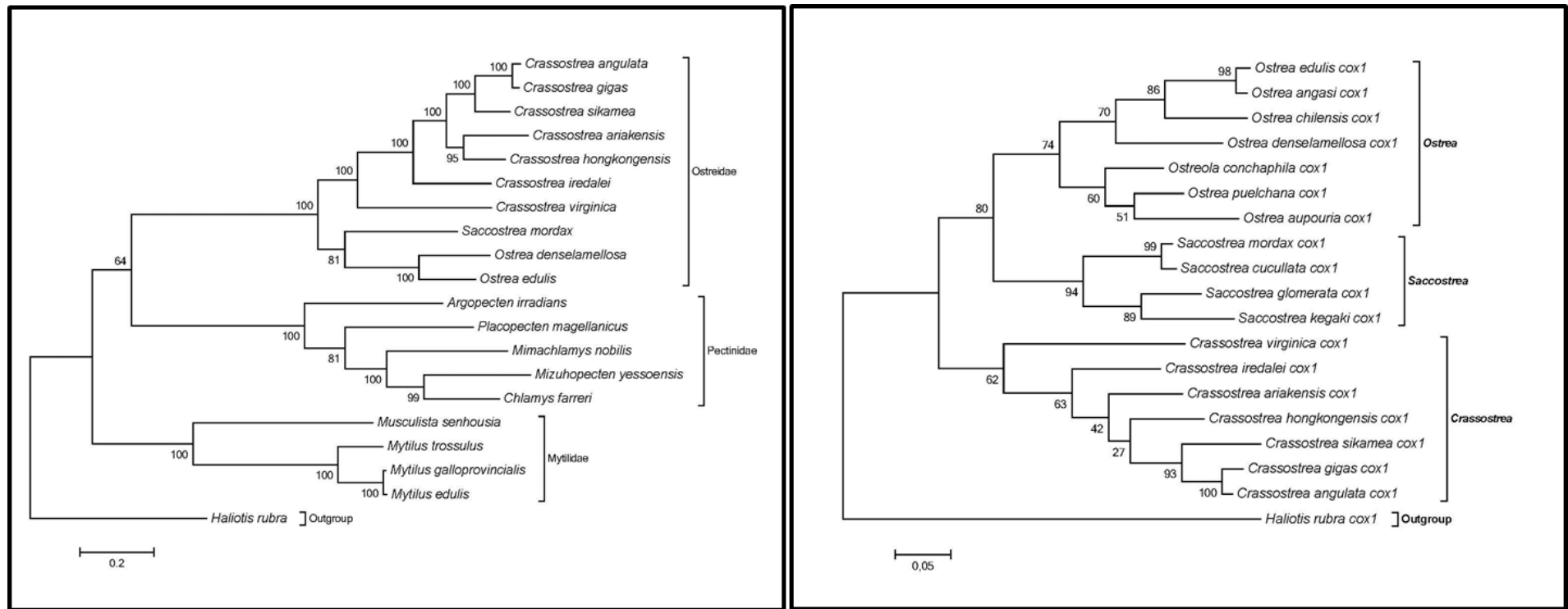
Danic-Tchaleu *et al.* (2011) then conducted two phylogenetic analyses: one using twelve concatenated protein coding genes from 19 bivalve mitogenomes; the second analysis used COI sequences from all published oyster species from the genera: *Ostrea*, *Ostreola*,

*Saccostrea* and *Crassostrea*. The phylogenetic analysis was conducted by aligning the genes using the program MUSCLE (Edgar, 2004) the tree was produced using Maximum Likelihood (ML) implemented in MEGA5 (Tamura *et al.* 2013). The phylogenetic trees from the study of Danic-Tchaleu *et al.* (2011) are reproduced in Figure 5.1.

The phylogenetic position of *O. chilensis* was examined in earlier work by Jozefowicz and Ó Foighil (1998) using a partial 16S rRNA sequence. In later work Ó Foighil (Ó Foighil, 2000) tested the phylogenetic relationships using 28S rRNA. The tree topology reported in these studies was subsequently confirmed by Danic-Tchaleu *et al.* (2011) with *Ostrea chilensis* being ancestral to the sister taxa *Ostrea angasi* and *Ostrea edulis*. Ó Foighil (Ó Foighil *et al.* 1999) also conducted a phylogeographic analysis of *O. chilensis* using samples from the North Island, South Island and Chile to conclude (along with fossil evidence) that the most probable explanation for the presence of a Chilean population of *O. chilensis* is by rafting from an ancestral North Island population.

There have not been any studies that have used multiple genes to determine the phylogenetic position of *O. chilensis*, even though the use of single gene regions has been criticised by a number of studies (for example, Brito and Edwards, 2009). Single-gene trees have been shown to have limited value because tree topology is often only weakly supported. Better statistical support tends to result from larger data sets that are comprised of several different gene regions.

In addition to their applications to phylogenetic reconstruction, there are many other benefits that can be gained from mitochondrial sequence data. Because the mitogenome is inherited through the maternal line (although this is not always the case in some bivalves (see Doucet-Beaupré, *et al.* 2010)) it enables sex-specific patterns of migration to be traced. Genetic drift occurs quicker in mtDNA due to a smaller effective population size than nuclear markers, meaning mtDNA markers can be more sensitive to bottlenecks (Moritz *et al.* 1987). Furthermore, mtDNA analyses tend to show monophyletic relationships whereas the nuclear genes often show poly- or paraphyletic relationships, which most likely arise because recombination can occur within genes in the nuclear genome (Zink and Barraclough, 2008).



**Figure 5.1:** The tree on the left was produced using 19 protein coding genes extracted from the full mitogenomes. The tree on the right was produced using COI fragments from all published *Ostrea*, *Ostreola*, *Saccostrea* and *Crassostrea*. Support values are the result of 100 Bootstraps

The goal of the research presented in this chapter is to provide the initial platform for future studies of *O. chilensis* that use mitochondrial DNA markers in areas such as phylogenetics, phylogeography and aquaculture genetics.

### **5.1.3 Hypotheses**

#### **Hypothesis One: Mitochondrial DNA**

Mitochondrial DNA is over-represented in 454 sequencing runs because there are multiple copies of the mitogenome in a cell compared to the nuclear genome. Therefore, it will be possible to recover many mitochondrial fragments.

#### **Hypothesis Two: The phylogenetic position of *Ostrea chilensis***

The placement of *O. chilensis* using protein coding mitochondrial genes developed in this study and analysed using Bayesian methodology will support the findings of previous work using single gene regions (Jozefowicz and Ó Foighil, 1998; Ó Foighil, 2000)

#### **Hypothesis Three: Origin of the Chilean population**

The study of Ó Foighil *et al.* (1999) suggests that the Chilean population of *O. chilensis* was established by rafting from the North Island of New Zealand. The inclusion of data developed in this study and data available from GenBank will support this suggestion.

## 5.2 Methods

### 5.2.1 Identification of mtDNA sequences

Mitochondrial sequences were identified in the 454 data set by using the BLAST search procedure of the sequences in the 454 DNA database against the published genomes of closely related taxa. A custom BLAST database was made using the program GENEIOUS, a batch of mitogenomes was then blasted against this database, maximum e-value was set to 1e-20, the e-value is a measure of the probability an alignment occurring by chance; increasing the e-value would increase the number of similar sequences (“hits”) that were returned, but are more likely to be biologically un-meaningful (not mitochondrial sequences). The mitogenomes used and the number of 454 sequences that matched with each mitogenome are shown in Table 5.1.

Once the 454 sequences of mitochondrial origin were identified the sequences were assembled - using the ‘*de novo*’ assembly function in GENEIOUS. The contigs were then blasted against the 454 data set in an attempt to find additional sequence matches. The contigs from each assembly were then mapped to the published mitogenome of *Ostrea edulis* (Genbank Accession Number JF274008), which was used as a reference genome, finally the 454 sequences were aligned - using the ‘*align to reference*’ function in GENEIOUS. The *O. edulis* genome was used as a reference genome as it had the highest number of hits from the 454 database (Table 5.1). The annotations of genes on the mitochondrial genome of *O. edulis* were used as a guide for the annotation of the draft genome of *O. chilensis*.

### 5.2.2 Phylogenetic reconstruction using oyster mitogenomes

All oyster mitogenomes published on GenBank were used for phylogenetic construction (Table 5.1). The gene sequences for Cytochrome oxidase 1 (COI), Cytochrome oxidase 2 (COII), ATP synthetase subunit 6 (ATP6), Cytochrome Oxidase 3 (COIII), Nicotinamide adenine dinucleotide 1 (ND 1), Nicotinamide adenine dinucleotide 2 (ND 2), Nicotinamide adenine dinucleotide 3 (ND3), Nicotinamide adenine dinucleotide 4 (ND4), Nicotinamide adenine dinucleotide 4L (ND 4L), Nicotinamide adenine dinucleotide 6 (ND6), Cytochrome B (CYTB) and were extracted from each of the published genomes using GENEIOUS. The partial gene sequences for ATP6, COX1, COX2, COX3, CYTB, NAD3, NAD4 and NAD4L were used for *O. chilensis* because not all genes on its mitogenome were recovered.

Each set of homologous gene regions were aligned separately using the alignment program ClustalW implemented in MEGA 6 (Tamura *et al.* 2013). The protein coding genes were analysed for the best fitting evolutionary model using jModeltest v 2.1.1 (Darriba *et al.* 2012) using four Gamma categories. The tree on which the models were optimised was produced using maximum likelihood. The corrected Akaike information criterion (AICc) was used to select the best models. The alignment was partitioned to take into account the different evolutionary models for each gene sequence. Partitions were chosen *a priori* based on gene identity, although more computationally intensive it has been shown to improve the estimated posterior probabilities for phylogenetic reconstruction (Brandley *et al.* 2005).

MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) was used to for Bayesian inference. Two independent runs with two different starting trees consisting of four Markov chains per run (three heated chains and one cold chain) of 1,100,000 steps were implemented. The first 100,000 steps of each are discarded as the burn-in period, the chain is then sampled every 200<sup>th</sup> step. Convergence of the MCMC chain was tested in three ways (1) visual inspection of the trace file to make sure there is regular swapping between low and high likelihoods (2) the standard deviation of parameter estimates is >1 (3) by testing the null hypothesis that mean of the first 10 percent of the Markov chain (after burn in is removed) is equal to the mean of the final 50 percent of the Markov chain, the test statistic for this is the Geweke diagnostic (Geweke, 1992). If there is no difference between means, the null hypothesis that there is no significant difference between the means is accepted. These tests were implemented using the R package CODA (Plummer *et al.* 2005).

Bayesian posterior probabilities of >95 were considered as significant support for a clade (Wilcox *et al.* 2002). The tree produced was a majority rule consensus tree, containing the posterior probability values for each clade these are the sum of posterior probabilities of all trees with that clade.

### **5.2.3 Phylogeographic analysis of *Ostrea chilensis* Cytochrome Oxidase I sequences**

Following the study Ó Foighil, (1999) the phylogeographic relationship of *O. chilensis* was investigated. The analysis included the COI sequence obtained in this thesis research, the sequences used in the Ó Foighil study and sequences from a Chilean study (unpublished data) sourced from GenBank. *Ostrea stentina* and *Ostrea angasi* are used as outgroups as in the Ó Foighil study. GenBank accessions for this study are listed in Table 5.2. A Bayesian

phylogenetic reconstruction was implemented using the same steps outlined in the previous section.

## 5.3 Results

### 5.3.1 Mitochondrial Genome Construction

The amount of mitochondrial DNA recovered from the 454 DNA database was compared to the work of Leese *et al.* (2012) and reproduced in Figure 5.2. A total of 10,086 bp of DNA sequence data from the 454 read set was aligned to the published mitochondrial genome of *O. edulis*. This represents 65.4 percent of the 16,320 bp *O. edulis* genome (Figure 5.3 and Table 5.3). The mean depth of coverage is the number of sequence reads that overlap. The coverage of the reference sequence refers to the percentage of the total genes as compared to the *Ostrea edulis* genes and finally the percentage pairwise identity refers to the match of DNA bases between the *Ostrea chilensis* gene region and the reference gene. The overall depth of coverage had a mean of 1.2, and the maximum depth was 7 sequences, this was generated with 38 sequences with a mean length of 490 bp. The levels of coverage for the mitochondrial coding regions are detailed in Table 5.3. Mean depth of coverage refers to the number of sequences that overlap a gene region from the 454 data, coverage to the reference sequence refers to the how much of the region is actually represented by the particular region compared to the reference sequence, the percentage pairwise identity refers to the similarity to the reference sequences and the 454 data.

There were nine gaps identified in the mitochondrial genome. PCR primers were designed to amplify the DNA sequences between these gap regions (Table 5.4) and, if required, they can be used to close the gaps using standard PCR and Sanger sequencing methods.

### 5.3.2 Best-fit models of gene evolution for oyster mitogenomes

Table 5.5 shows the best-fit models of sequence evolution for each of the mitochondrial genes. For the concatenated alignment the best fitting model selected was the GTR+I+G model. However, this model does not apply to all genes, which justifies the use of a partitioned analysis.

### 5.3.3 Phylogenetic analysis based on all oyster mitogenomes

Figure 5.4 shows a plot of the MCMC chain, after the burn-in was discarded. It shows regular mixing between high and low likelihoods. The standard deviation of the parameter estimates was 0.13. The Geweke diagnostic gave a value of 1.389 (values greater than two are usually taken as an indication of lack of convergence).

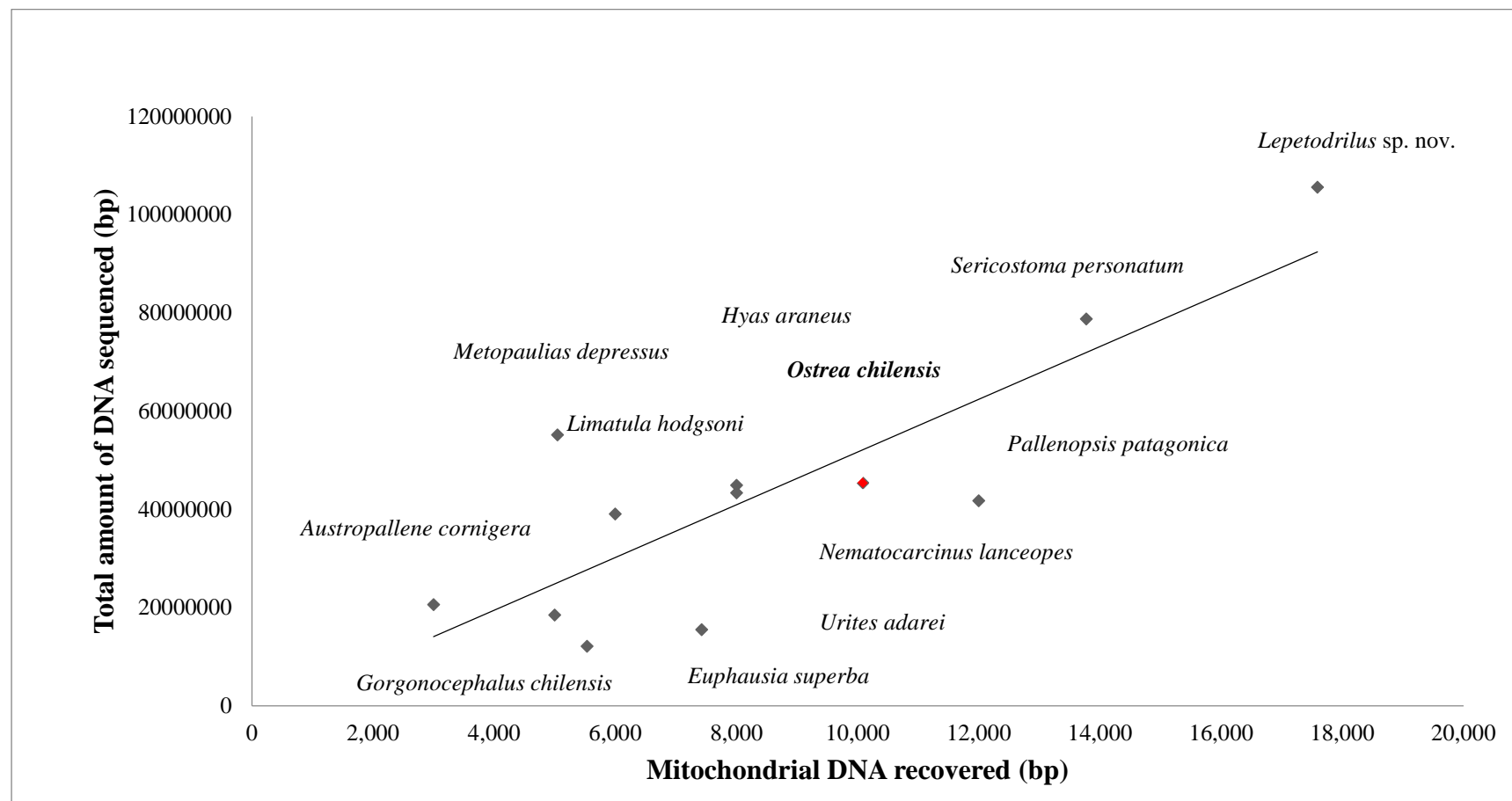
The phylogenetic tree produced (Figure 5.6) has a similar topology to that produced by Danic-Tchaleu *et al.* (2011). With the addition of *O. chilensis* and *O. lurida* this analysis places *O. deselamelonosa* in a more basal position with *O. lurida* being ancestral to all *Ostrea* species.

### 5.3.4 Phylogeographic analysis

The best model for the alignment of COI genes was GTR+I+G. The trace file shown in Figure 5.5 shows a plot of the MCMC chain, after the burn in was discarded. It shows regular mixing between high and low likelihoods. The standard deviation of the parameter estimates was 0.12. The Geweke diagnostic was 1.069, demonstrating that the chains converged. The un-rooted tree produced by Bayesian analysis (Figure 5.7) shows three clades. One clade contains the two species *O. angasi* and *O. stentina*. A second clade containing South Island *O. chilensis* samples was split from a clade containing North Island and Chilean populations; however the support values are low for this split.

**Table 5.1:** Mitogenomes used in the study and the number of hits when each genome was blasted against the 454 database

Species	Accession	Genome Size (bp)	Number of BLAST hits
<b>Crassostrea spp.</b>			
<i>Crassostrea angulata</i>	NC_012648	18,225	16
<i>Crassostrea ariakensis</i>	NC_012650	18,414	16
<i>Crassostrea gigas</i>	NC_001276	18,224	16
<i>Crassostrea hongkongensis</i>	NC_011518	18,622	17
<i>Crassostrea iredalei</i>	NC_013997	22,446	15
<i>Crassostrea nippona</i>	NC_015248	20,030	14
<i>Crassostrea sikamea</i>	NC_012649	18,243	19
<i>Crassostrea virginica</i>	NC_007175	17,244	14
<b>Ostrea spp.</b>			
<i>Ostrea denselamellosa</i>	NC_015231	16,277	37
<i>Ostrea edulis</i>	JF274008	16,320	40
<i>Ostrea lurida</i>	NC_022688	16,344	38
<i>Saccostrea mordax</i>	NC_013998	16,532	19
<b>Outgroup</b>			
<i>Argopecten irradians</i>	NC_009687	16,221	NA



**Figure 5.2:** Total number (bp) of mitochondrial DNA sequences recovered compared to total genomic sequences from various next generation sequence datasets, adapted from Leese *et al.* (2012). Amount of mitochondrial DNA recovered for *Ostrea chilensis* is highlighted in red

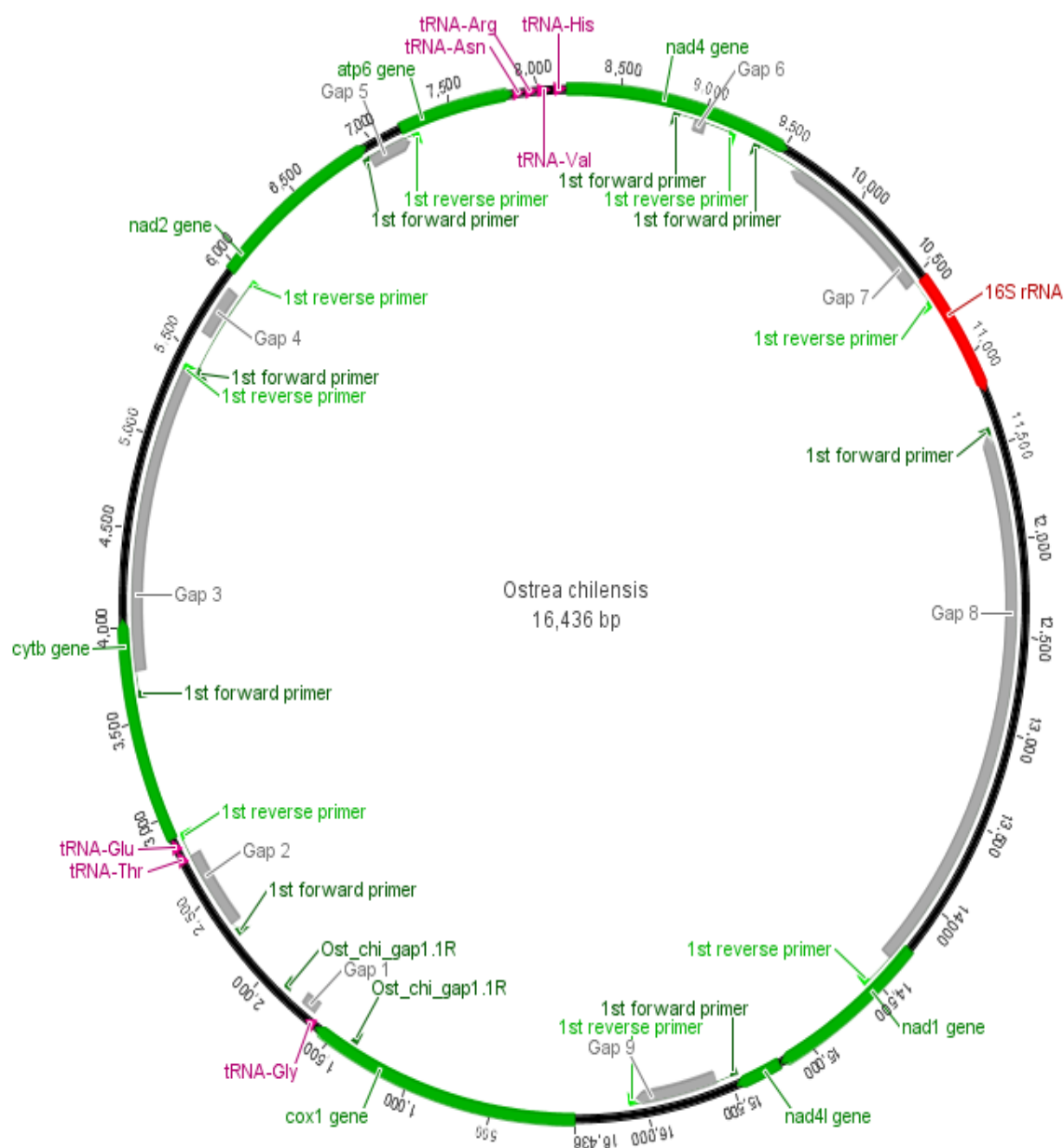
**Table 5.2:** partial COX1 sequences used in the phylogeographic study

Species	Location	Accession	Study
<i>O. chilensis</i>	Marlborough Sounds, South Island, NZ	NA	This study
<i>O. chilensis</i>	Foveaux Strait, South Island, NZ	AF112285	Ó Foighil (1999)
<i>O. chilensis</i>	Moturekareka, North Island, NZ	AF112289	Ó Foighil (1999)
<i>O. chilensis</i>	Quempillen, Chile	AF112286	Ó Foighil (1999)
<i>O. chilensis</i>	Chile	JF301917	Unpublished
<i>O. chilensis</i>	Chile	JF301920	Unpublished
<i>O. chilensis</i>	Chile	JF301921	Unpublished
<i>O. chilensis</i>	Chile	JF301922	Unpublished
<i>O. chilensis</i>	Chile	JF301924	Unpublished
<i>O. chilensis</i>	Chile	JF301929	Unpublished
<i>O. chilensis</i>	Chile	JF301931	Unpublished
<i>O. angasi</i>	NZ	AF112287	Ó Foighil (1999)
<i>O. stentina</i>	NZ	AF112288	Ó Foighil (1999)

**Table 5.3:** Summary Statistics for mitochondrial genome coverage

Coding Region	Mean Depth of Coverage	Coverage of Reference Sequence (bp)	Percentage Pairwise Identity (bp)
COI	2.6	61.8	79
tRNA-Gly	3.4	100	79
COII	1.2	59	58.1
tRNA-Ile	0	0	0
tRNA-Thr	1	71.8	93.5
tRNA-Glu	1	100	89.9
CYT B	3.8	76.7	83.5
COIII	0	0	0
tRNA-Met	0	0	0
tRNA-Ser	0	0	0
tRNA-Leu	0	0	0
tRNA-Pro	0	0	0
16 S RNA	1.1	47.5	64.3
NAD 2	1.1	98.7	76
tRNA-Cys	0	0	0
tRNA-Tyr	0	0	0
tRNA-Arg	2	100	94.9
tRNA-Val	2	100	82.5
tRNA-His	2	100	93.8
APT 6	1	94.8	66.3
tRNA-Asn	1	100	58.9
NAD 4	1.2	100	79.7
12 S RNA	0.1	6.5	98.3
16 S RNA	2.2	100	87.1
NAD 5	0	0	0
NAD 6	0	0	0
tRNA-Gln	0	0	0

NAD 3	0	0	0
tRNA-Lys	0	0	0
tRNA-Leu	0	0	0
tRNA-Phe	0	0	0
tRNA-Ala	0	0	0
NAD 1	0.8	77.5	82.4
NAD4l	1	100	84.2
tRNA-Trp	1	62.5	75.6
tRNA-Asp	0	0	0



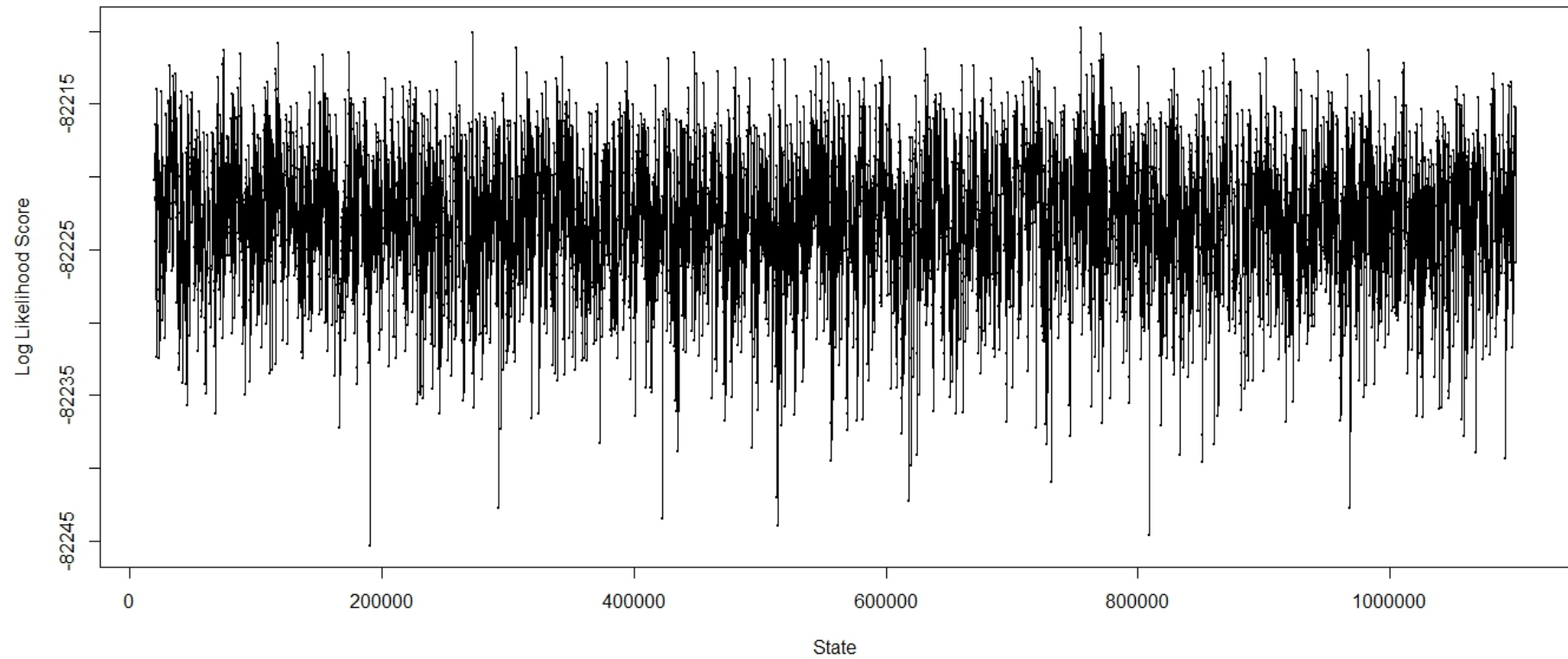
**Figure 5.3:** partial draft mitochondrial genome of *Ostrea chilensis*

**Table 5.4:** PCR primers designed to amplify DNA in the gap regions found in the mitochondrial genome

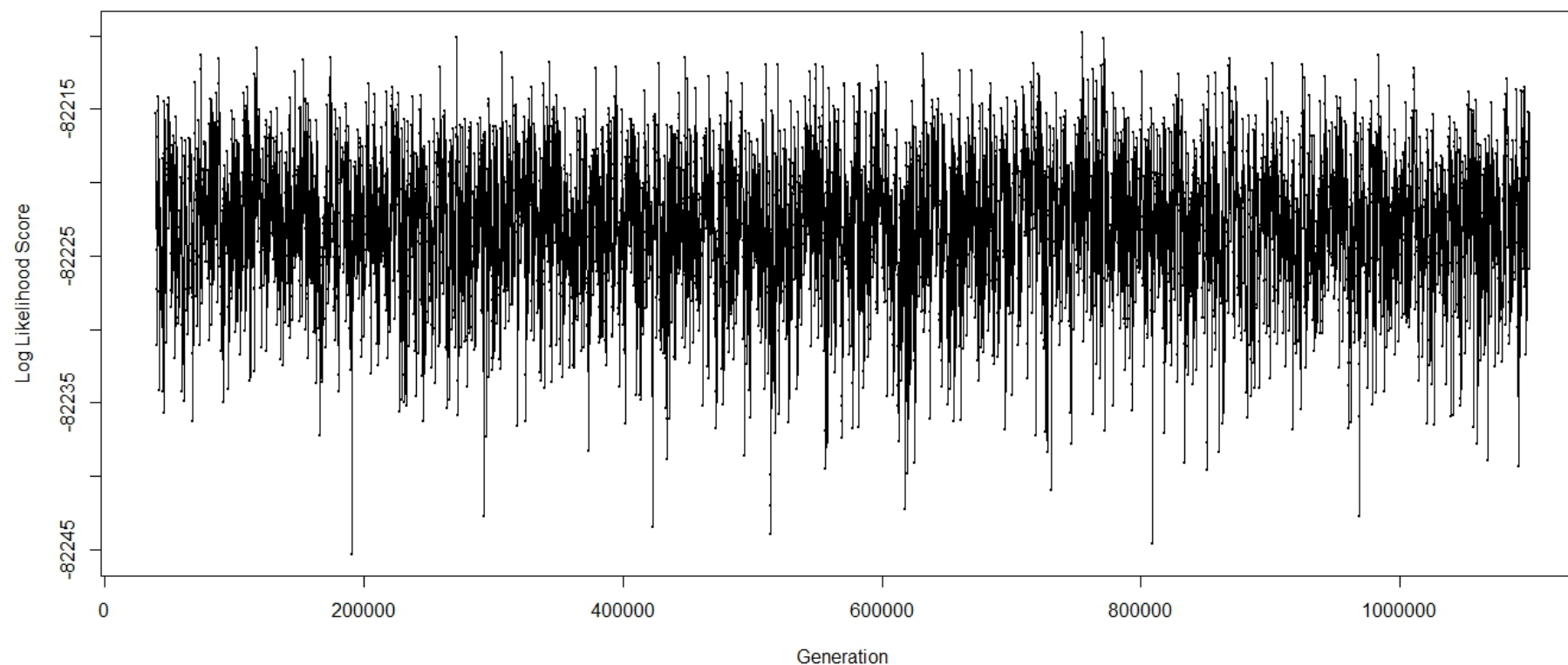
Gap	Size (bp)	Forward primers	Sequence	Reverse primers	Sequence
1	120	<i>Ost_chi_Gap1.1F</i>	CCGCGTCGAGTTGTTGACTA	<i>Ost_chi_Gap1.1R</i>	ACATTCGCCGCCTTACATGA
2	440	<i>Ost_chi_Gap2.1F</i>	GCTCCGCTTACGACTAACACT	<i>Ost_chi_Gap.2.1R</i>	CATCTAGGCGGGGAAGTTCC
3	1614	<i>Ost_chi_Gap3.1F</i>	GCAGATCCCTTACAGTGGCA	<i>Ost_chi_Gap.3.1R</i>	CAAGGCACACAGGGGATAGG
4	281	<i>Ost_chi_Gap4.1F</i>	GAGAGACGCCGACCACTCTA	<i>Ost_chi_Gap.4.1R</i>	CCACAAATGTAAGTCTACCAAAAGA
5	259	<i>Ost_chi_Gap5.1F</i>	GAAACGTGTGCGGAGGAATG	<i>Ost_chi_Gap.5.1R</i>	CATACAGCAGCTGACGGGAA
6	73	<i>Ost_chi_Gap6.1F</i>	GGTCAGTGGTTCTGCTGGAT	<i>Ost_chi_Gap.6.1R</i>	GGTCAGTGGTTCTGCTGGAT
7	935	<i>Ost_chi_Gap7.1F</i>	ACTTAGTAGTGAGCGGGGGT	<i>Ost_chi_Gap.7.1R</i>	GGCGGCCGTTTACAAGTTTT
8	2923	<i>Ost_chi_Gap8.1F</i>	TAGAGCTTGGCCATGTGTCC	<i>Ost_chi_Gap.8.1R</i>	GCAGAGAAGCCACTTGAAGG
9	490	<i>Ost_chi_Gap9.1F</i>	TAGCCAGGGGGAGTAAGCTT	<i>Ost_chi_Gap.9.1R</i>	TAGCCAGGGGGAGTAAGCTT

**Table 5.5:** Comparison of best fit models for each of the protein coding genes

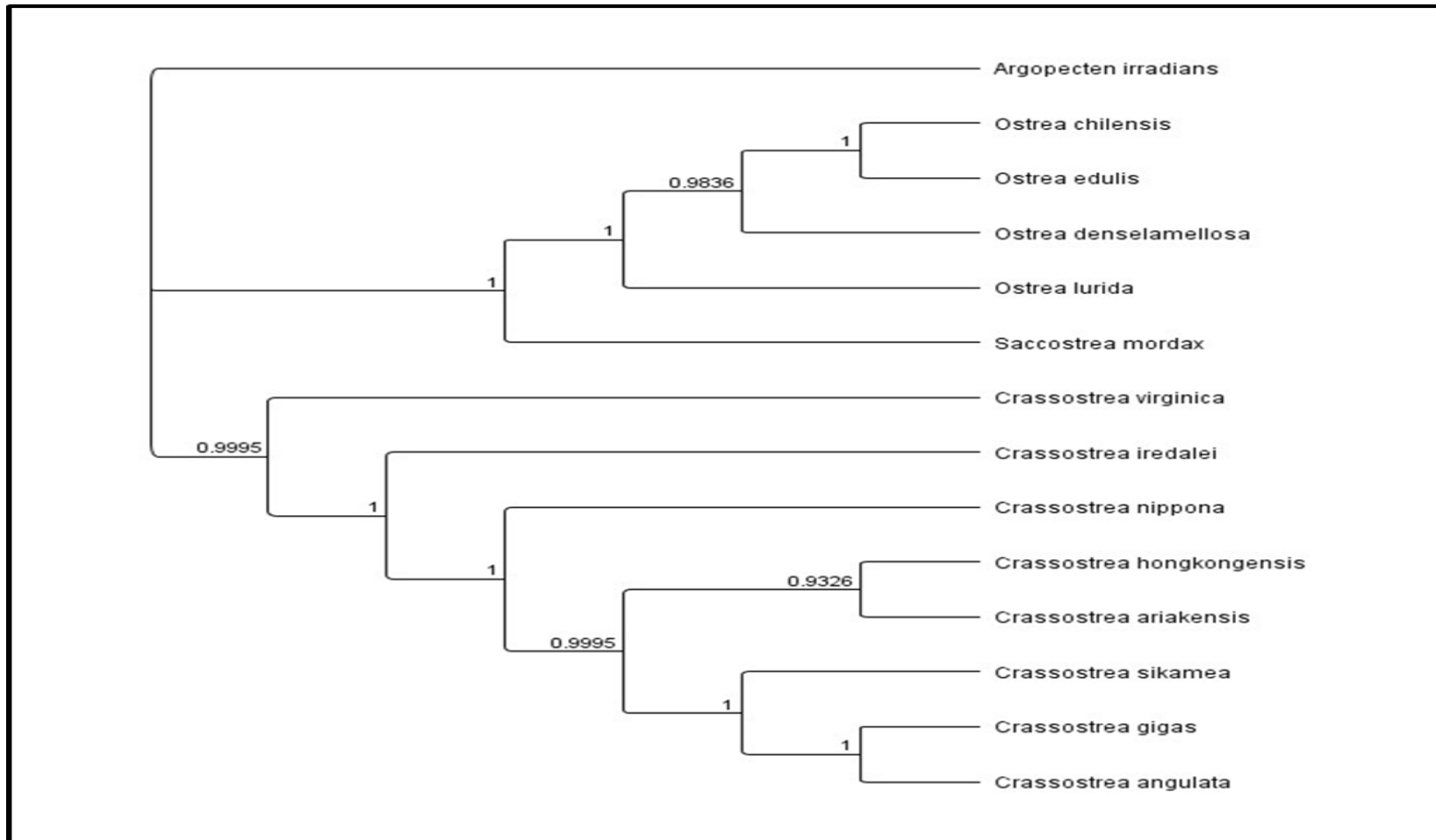
<i>Gene</i>	<i>Model</i>	<i>InL</i>	<i>p-inv</i>	<i>gamma</i>
Concatenated Alignment	GTR+I+G	82201.3	0.041	0.399
ATP6	TVM+I+G	4628.78	0.275	0.661
COI	GTR+I+G	9811.63	0.458	0.754
COII	TIM3+I+G	4316.94	0.44	0.918
COIII	TVM+I+G	6673.68	0.233	0.734
CYTB	GTR+I+G	8536.62	0.3	0.861
NAD4l	TVM+I+G	4628.78	0.275	0.661
ND1	TVM+I+G	6601.22	0.296	0.891
ND2	TVM+I+G	8921.33	0.152	1.198
ND3	TVM+I+G	2600.6	0.3	1.073
ND4	GTR+I+G	10667.5	0.266	0.947



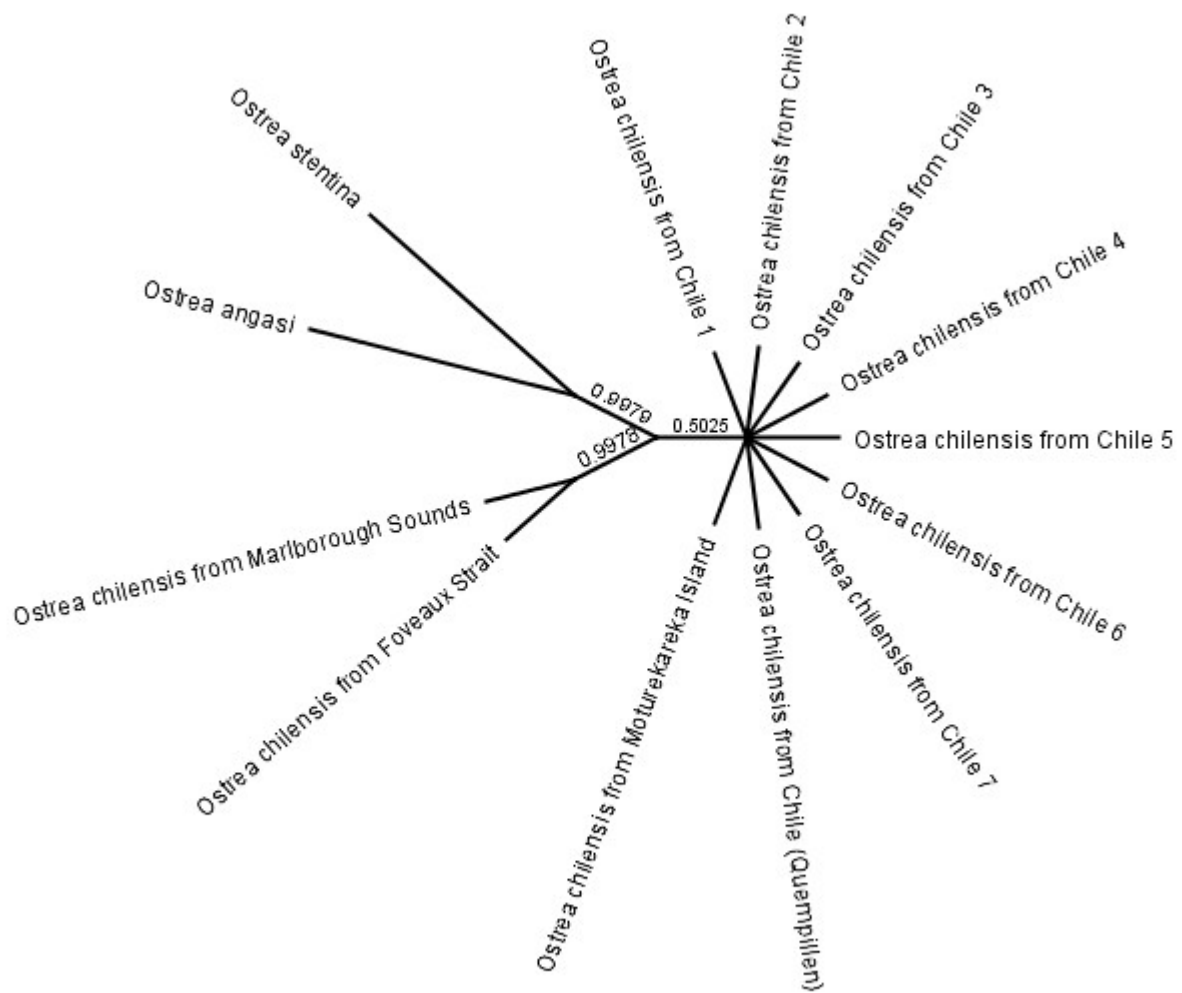
**Figure 5.4:** Trace plot of the MCMC chain for the mitogenome phylogenetic reconstruction



**Figure 5.5:** Trace plot of the MCMC chain for phylogeography analysis



**Figure 5.6:** Phylogenetic tree produced using Bayesian methods. Branch supports are Bayesian posterior probabilities. Out group is *Argopecten irradians*



**Figure 5.7:** Phylogenetic tree based on Bayesian analysis using partial cytochrome oxidase 1 gene (COI) alignments for *Ostrea chilensis*. *Ostrea stentina* and *Ostrea angasi* are used as outgroups.

## 5.4 Discussion

The first hypothesis is accepted, 10,086 bp of mitochondrial DNA was found in the 454 DNA database. Based on the size of the *O. edulis* mitogeneome, it was estimated that this represents around 65% of the full mitogenome sequence of *O. chilensis*. Leese *et al.* (2012) found a significantly positive linear correlation between genomic library size and amount of mitochondrial DNA sequences recovered. The amount of mitochondrial DNA recovered in the present study was compared to results reported by Leese *et al.* (2012) (Figure 5.1). The finding here suggests that around  $1 \times 10^8$  bp of DNA sequence reads are required from a genomic sequencing run to recover the full mitogenome sequence for an average sized mitogenome of 16 – 20kbp.

The second hypothesis was also accepted: the phylogenetic analysis using mitochondrial protein coding genes (Figure 5.6) showed that *O. chilensis* forms a sister taxa relationship with *O. edulis*. *O. lurida* and *O. densellamellosa* are ancestral to this relationship. However, *O. angasi*, was found to be sister taxa to *O. edulis* in the study of Danic-Tchaleu *et al.* (2011), using only the COI gene, this species was not included in the present study as it is yet to have its mitogenome fully sequenced. A greater taxonomic coverage within the genus is needed to clarify the phylogenetic position and test scenarios for the evolutionary origins of *O. chilensis*. These questions can be further investigated when the whole mitogenome of *O. chilensis* and all oyster species has been completed. A limitation of this part of study is that there are still many mitogenomes to be constructed within the oyster family; the final topology of the phylogenetic tree will be dependent on the taxa used and the level of genomic information obtained. The phylogenetic position of *O. chilensis* was determined using some missing data; it is unlikely the phylogenetic position of the species will change with greater genetic data as studies have shown that topology of a phylogeny does not change due to missing data (Lemmon *et al.* 2009).

The phylogeographic tree reported in the present study (Figure 5.7) using a COI sequence recovered from the 454 DNA database (*O. chilensis* from Marlborough Sounds), and seven Chilean samples (from an unpublished study found on GenBank), supported the suggestion of Ó Foighil *et al.* (1999): that the Chilean population arrived by rafting from the North Island of New Zealand. The North Island sample (from Moturekaraka) is more closely related to the

Chilean individuals (from Quempillen) compared to the two South Island individuals (Marlborough Sounds and Foveaux Strait). Also, the Chilean samples cluster in a ‘star-like’ phylogeny this suggests a recent arrival from a limited number of individuals. Ó Foighil *et al.* (1999) used COI sequences to suggest that *O. chilensis* rafted to Chile from a North Island ancestor. The idea of a vicariant origin of this species was ruled out as the only fossil evidence found post-dates New Zealand’s separation from Gondwanaland. The possibility of anthropogenic transport was also ruled out because radio-carbon dating of Chilean fossils suggests they were present before the arrival of humans in New Zealand. Although to make the conclusion in the absence of fossil data that the Chilean population was established by rafting from the North Island, might rest on too many assumptions about the genetic data as the sample size is very small, and may not include the full extent of genetic variation of the species, inclusion of further samples could result in a different tree topology. More gene regions and a greater number of individuals would be needed to more thoroughly test this idea.

Although the full mitogenome of *O. chilensis* has not yet been fully sequenced, this study was an important first step towards a greater understanding of mitochondrial genome evolution in oysters, and a step towards full mitogenome sequencing as primers have been designed to amplify the remaining gaps. The mitochondrial DNA can be used for further studies, such as the investigation of female mediated gene flow in *O. chilensis*, which could reveal patterns that are not apparent when using nuclear markers. In a study of the population genetics of *Ostrea edulis* using mitochondrial DNA (Hedgecock *et al.* 2007), mitochondrial haplotype diversity was found to have high among population variance, reflecting smaller effective population size in some locations. Also a ten-fold difference was observed in the same study between  $F_{st}$  obtained from mitochondrial and nuclear genomes. This could be due to variance in reproductive success between males and females. Analyses of mtDNA markers have been used to identify stock structure in a variety of fishes, including eels (Awise *et al.* 1986), bluefish (Graves *et al.* 1992), red drum (Gold *et al.* 1993), and snapper (Chow *et al.* 1993). More recently a combination of mitochondrial and microsatellite markers have been used to identify brood stock of grouper for aquaculture purposes (Kuo *et al.* 2014).

Some studies have shown a link between mitochondrial DNA polymorphisms and performance traits such as meat quality (Mannen 2003) and fertility (Cummins *et al.* 2002) in

Cattle. It remains to be seen whether mitochondrial polymorphism can be found to be associated with traits in aquaculture species.

This chapter has demonstrated the utility of 454 DNA sequencing datasets for developing genomic resources. There are further avenues of research that could be explored within the 454 data set, such as the identification of transposons, identification of rRNA genes and nuclear genes – there are a range options for further research.

## 5.5 Summary

1. This work provides a useful first step to full mitogenome sequencing and further studies into the evolution of the mitogenome in oysters. The analyses presented in this chapter showed that mitochondrial DNA fragments are present in the 454 DNA database – these fragments were used to assemble the partial mitogenome of *Ostrea chilensis*. PCR primers have been designed to amplify the remaining gaps in the genome
2. Mitochondrial gene fragments developed in this study were used to test phylogenetic relationships reported in previous studies, this work confirmed earlier phylogenetic work concerning the phylogenetic position of *Ostrea chilensis* (Jozefowicz and Ó Foighil, 1998; Ó Foighil, 2000).
3. A phylogeographic analysis supported the suggestion of Ó Foighil *et al.* (1999) that the Chilean populations of *Ostrea chilensis* arrived by rafting from an ancestral New Zealand population from the North Island.

## 5.6 References

- Awise JC, Helfman GS, Saunders NC, Hales LS. (1986). Mitochondrial DNA differentiation in North Atlantic eels: population genetic consequences of an unusual life history pattern. *Proceedings of the National Academy of Sciences*, 83(12), 4350-4354.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research* 27: 1767–1780.
- Brandley MC, Schmitz A, Reeder TW (2005). Partitioned Bayesian Analyses, Partition Choice, and the Phylogenetic Relationships of Scincid Lizards. *Systematic Biology*, 373-390.
- Brito P, Edwards S (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* 135: 439–455
- Cummins JM, Greeff, J, Lymbery AJ (2002). Mitochondrial DNA polymorphisms and fertility in beef cattle. *Theriogenology*, 57(6), 1603-1610.
- Danic-Tchaleu G, Heurtebise S, Morga B, Lapègue S (2011). Complete mitochondrial DNA sequence of the European flat oyster *Ostrea edulis* confirms Ostreidae classification. *BMC research notes*, 4(1), 400
- Darriba D, Taboada GL, Doallo R, Posada D (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8), 772-772.
- Doucet-Beaupré H, Breton S, Chapman EG, Blier PU, Bogan A.E, Stewart DT, Hoeh WR. (2010). Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evolutionary Biology*, 10(1), 50.
- Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113.
- Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, JM *et al.* (eds.) Bayesian Statistics 4. Clarendon Press, Oxford UK
- Gold JR, Richardson LR, Furman C, King TL. (1993). Mitochondrial DNA differentiation and population structure in red drum (*Sciaenops ocellatus*) from the Gulf of Mexico and Atlantic Ocean. *Marine Biology*, 116(2), 175-185.

- Graves JE (1998). Molecular insights into the population structures of cosmopolitan marine fishes. *Journal of Heredity*, 89(5), 427-437.
- Hedgecock D, Launey S, Pudovkin AI, Naciri Y, Lapegue S, Bonhomme F. (2007). Small effective number of parents ( $N_b$ ) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. *Marine Biology*, 150(6), 1173-1182.
- Jozefowicz CJ, Ó Foighil D. (1998). Phylogenetic analysis of southern hemisphere flat oysters based on partial mitochondrial 16S rDNA gene sequences. *Molecular Phylogenetics and Evolution*, 10(3), 426-435.
- Kuo HC, Hsu HH, Chua CS, Wang TY, Chen YM, Chen TY. (2014). Development of Pedigree Classification Using Microsatellite and Mitochondrial Markers for Giant Grouper Broodstock (*Epinephelus lanceolatus*) Management in Taiwan. *Marine Drugs*, 12(5), 2397-2407.
- Leese F, Brand P, Rozenberg A, Mayer C, Agrawal S, Dambach J, Sands C J. (2012). Exploring Pandora's Box: potential and pitfalls of low coverage genome surveys for evolutionary biology. *PloS one*, 7(11)
- Lowe TM, Eddy S R (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 0955-964.
- Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. (2009). The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology*, 58(1), 130-145.
- Mannen H, Morimoto M, Oyama K, Mukai F, Tsuji S. (2003). Identification of mitochondrial DNA substitutions related to meat quality in Japanese Black cattle. *Journal of Animal Science*, 81(1), 68-73.
- Moritz C, Dowling TE, Brown WM. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, 269-292.
- Ó Foighil D, Taylor DJ. (2000). Evolution of parental care and ovulation behaviour in oysters. *Molecular Phylogenetics and Evolution*, 15(2), 301-313.

ÓFoighil D, Marshall BA, Hilbish TJ, Pino MA. (1999). Trans-Pacific range extension by rafting is inferred for the flat oyster *Ostrea chilensis*. *Biological Bulletin*, 122-126.

Plummer M (2006). CODA: output analysis and diagnostics for Markov Chain Monte Carlo simulations. <http://cran.Rproject.org>

Rasmussen DA, Noor MA (2009). What can you do with 0.1× genome coverage A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC Genomics*, 10(1), 382.

Ren J, Liu X, Jiang F, Guo X, Liu B (2010). Unusual conservation of mitochondrial gene order in Crassostrea oysters: evidence for recent speciation in Asia. *BMC Evolutionary Biology* 10:394

Rombel IT, Sykes KF, Rayner S, Johnston S A. (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene*, 282(1), 33-41.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Huelsenbeck JP (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539-542.

Shadel GS, Clayton DA (1997). Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry*, 66:409-435

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

Wilcox TP, Zwickl DJ, Heath TA, Hillis DM (2002). Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular phylogenetics and evolution*, 25(2), 361-371.

Wyman SK, Jansen RK, Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17), 3252-3255.

Xiao S, Wu X, Li L, Yu Z. (2013). Complete mitochondrial genome of the Olympia oyster *Ostrea lurida* (Bivalvia, Ostreidae). *Mitochondrial DNA*, (0), 1-2

Yuan Y, Li Q, Yu H., Kong L. (2012). The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications. *PloS one*, 7(2)

Zeng C, Thomas, LJ, Kelly M, Gardner JPA. (2014). The complete mitochondrial genome of the deep-sea sponge *Poecillastra laminaris* (Astrophorida, Vulcanellidae). *Mitochondrial DNA*, (0), 1-2.

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

# Chapter 6: Molecular Evolution in Oyster Mitochondrial Genes

---

*'As far as is known, synonymous mutations are truly neutral with respect to natural selection'*  
*King and Jukes*

## 6.0 Aims

The aims of this chapter were to describe the patterns and processes of molecular evolution in oyster mitogenomes. The focus is on the effect of transition and transversion mutations on evolutionary distance between oyster species and the codon usage in mitochondrial protein coding genes. The findings reported in this chapter will be compared to previously published studies of molecular evolution in other taxa in order to identify common patterns of genetic variation.

## 6.1 Introduction

An assumption about the mitochondrial genome (mtDNA) is its rate of DNA sequence evolution is constrained due to its products having important functions in the electron transport of cellular energy production. However, it has been shown that mtDNA appears to evolve more quickly than single copy nuclear DNA (Brown *et al.* 1979). Most of the evolutionary change is thought to be neutral to selection. The most common type of neutral changes to protein-coding DNA sequences are synonymous mutations, which do not change the type of amino acid encoded by a particular codon. A DNA substitution that results in an amino acid change is termed a non-synonymous mutation, and can either have no impact on fitness, or in some cases be deleterious or slightly deleterious. The neutral theory of molecular evolution suggests that most fixed alleles were never subjected to positive natural selection and that evolutionary change primarily occurs by neutral (or nearly-neutral) alleles randomly drifting to fixation (Kimura, 1984).

A widely used test for selection is to compare the ratio of synonymous (dN) vs non-synonymous (dS) mutations observed in a DNA sequence of a protein-coding gene. If there is no selection, synonymous and non-synonymous substitutions are expected to occur at the same rate; the expected ratio is  $dN/dS = 1$ . If  $dN/dS$  is  $>1$  then positive selection is thought to have acted on the gene during its evolutionary history. If the  $dN/dS$  ratio is  $<1$  then negative

selection is expected to have occurred (Mugal *et al.* 2013). Using this method positive selection was demonstrated to have acted on the *atp-2* gene in *Escherichia coli* (Endo *et al.* 1996) and in the *COX4* gene of simian primates (Wu *et al.* 1997).

Evolutionary distance between two species can be measured using the p-distance. This is simply the proportion of nucleotides that are different between two sequences. There are two types of substitution that can occur in a nucleotide strand depending on the chemical properties of the nucleotide base – that are either purines or pyrimidines. Substitutions within one structural class are termed transitions; substitution from one structural class to another are termed transversions. A transitional bias has been identified between most homologous sequences. Interestingly, this bias has been found to be more pronounced in mitochondrial genes than in nuclear genes (Wakeley, 1996). The investigation of differences between transitions and transversions is important in the understanding of patterns of molecular evolution, and the information can be used to inform phylogenetic methods aiding a greater understanding of sequence evolution. The p-distance can be calculated using just transitions or just transversions in order to compare the relative effect on the overall genetic distance between two species. This is useful, as rates of transistions and transversions are not always universal across evolutionary history.

Finally, the research in this chapter seeks to describe the patterns of codon usage in the oyster mitochondrial genes. The 20 commonly occurring amino acids in proteins are encoded by 61 different codons – this redundancy in the genetic code means that there are synonymous codons that encode the same amino acid. Codon usage bias is preference for a particular synonymous codon. It would seem logical codon usage would not come under selection since the resulting protein sequence is the same. However, the general consensus is that codon preference reflects a balance between mutation and selection for translational optimisation. Evidence for this has come from a number of studies: it has been shown that genomes have their own patterns of codon bias, in some genomes highly expressed genes – such as mitochondrial genes – have shifted there codon usage to a set of particular codons compared to other less expressed genes (Grantham *et al.* 1980). In many cases the pattern of codon usage has been shown to reflect the tRNA abundance in the cytoplasm (Ikemura, 1981). This seems to suggest that relative abundance of tRNA is the selective pressure driving codon usage.

In this chapter protein-coding genes will be examined and the patterns of evolution described using the full mitogenomes of *Ostrea* and *Crassostrea* species and comparing them to a model organism *Drosophila melanogaster* (this organism is used simply because it is a well-known model organism). The patterns of molecular evolution are also compared using the alignment of a partial cytochrome oxidase I (COI) sequence using all the oyster species with a full mitogenome on GenBank, and the *O. chilensis* sequence obtained in this study.

### 6.1.1 Hypotheses

**Hypothesis One: Levels of genetic distance are related to one of the positions in the codon.**

Transition mutations are more common when the evolutionary distance between species is small, transversion mutations become more common deeper into evolutionary history. Suggesting that, transition mutations reach a saturation point, because of back mutations and/or multiple mutations occurring at a site, or transversions become more common when evolutionary distance is larger.

**Hypothesis Two: Mitochondrial genes are not influenced by positive selection.**

Mitochondrial protein coding genes are functionally constrained and the majority of the DNA substitutions will be synonymous-type changes. A ratio of  $dN/dS \leq 1$  is expected for protein-coding genes of the oyster mitogenomes.

**Hypothesis Three: There is a taxon specific codon bias.**

Previous studies have reported a codon bias in some taxa, and this is thought to be due to variation in mRNA translation efficiency. The suggested mechanism is that it would be inefficient to produce a range of different tRNAs to translate the same amino acid. It is expected that when comparing the codon usage from the COI genes of all *Ostrea*, *Crassostrea* species and *D. melanogaster* there will be a different bias in each group. Furthermore, the codon bias is expected to be similar for each gene within the *Ostrea* mitogenomes.

## 6.2 Methods

### 6.2.1 Assessment of codon bias

To determine the pattern of non-random usage of synonymous codons in mitochondrial genes the relative synonymous codon usage (RSCU) was computed, this is the observed frequency of a codon divided by the expected frequency under the assumption of equal codon usage (Sharp *et al.* 1986). This is given by:

$$\text{RSCU} = X_i / \bar{X}.$$

$X_i$  is the observed number of the  $i$ -th codon for the amino acid, and  $\bar{X}$  is the average of  $X_i$  over all codons - that is -  $\bar{X} = \sum \frac{i}{X_i} / m$  where  $m$  is the number of different codons for the particular amino acid (Nei and Kumar, 2000). An  $\text{RSCU} > 1$  represents codons that are used more frequently than expected by chance whereas  $\text{RSCU} < 1$  represents codons that are used less frequently than expected by chance. This analysis was implemented in MEGA6 (Tamura *et al.* 2013).

The RSCU of individual codons was compared between the groups of species for each of the protein coding genes. The groups tested were all *Ostrea/Saccostrea* and *Crassostrea* species for which a full mitogenome has been sequenced and the model organism *Drosophila melanogaster*. The model organism *D. melanogaster* was used for comparison to see if a taxon specific bias could be detected (this species was chosen because it is a well-known invertebrate model species). Taxa used for this analysis are listed in chapter five (Table 5.1). The GenBank accession number for the *D. melanogaster* mitogenome is U37541.1.

### 6.2.2 Comparison of genetic distances

The measure of sequence divergence was assessed using p-distances, which is a simple measure of the number of amino acid differences between pairwise comparisons divided by the total number of amino acids in the sequence alignment. The p-distances were estimated using MEGA6.

The protein-coding genes from the mitogenomes of *Crassostrea* and *Ostrea* species were extracted from the full genome sequence using GENEIOUS and aligned using ClustalW (Larkin *et al.* 2007) using the codon alignment option implemented in MEGA6. The codons were converted into amino acids to check that there were no stop or nonsense codons within the sequence.

The number of non-synonymous substitutions per non-synonymous site from averaging over all sequence pairs was estimated using MEGA6. Standard error estimates were obtained by a bootstrap procedure (500 replicates). Analyses were conducted using the Nei-Gojobori model (Nei and Gojobori, 1986). All ambiguous positions were removed for each sequence pair for each of the protein coding genes.

## 6.3 Results

### 6.3.1 Codon usage bias

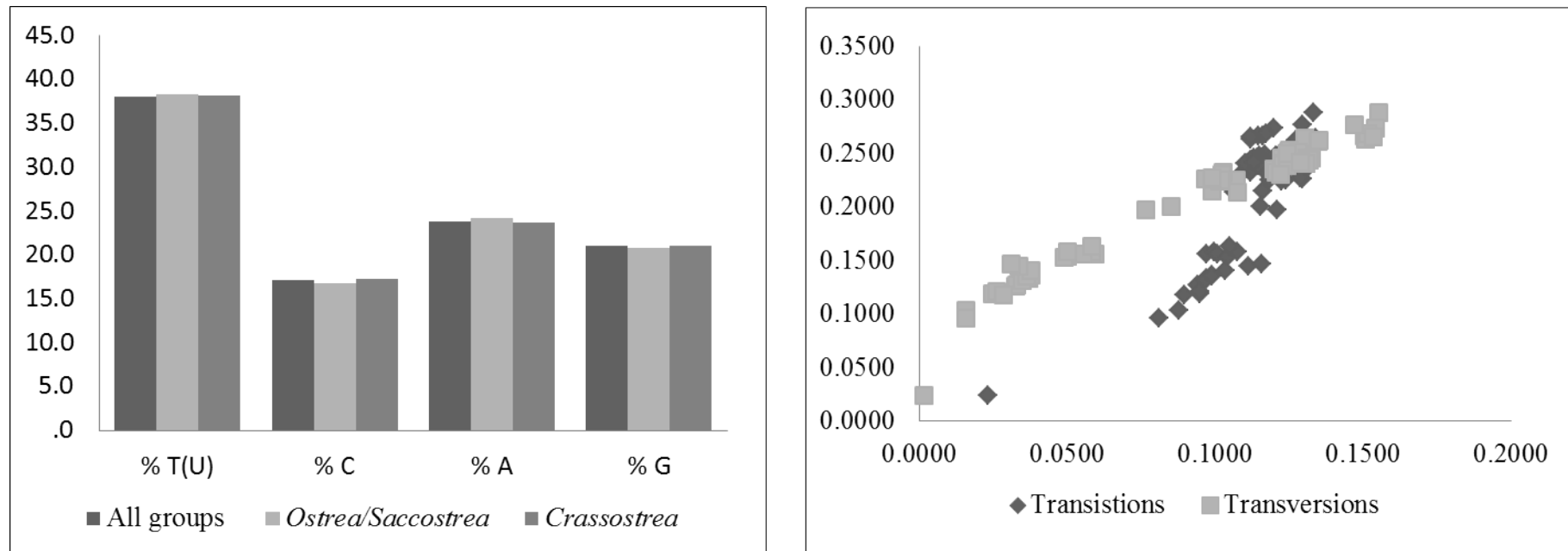
The codon biases that were calculated for each gene and compared between the three taxonomic groupings are presented in Table 6. 2 for the COI gene (summary statistics for all other genes are shown in the appendix). Table 6.3 shows a comparison of the codons that have the highest bias for each amino acid for each mitochondrial gene compared for all of the *Ostrea* species.

### 6.3.2 Genetic distance

Across all genes and taxa the pattern of base composition was very similar, and all genes were AT-rich. Figure 6.1a shows the base composition for the COI gene (all the other genes are shown in the appendix). The NAD2 gene had the highest number of variable sites (72%) among taxa and COI had the lowest number of variable sites among taxa (44%). In all genes the third codon position had the largest p-distance. Within in the *Ostrea/Saccostrea* taxa the gene COIII had the largest p-distance (0.290), this is contrasted within the *Crassostrea* group which has the largest p-distance for the gene NAD2 (0.328). Between the two groups *Ostrea/Saccostrea* and *Crassostrea* NAD2 had the largest p-distance (0.452).

### **6.3.3 Test of Neutrality - dN/dS ratios of mitochondrial genes among *Ostrea* and *Crassostrea* spp.**

For all of the protein coding genes examined the dN/dS ratio was less than 1 (Table 6.1a displays the dN/dS ratio for the COI gene data, all other genes are displayed in the appendix). The lowest ratio was 0.069 in the COII gene and the highest was 0.261 in the NAD2 gene.



**Figure 6.1 a (left):** Base composition of all CO1 genes compared for *Ostrea* and *Crassostrea* species. **Figure 6.1b (right):** pairwise comparison of all CO1 genes for all oyster species. Y-axis shows p-distance for either transitions or transversions the X axis shows the overall p-distance.

**Table 6.1 a:** Summary statistics for the COI gene compared between all oysters

Length of COI sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
1677	53	44	34	9	0.039

**Table 6.1b:** Comparison of p-distance for the COI gene compared between different codon positions and at different taxonomic levels

Overall Distance	p-distance	SE
All substitutions	0.1972	0.0006
Transitions	0.1094	0.0040
Transversions	0.0879	0.0038
1st position	0.0889	0.0074
2nd position	0.0401	0.0049
3rd position	0.4627	0.0082
<i>Ostrea/Saccostrea</i> within group mean	0.2007	0.0072
<i>Crassostrea</i> within group mean	0.1542	0.0045
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.2454	0.0066

**Table 6.2:** Relative synonymous codon usage (RSCU) for the COI gene in two oyster genera and *D. melangaster*. The first column of RSCU values is the *Ostrea* group, the second is the *Crassostrea* group and the third column is *D. melanogaster*.

<b>Phe</b>	UUU	1.64	1.63	1.85	<b>Ser</b>	UCU	2.64	2.41	2.20	<b>Tyr</b>	UAU	1.21	1.46	1.44	<b>Cys</b>	UGU	1.88	1.42	0.00
	UUC	0.36	0.37	0.15		UCC	0.21	0.14	0.20		UAC	0.79	0.54	0.56		UGC	0.12	0.58	0.00
<b>Leu</b>	UUA	2.46	1.82	4.89		UCA	2.05	2.01	3.60	<b>STOP</b>	UAA	2.00	1.50	2.00	<b>Trp</b>	UGA	1.38	1.22	2.00
	UUG	1.15	1.05	0.09		UCG	0.21	0.2	0.20		UAG	0.00	0.50	0.00		UGG	0.62	0.78	0.00
	CUU	1.05	2.12	0.37	<b>Pro</b>	CCU	2.29	2.02	2.24	<b>His</b>	CAU	1.33	1.38	1.78	<b>Arg</b>	CGU	1.41	1.45	0.8
	CUC	0.20	0.09	0.00		CCC	0.23	0.32	0.16		CAC	0.67	0.62	0.22		CGC	0.15	0.18	0.00
	CUA	0.71	0.70	0.65		CCA	0.91	1.25	1.44	<b>Gln</b>	CAA	1.42	1.19	2.00		CGA	1.63	1.55	3.20
	CUG	0.44	0.21	0.00		CCG	0.57	0.42	0.16		CAG	0.58	0.81	0.00		CGG	0.81	0.82	0.00
<b>Ile</b>	AUU	1.87	1.8	1.96	<b>Thr</b>	ACU	1.92	2.15	2.06	<b>Asn</b>	AAU	1.52	1.40	2.00	<b>Ser</b>	AGU	0.67	0.76	0.80
	AUC	0.13	0.2	0.04		ACC	0.47	0.38	0.24		AAC	0.48	0.6	0.00		AGC	0.08	0.35	0.00
<b>Met</b>	AUA	0.92	0.9	1.91		ACA	1.14	1.06	1.58	<b>Lys</b>	AAA	1.34	1.15	1.67		AGA	1.09	1.06	1.00
	AUG	1.08	1.1	0.09		ACG	0.47	0.41	0.12		AAG	0.66	0.85	0.33		AGG	1.05	1.06	0.00
<b>Valine</b>	GUU	1.32	1.71	1.82	<b>Ala</b>	GCU	1.98	1.53	3.14	<b>Asp</b>	GAU	1.21	1.21	1.87	<b>Gly</b>	GGU	1.31	0.99	0.51
	GUC	0.39	0.45	0.00		GCC	0.6	0.32	0.00		GAC	0.79	0.79	0.13		GGC	0.38	0.47	0.00
	GUA	1.28	1.01	2.06		GCA	0.96	1.59	0.76	<b>Glu</b>	GAA	1.28	1.00	2.00		GGA	1.02	1.24	3.40
	GUG	1.02	0.84	0.12		GCG	0.45	0.55	0.11		GAG	0.72	1.00	0.00		GGG	1.29	1.30	0.09

**Table 6.2:** Codon usage bias for the cytochrome oxidase II gene (COII)

	<i>Ostrea edulis</i>	<i>Ostrea chilensis</i>	<i>Ostrea denselamell-osa</i>	<i>Crassostrea ariakensis</i>	<i>Crassostrea gigas</i>	<i>Crassostrea ariakensis</i>	<i>Ostrea denselamellosa</i>	<i>Crassostrea hongkongensis</i>	<i>Crassostrea ariakensis</i>	<i>Crassostrea virginica</i>	<i>Saccostrea mordax</i>	<i>Argopecten irradians</i>
<b>Phe</b>	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU
<b>Leu</b>	UUG	UUA	UUG	UUA	UUA	UUA	UUA	UUG	UUG	UUA	UUG	UUA
<b>Ile</b>	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU
<b>Met</b>	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG
<b>Val</b>	GUU	GUU	GUA	GUU	GUU	GUU	GUU	GUU	GUU	GUU	GUU	GUU
<b>Ser</b>	UCU	UCA	UCU	UCU	UCU	UCU	UCU	UCU	UCU	UCU	UCU	UCU
<b>Pro</b>	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCU
<b>Thr</b>	ACA	ACA	ACU	ACU	ACU	ACU	ACU	ACU	ACU	ACU	ACU	ACU
<b>Ala</b>	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU
<b>Tyr</b>	UAU	UAC	UAU	UAU	UAU	UAU	UAU	UAU	UAU	UAU	UAU	UAU
<b>His</b>	CAU	CAU	CAC	CAU	CAU	CAU	CAU	CAU	CAU	CAU	CAU	CAU
<b>Gln</b>	CAA	CAG	CAG	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA
<b>Asn</b>	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU
<b>Lys</b>	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
<b>Asp</b>	GAU	GAU	GAU	GAC	GAC	GAU	GAU	GAU	GAU	GAU	GAU	GAU
<b>Glu</b>	GAG	GAG	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA
<b>Cys</b>	UGU	UGU	UGU	UGU	UGU	UGC	UGC	UGC	UGC	UGC	UGU	UGC
<b>Trp</b>	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG
<b>Arg</b>	CGU	CGA	CGG	CGG	CGG	CGC	CGC	CGC	CGC	CGC	CGC	CGG
<b>Ser</b>	AGU	AGC	AGC	AGC	AGC	AGC	AGC	CGC	CGC	CGC	CGC	AGU
<b>Gly</b>	GGU	GGU	GGU	GGU	GGU	GGG	GGG	GGG	GGG	GGG	GGG	GGG

**Table 6.4:** Comparison of the codon bias for each protein-coding gene from the mitochondrial genome of four *Ostrea* species. Codons shown are those with the highest RSCU values. Highlighted values are those which have a different codon to the other genes

	<i>COI</i>	<i>COII</i>	<i>COII</i>	<i>ND1</i>	<i>ND2</i>	<i>ND4L</i>	<i>CYTB</i>	<i>ATP6</i>
<b>Phe</b>	UUU	UUU	UUU	UUU	UUU	UUU	UUU	UUU
<b>Ile</b>	AUU	AUU	AUU	AUU	AUU	AUU	AUU	AUU
<b>Met</b>	AUA	AUA	AUA	AUA	AUA	AUA	AUA	<b>AUG</b>
<b>Valine</b>	GUU	GUU	GUU	GUU	GUU	<b>GUA</b>	GUA	GUU
<b>Ser</b>	UCU	UCU	UCU	UCU	UCU	UCU	UCU	UCU
<b>Pro</b>	CCU	CCU	CCU	CCU	CCU	CCU	CCU	CCG
<b>Thr</b>	ACU	ACA	ACU	ACU	ACU	ACU	ACU	ACU
<b>Ala</b>	GCU	GCU	GCU	GCU	GCU	GCU	<b>ACU</b>	GCU
<b>Tyr</b>	UAU	UAU	UAU	UAU	UAU	UAU	UAU	UAU
<b>His</b>	CAU	CAU	CAU	CAU	<b>CAC</b>	CAU	CAU	CAU
<b>Gln</b>	CAA	CAA	CAA	CAA	CAA	CAA	CAA	CAA
<b>Asn</b>	AAU	AAU	AAU	AAU	AAU	AAU	AAU	AAU
<b>Lys</b>	AAA	AAA	AAA	AAA	<b>AAG</b>	AAA	AAA	<b>AAA/AAG</b>
<b>Asp</b>	GAU	GAU	GAU	GAU	GAU	GAU	GAU	GAU
<b>Glu</b>	GAA	GAA	GAA	GAA	<b>GAG</b>	GAA	<b>GAG</b>	<b>GAG</b>
<b>Cys</b>	UGU	UGU	UGU	<b>UGC</b>	UGU	<b>UGC</b>	UGU	UGU
<b>Trp</b>	UGA	UGA	UGA	<b>UGG</b>	UGA	<b>UGG</b>	UGA	UGA
<b>Arg</b>	CGA	CGA	<b>CGU</b>	CGA	<b>CGU</b>	CGA	<b>CGU</b>	CGA
<b>Ser</b>	AGA	AGA	<b>AGU</b>	AGA	<b>AGU</b>	AGU	<b>AGG/AGA</b>	AGA
<b>Gly</b>	<b>GGU</b>	<b>GGU</b>	<b>GGG</b>	<b>GGG</b>	<b>GGA</b>	<b>GGU</b>	<b>GGA/GGG</b>	<b>GGG</b>

## 6.4 Discussion

As expected the overall p-distance among all taxa was largest for the 3<sup>rd</sup> codon position. The smallest p-distances were observed at the second codon position (Table 6.1b), which was also expected. Most substitutions in the 3<sup>rd</sup> codon position are synonymous, whereas all 2<sup>nd</sup> codon position substitutions are non-synonymous and these latter mutational-types are likely to be selected against. Across all taxa the gene NAD2 had the largest p-distance value (0.376), this pattern is not repeated at different taxonomic levels, within *Ostrea* and *Saccostrea* the gene with the highest p-distance value was COIII (0.290), this suggests that genes are evolving at different rates at different taxonomic levels. The typical pattern of more transitions compared to transversion mutations was observed in each of the genes: when the p-distance between taxa is small, this distance is mostly due to transition-type mutations between the more closely related species – the pattern becomes more complex between distantly related organisms as transversions tend to contribute equally or more to the overall distance (Figure 6.1b). This suggests that transition mutations reach a saturation point because of back mutations and/or multiple mutations occur at a site, or transversions become more common in deeper evolutionary history.

When comparing *Ostrea*, *Crassostrea* and the model organism *Drosophila melanogaster* (although there are some exceptions), there appears to be a conservation of codon bias (Table 6.2). The same codon bias seems to exist when comparing all protein coding mitochondrial genes in *Ostrea* (Table 6.3). The results found here with oysters are consistent with that found in other species: codon bias is a phenomenon found in many studies (for a review see Sharp *et al.* (1988)). The reasons for these biases was first discussed by Ikemura (1981, 1985) he showed in *Escherichia coli* that the frequency of codon usage is correlated with the relative abundance of isoaccepting tRNAs in the cell. This finding suggests that the translational machinery tends to use the most abundant tRNAs to produce the proteins. Other mechanisms of selection at the level of translation have been suggested. Kudla (2009) showed that folding of mRNA strands can be altered by different codons and that strong folding around the binding site of the ribosome can inhibit transcription. Splice sites on the mRNA strand can be altered by codon usage – synonymous mutations can introduce splice sites that subsequently affect the phenotype, therefore, bringing codon usage under selection (Pagani and Baralle (2004)). However, the most widely supported mechanism of codon bias is that proposed by Ikemura (1981, 1985). The mechanism predicts that the codon usage bias

should be the same across all genes. This prediction appears to be supported when comparing codon usage for each gene in the *Ostrea* species (Table 6.3), where there were exceptions and the difference between RSCU values was small (*data not shown*). This observation would support the hypothesis that the genes are under purifying selection - genes that accumulate mutations for non-synonymous mutations could still be eliminated by selection if they were inefficiently synthesising proteins. Purifying selection is also suggested by the consistent  $dN/dS = < 1$ , which was found for all gene regions in this study, this is consistent with the predictions of the neutral theory of evolution. Another possible cause of codon bias is mutational bias – mutational biases arise because there are underlying mechanisms that favour certain mutations. It has been shown that there is a correlation between the GC content of a genome and the GC content of codons (Muto and Osawa, 1987). Knight *et al.* (2001) found this relationship in the three major domains of life (archaea, bacteria and eukaryotes) they suggest that the GC content of the genome drives the GC content of the codons. This theory could not be examined using the oyster mitogenomes because there were not large differences between the GC content of each genome.

There was not a clear taxonomic pattern of codon bias within oysters, although this could be investigated using more sensitive statistical methods of ordination, for example, principle component analysis (PCA) or correspondence analysis (for example Zhang *et al.* 2011). The one limitation of this study is that I was not able to include all the gene regions for *O. chilensis* in the analysis due to them not being completely sequenced (reported in Chapter Five). However, it is not likely that inclusion of the full mitogenome of *O. chilensis* will greatly affect the results since a similar pattern of codon bias has been found when analysing a partial COII sequence from *O. chilensis* (Table 6.3).

The quote at the beginning of this chapter does not seem to hold - a conservation of codon bias across species and between genes has been found - it seems likely that synonymous changes are under selection. The selective pressure that has caused the conservation of codon bias would need further study to investigate and properly describe the mechanisms.

## **6.5 Summary**

1. This study suggests that evolution of the oyster mitochondrial genes is non-neutral. There are codon biases in mitochondrial genes and these biases are likely to be the result of purifying selection. The mechanistic cause of this bias was not studied, but translational selection is the most likely explanation.
2. As expected it has been shown that genetic distance varies depending on the gene region and codon position examined. This supports the suggestion made in the previous chapter that phylogenies should be produced using multiple gene regions due to different evolutionary rates.
3. A similar codon bias has been found within oyster mitogenomes – this is likely to be driven by transitional selection. Further studies into the codon usage could be conducted to see whether there is a taxon specific bias using ordination methods.

## 6.6 References

- Brown WM, George M, Wilson AC. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76(4), 1967-1971
- Endo T, Ikeo K, Gojobori T. (1996). Large-scale search for genes on which positive selection may operate. *Molecular Biology and Evolution*, 13(5), 685-690.
- Grantham R, Gautier C, Gouy C. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Research*, 8, 1893–1912
- Ikemura T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology*, 151(3), 389-409.
- Ikemura T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2(1), 13-34.
- Kimura M. (1984). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Knight RD, Freeland SJ, Landweber LF. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology*, 2(4)
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Higgins DG. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948.
- Mugal CF, Wolf JB, Kaj I. (2013). Why time matters: Codon evolution and the temporal dynamics of dN/dS. *Molecular Biology and Evolution*, 31(1):212–231
- Muto A, Osawa S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences*, 84(1), 166-169.
- Nei M, Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3:418-426.
- Nei M, Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press.

Pagani F, Baralle FE. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Reviews Genetics*, 5(5), 389-396.

Sharp PM, Li WH. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281-1295.

Sharp PM., Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F. (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Research*, 16(17), 8207-8211.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729

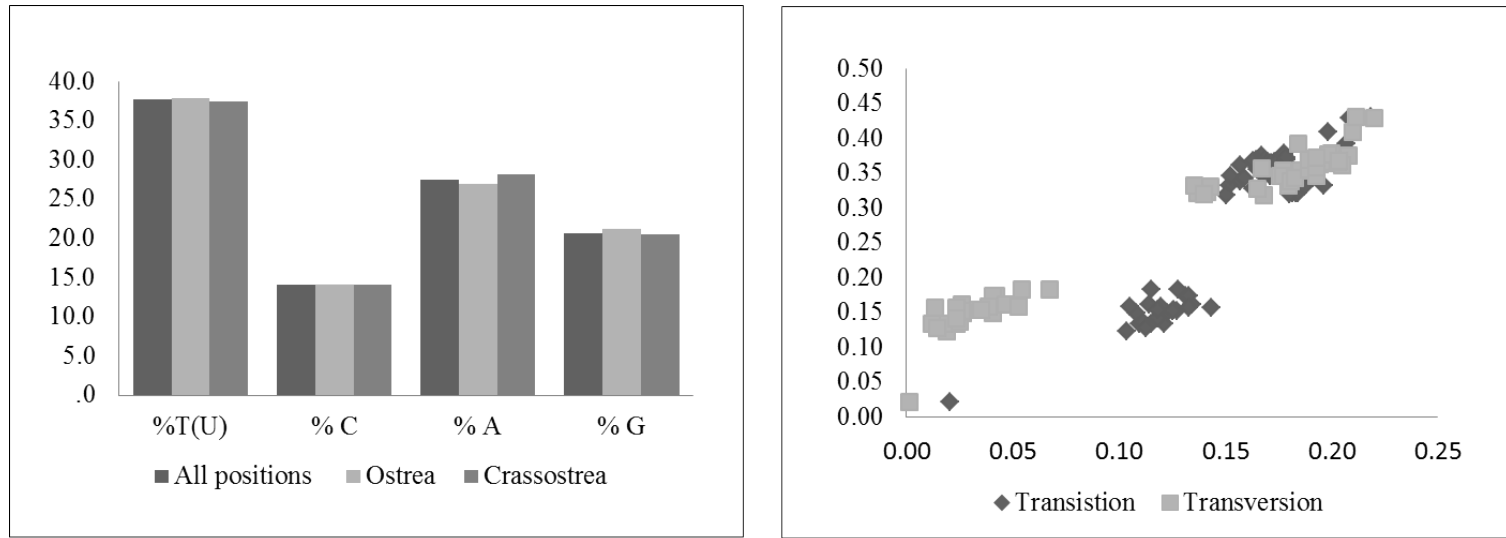
Wakeley, J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology and Evolution* 11:158-163

Wu W, Goodman M, Lomax MI, Grossman LI. (1997). Molecular evolution of cytochrome c oxidase subunit IV: evidence for positive selection in simian primates. *Journal of Molecular Evolution*, 44(5), 477-491.

Yang Z, Yoder AD. (1999). Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution*, 48(3), 274-283.

Zhang J, Wang M., Liu WQ, Zhou JH, Chen HT, Ma LN, Liu YS. (2011). Analysis of codon usage and nucleotide composition bias in polioviruses. *Virology Journal*, 8(1), 146.

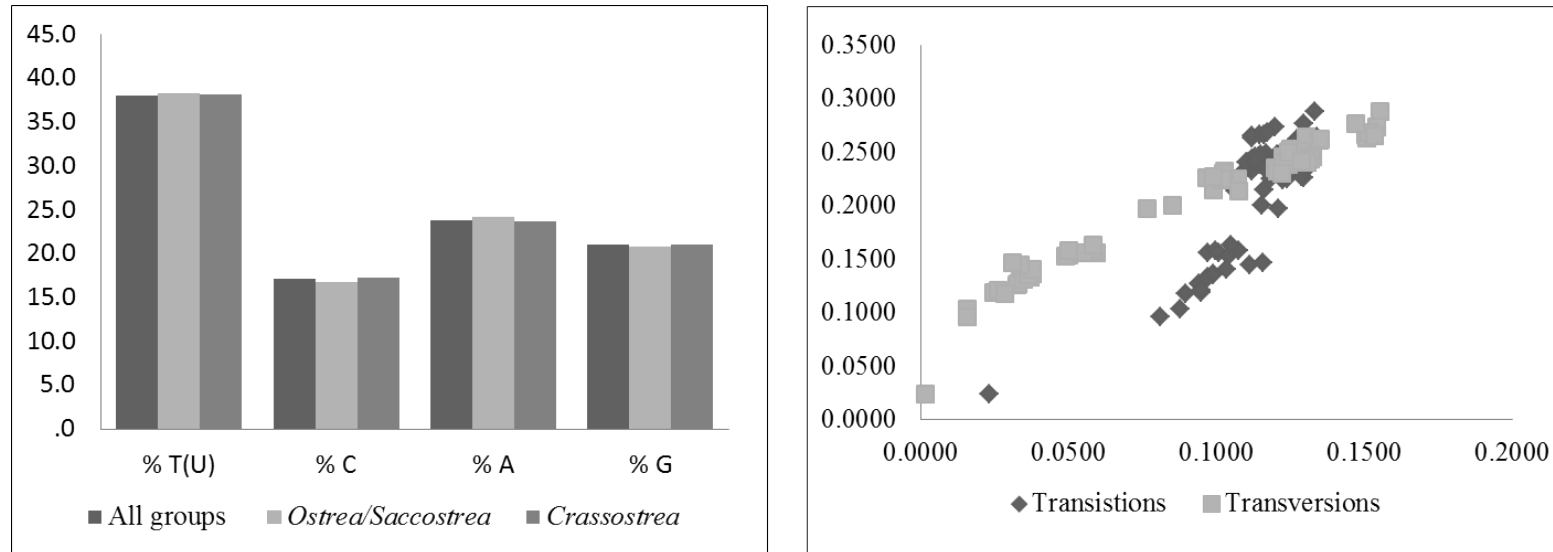
## 6.7 Appendix



Length of ATP6 sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
696	40	59	48	10	0.115

Overall Distance	p-distance	SE
All substitutions	0.281	0.011
Transitions	0.153	0.007
Transversions	0.127	0.008
1st position	0.218	0.016
2nd position	0.126	0.015
3rd position	0.490	0.011
<i>Ostrea/Saccostrea</i> within group mean	<b>0.260</b>	<b>0.013</b>
<i>Crassostrea</i> within group mean	0.188	0.009
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.366	0.0146

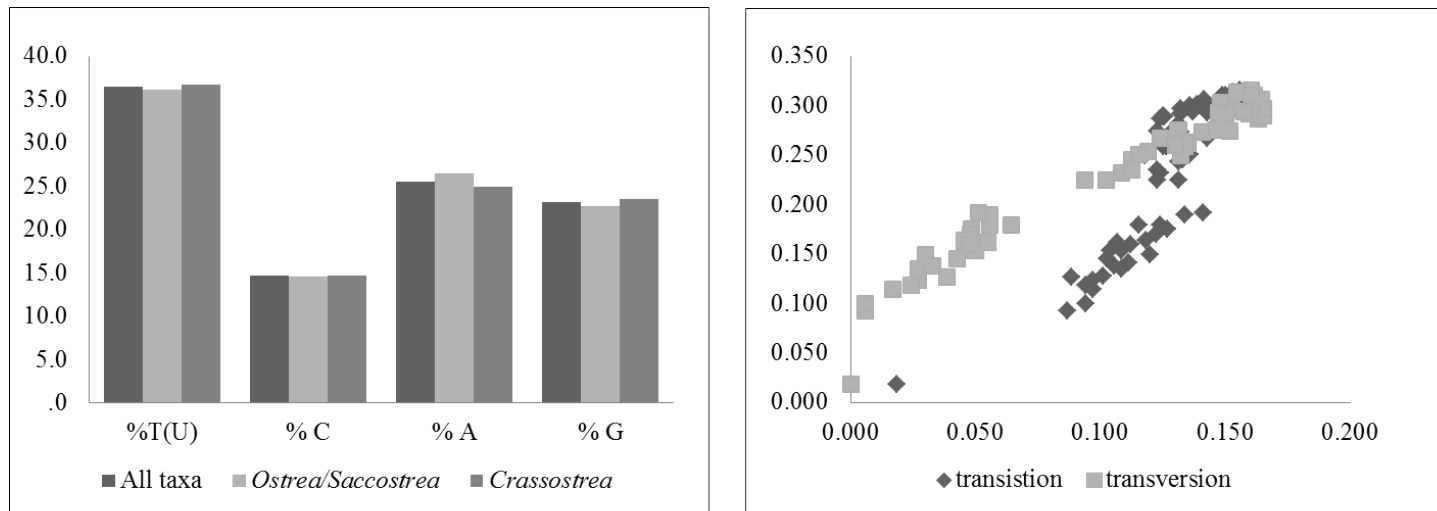
### Appendix 6.1: Summary statistics for the ATP6 mitochondrial gene



Length of COI sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
1677	53	44	34	9	0.039

Overall Distance	p-distance	SE
All substitutions	0.1972	0.0006
Transitions	0.1094	0.0040
Transversions	0.0879	0.0038
1st position	0.0889	0.0074
2nd position	0.0401	0.0049
3rd position	0.4627	0.0082
<i>Ostrea/Saccostrea</i> within group mean	0.2007	0.0072
<i>Crassostrea</i> within group mean	0.1542	0.0045
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.2454	0.0066

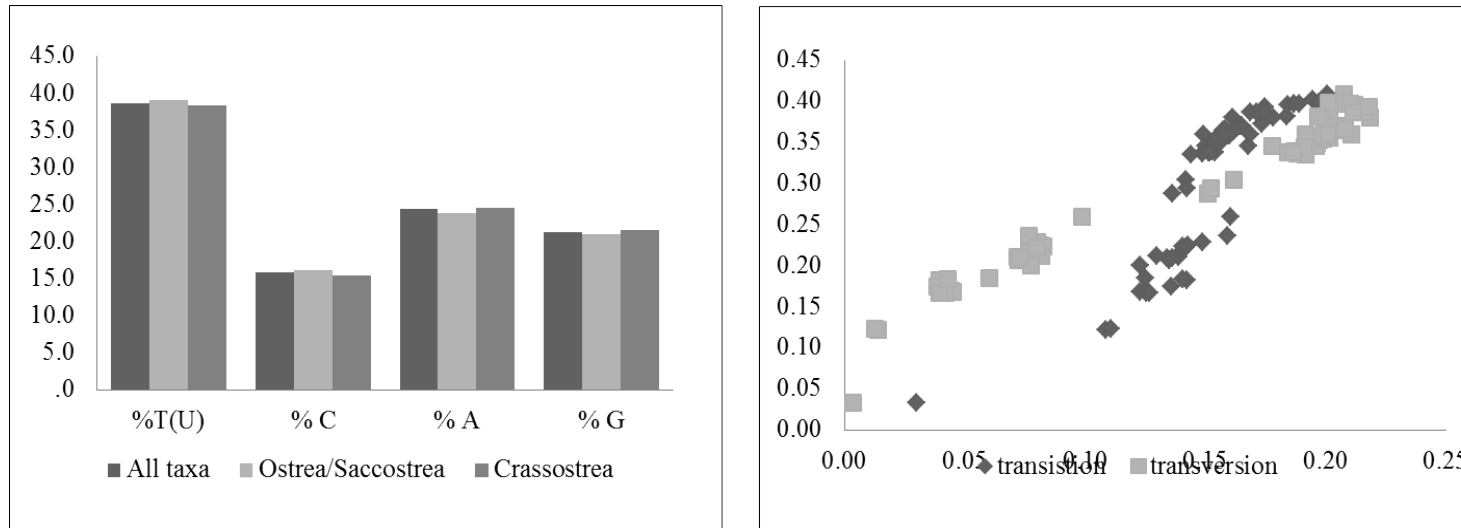
**Appendix 6.2:** Summary statistics for the COI mitochondrial gene



Length of COII sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
713	51	48	54	8	0.069

Overall Distance	p-distance	SE
All substitutions	0.2268	0.009
Transitions	0.1248	0.0062
Transversions	0.1038	0.006
1st position	0.1450	0.014
2nd position	0.0628	0.0106
3rd position	0.4778	0.0115
<i>Ostrea/Saccostrea</i> within group mean	0.2366	0.0113
<i>Crassostrea</i> within group mean	0.1620	0.008
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.2659	0.011

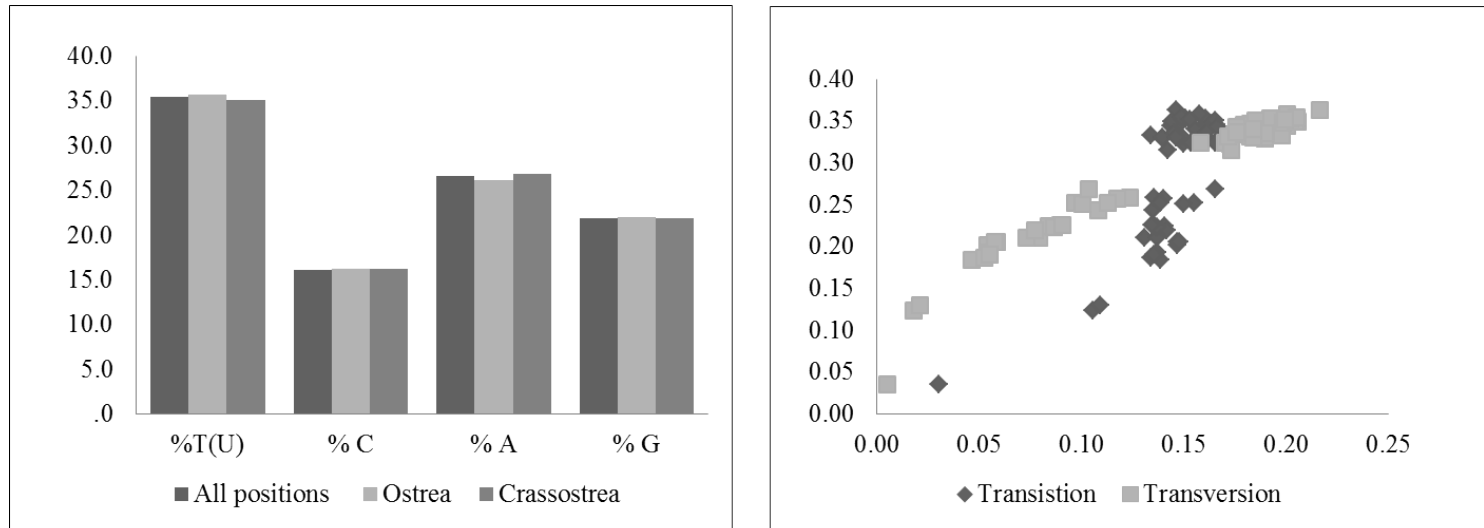
### Appendix 6.3: Summary statistics for the COII mitochondrial gene



Length of COIII sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
904	37	62	48	14	0.134

Overall Distance	p-distance	SE
All substitutions	0.298	0.008
Transitions	0.151	0.006
Transversions	0.147	0.006
1st position	0.231	0.015
2nd position	0.1590	0.013
3rd position	0.504	0.011
<i>Ostrea/Saccostrea</i> within group mean	0.290	0.010
<i>Crassostrea</i> within group mean	0.2457	0.010
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.360	0.011

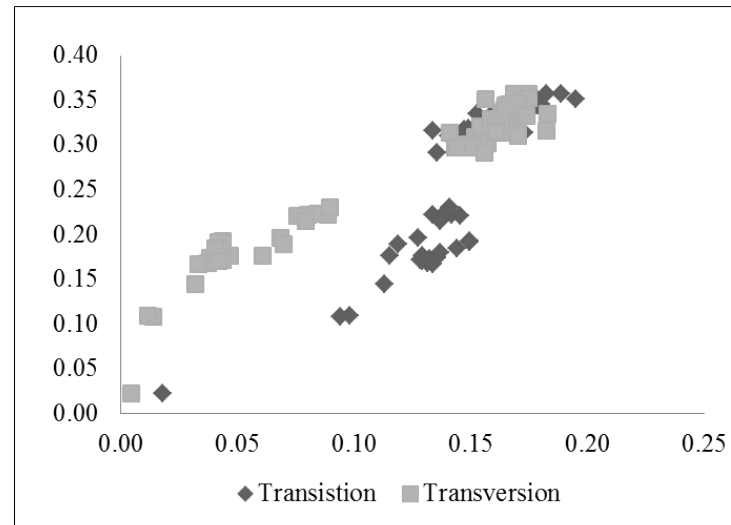
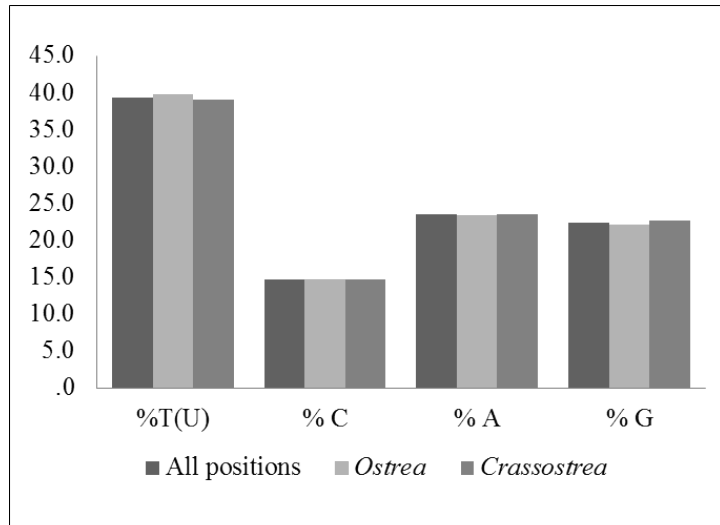
**Appendix 6.4:** Summary statistics for the COIII mitochondrial gene



Length of CYTB sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
1275	39	58	44	14	0.114

Overall Distance	p-distance	SE
All substitutions	0.282	0.007
Transitions	0.144	0.005
Transversions	0.137	0.005
1st position	0.217	0.013
2nd position	0.128	0.010
3rd position	0.500	0.009
<i>Ostrea/Saccostrea</i> within group mean	0.289	0.009
<i>Crassostrea</i> within group mean	0.201	0.007
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.341	0.010

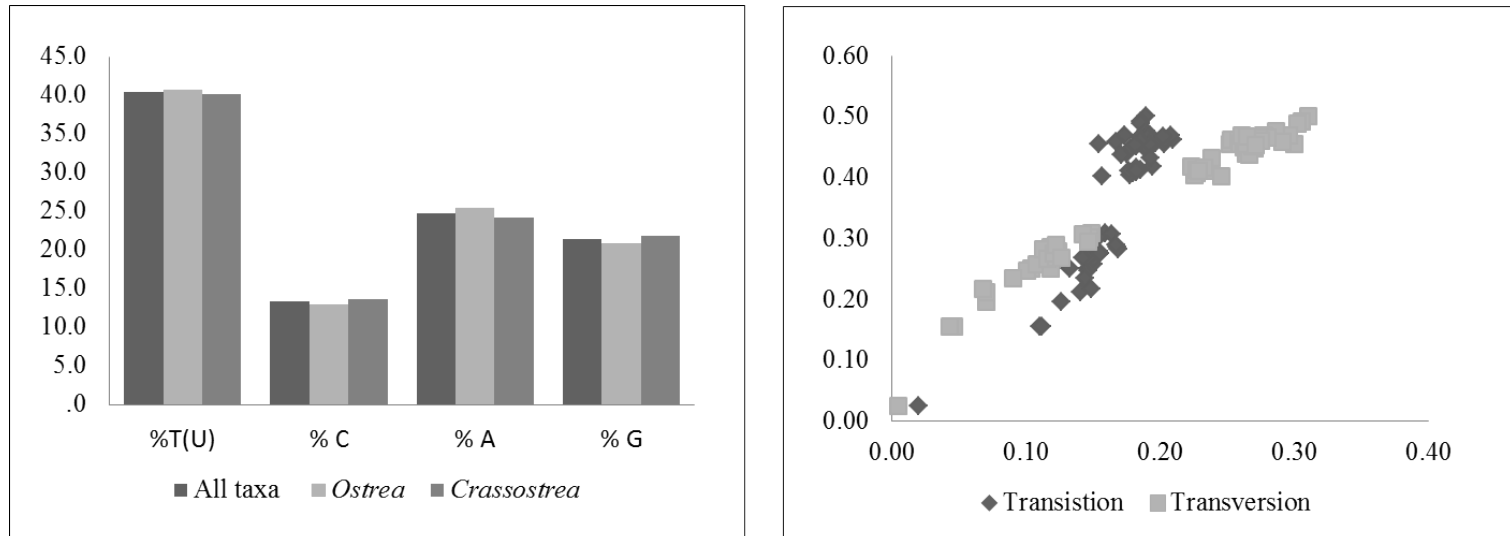
#### Appendix 6.5: Summary statistics for the CYTB mitochondrial gene



Length of NAD1 sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
948	41	58	44	13	0.110

Overall Distance	p-distance	SE
All substitutions	0.270	0.008
Transitions	0.148	0.005
Transversions	0.121	0.006
1st position	0.210	0.014
2nd position	0.100	0.010
3rd position	0.484	0.010
<i>Ostrea/Saccostrea</i> within group mean	0.284	0.009
<i>Crassostrea</i> within group mean	0.208	0.077
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.304	0.009

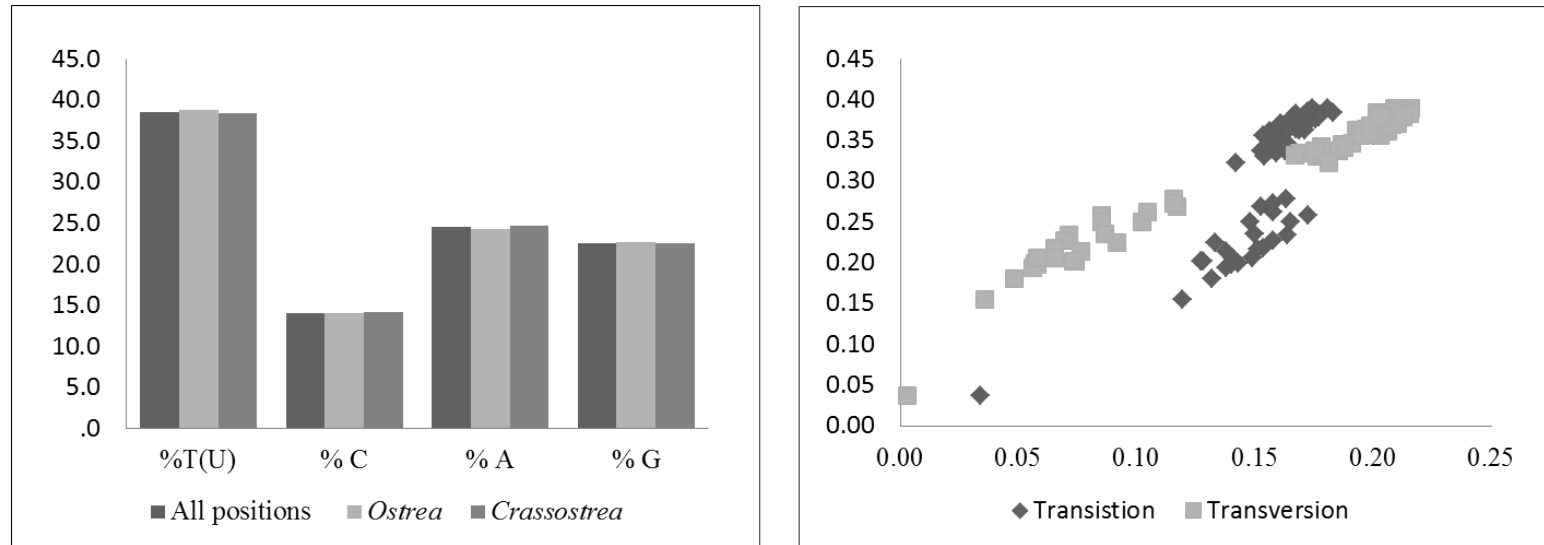
**Appendix 6.6:** Summary statistics for the NAD1 mitochondrial gene



Length of NAD2 sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
1022	25	73	60	13	0.261

Overall Distance	p-distance	SE
All substitutions	0.376	0.008
Transitions	0.170	0.005
Transversions	0.205	0.006
1st position	0.353	0.015
2nd position	0.243	0.013
3rd position	0.530	0.009
<i>Ostrea/Saccostrea</i> within group mean	0.258	0.009
<i>Crassostrea</i> within group mean	0.328	0.007
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.452	0.01

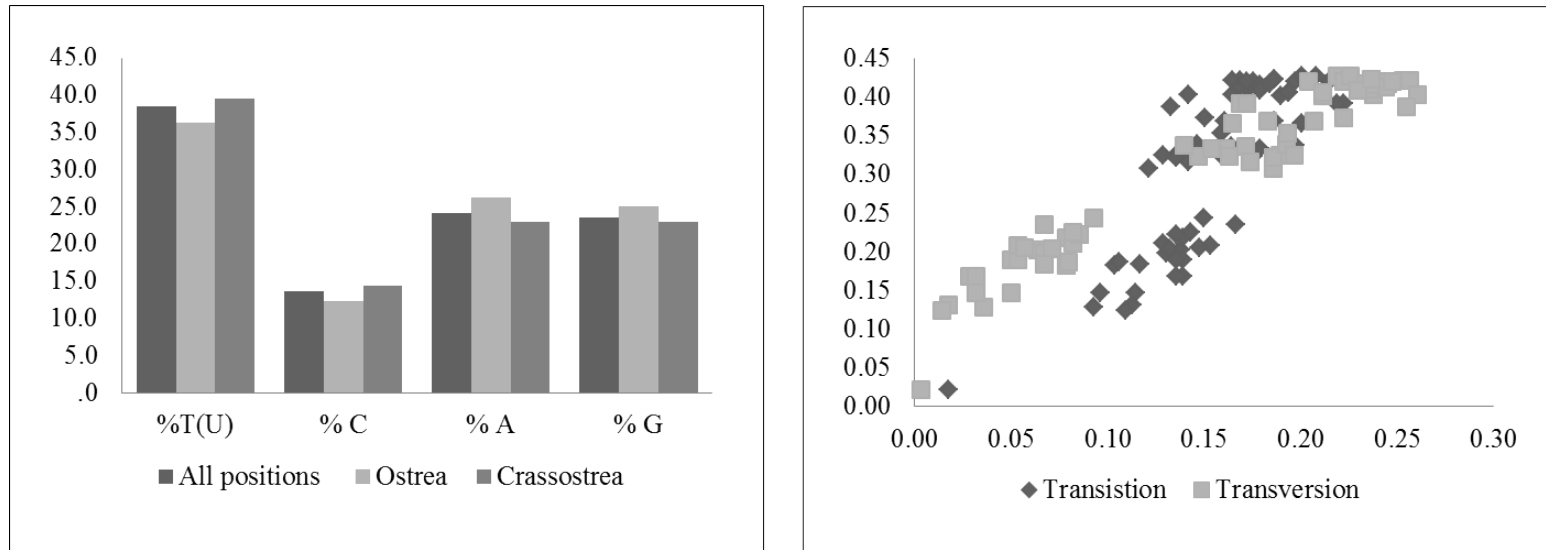
#### Appendix 6.7: Summary statistics for the NAD2 mitochondrial gene



Length of NAD4 sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
1356	32	63	51	11	0.142

Overall Distance	p-distance	SE
All substitutions	0.310	0.006
Transitions	0.157	0.004
Transversions	0.153	0.005
1st position	0.249	0.012
2nd position	0.157	0.010
3rd position	0.526	0.008
<i>Ostrea/Saccostrea</i> within group mean	0.270	0.006
<i>Crassostrea</i> within group mean	0.248	0.007
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.372	0.009

**Appendix 6.8:** Summary statistics for the NAD4 mitochondrial gene



Length of NAD4l sequence	% Conserved Sequences	% Variable Sites	% Parsimony informative site	% Singleton	dN/dS
286	38	61	51	11	0.161

Overall Distance	p-distance	SE
All substitutions	0.307	0.015
Transitions	0.156	0.010
Transversions	0.154	0.010
1st position	0.247	0.025
2nd position	0.129	0.020
3rd position	0.545	0.017
<i>Ostrea/Saccostrea</i> within group mean	0.269	0.018
<i>Crassostrea</i> within group mean	0.214	0.014
<i>Ostrea/Saccostrea</i> vs <i>Crassostrea</i>	0.396	0.021

#### Appendix 6.9: Summary statistics for the ATP6 mitochondrial gene

# Chapter Seven: General Discussion

---

## 7.0 Aims

The overall aim of this thesis was to understand the evolutionary processes that are operating on *Ostrea chilensis* at a molecular and population genetics level. This Chapter assimilates the key findings of the thesis and compares them to other studies. While not the focus of this research I consider how the findings can inform management practices. I make suggestions for further research in the light of the findings of this thesis. Lastly, there have been rapid advances in technology since the onset of this thesis; recommendations for future research are based on the advances made with other important oyster species.

## 7.1 Summary of Key Findings

The genomic data set developed for *O. chilensis* has been annotated with 16,338 novel microsatellite sequences. There were, however, difficulties in designing PCR primers to amplify many of the DNA sequences. Most of the primers for candidate loci would not amplify a single, clear PCR product, those loci that did amplify, could not be properly genotyped without problems such as excessive stutter of the PCR on the alleles. This seems to be a common problem in molluscs (e.g. Reece *et al.* 2004). However, I did (despite these difficulties) develop a novel set of four pairs of microsatellite DNA markers for *O. chilensis*. Further comparisons revealed that some microsatellite loci were potentially orthologous as they were found in a number of other oyster species. Primers designed for these sequences can be used to test for the presence of length polymorphic loci in *O. chilensis*. Of particular interest is the presence of a potentially orthologous loci in *Crassostrea virginica*, on closer inspection of the BLAST hit I could see that the microsatellite loci of both species was well aligned between both species. The more recently derived species *O. chilensis* seems to have evolved substitutions in this sequence. PCR primers designed for *C. virginica* (in the study of Reece *et al.* 2004) are likely to amplify the potentially orthologous sequences in other oyster species, giving the opportunity to study the evolutionary processes at the phylogenetic level of this particular locus. Further analysis of the genomic data set revealed the presence of *BclI* satellite DNA. This is the first time this transposon-like sequence has been identified in *O. chilensis* and has potential to be used as a phylogenetic marker (e.g. López-Flores *et al.* 2004), as well as for studies of transposon associated evolution of microsatellite sequences

(e.g. Zhang, 2004). However, due to the unresolvable problems with the microsatellite markers, which limited the number of useable loci, randomly amplified polymorphic DNA (RAPDs) were used as an alternative genetic marker to assess population genetic structure.

Using RAPD markers I found significant levels of genetic differentiation between *O. chilensis* populations ( $F_{st} = 0.1941$ ,  $p = <0.0001$ ), gene flow between populations was mostly low ( $N_m = < 1$ ), however, populations in the Foveaux strait are more open to gene flow ( $N_m = 16.4$ ). Using a matrix of Nei's genetic distance to conduct a principle component analysis (PCA) showed that many of the populations appeared to be inconsistent with their geographical locations. At most spatial scales I found the genetic data was not explained by an isolation-by-distance profile except when outlier loci were removed. Overall, the genetic structure found in *O. chilensis* seems to be the result of local recruitment and limited to gene flow between sub-populations. The New Zealand wide genetic structure was then tested for a correlation with environmental variables that were obtained from the Marine Environmental Classification system (Snelder *et al.* 2005), using the routine BEST from the software package PRIMER v6.0 (Clarke and Gorley, 2005) revealed that the genetic structure correlates with a combination of two variables: sediment and a metric of sea surface temperature ( $R_s = 0.263$ ), with outlier loci removed the most important environmental variables in explaining genetics structure were the same but the overall correlation was reduced ( $R_s = 0.181$ ). A more complicated relationship was revealed when the data was analysed using a Generalised linear model – as most of the environmental variables were present in the top models. However, this more complicated relationship was attributed to the loss of resolution in the genetic data by using a mean of the population genetic diversity ( $F_{st}$  or  $\Phi_{st}$ ) rather than using the raw allelic frequencies that were used in the BEST analysis.

The *O. chilensis* genomic dataset was searched for the presence of mitochondrial DNA sequences. Using a number of whole genome sequences that are available on GenBank for comparison, I was able to identify 10,086 bp of mitochondrial DNA within the genomic data set. Using these sequences I assembled a partial mitogenome representing around 65% of the full mitogenome and PCR primers were designed to enable the remaining gaps to be closed by a Sanger DNA sequencing approach. The partial mitogenome was used to confirm the phylogenetic position of *O. chilensis* using a partitioned Bayesian reconstruction – this analysis confirmed earlier phylogenetic studies of oysters (Jozefowicz and Ó Foighil, 1998; Ó Foighil, 2000). The cytochrome oxidase I sequence (COI) found in the genomic data set was used to test the suggestion of Ó Foighil, (1999) that the Chilean populations of *O.*

*chilensis* were established by rafting from an ancestral North Island population. With the inclusion of new Chilean sequences from GenBank as well as genetic data from my analysis I produced a phylogenetic tree that supports this suggestion.

Further analysis of the mitochondrial protein coding sequences revealed a codon bias that is probably influenced by transcriptional or translational selection pressures. Translational selection is thought to be a universal occurrence in all genomes of all species, and is found to be most strongly associated with highly expressed genes such as mitochondrial genes.

## 7.2 Context

### *(1) Comparison to other population genetic studies*

Overall, the factors that have shaped the population genetic structure in *O. chilensis* appear to be a combination of historical changes to habitats and the environment, restricted gene flow between populations (perhaps due to a limited pelagic larval duration (PLD), and the influence of environmental variability.

Most of the studies of marine taxa sampled from around the New Zealand coast (reviewed by Ross *et al.* 2009; Gardner *et al.* 2010) show some form of population genetic structure. Studies that found panmixia tended to use less discriminatory markers such as allozymes (e.g. Apte and Gardner, 2001). Studies that previously found panmixia in two species of lobster (*Jasus edwardsii* and *Jasus verreauxi*) (e.g. Smith and McKoy 1980; Ovenden *et al.* 1992; Brasher *et al.* 1992) have since been superseded by a more recent study that found some evidence of population genetic structure in *J. edwardsii* populations using microsatellite loci (Thomas and Bell, 2013). An exception to this trend was the study of Waters and Roy (2003) they unexpectedly did not find not find structure for *Coscinasterias muricata* (which has PLD of around 30 days) based on an analysis of mtDNA markers.

A phylogeographic break at around 42° south is a key finding for many NZ marine taxa (reviewed by Ross *et al.* 2009, Gardner *et al.* 2010). This phylogenetic break was not found in *O. chilensis* populations. The absence of this particular phylogenetic break could be due to a number of reasons: the species has reached equilibrium or there has been human-mediated stock transfer both of which would mask the phylogeographic break.

In the present study two environmental variables were shown to be associated with genetic structure in *O. chilensis*. One of which: a metric of sea surface temperature SST (specifically

spatial gradient of annual mean SST) was also found to be correlated with the genetic structure of *Perna canaliculus* in the study of Wei *et al.* (2013b). Wei *et al.* (2013b) suggested a mode of action for temperature could be selection for tolerance to thermal stress, and the likely candidates could be the family of heat shock proteins. Because a similar association with sea surface temperature has been found in *O. chilensis* there could be similar thermal tolerance selection pressures. Further investigation of variation in genes that relate to thermal tolerance would be an interesting avenue for future research.

Perhaps the main contribution to high levels of genetic structure in *O. chilensis* has been the extremely long brooding period of six to twelve weeks, which allows the larvae to settle almost immediately after their release into the environment (Videla *et al.* 1998). In that study the global multi-locus measure of genetic differentiation was an order of magnitude higher ( $F_{st} = 0.194$ ) compared to the closely related oyster species *O. edulis* ( $F_{st} = 0.019$ ) (Launey, 2002), which has a much shorter pelagic larval duration of 8-10 days.

## **(2) Studies of molecular evolution of microsatellites**

An interesting finding was the many potentially orthologous microsatellite DNA sequences, there are a number of studies that have looked for orthologous sequences with the aim to make inferences about how the genomic region might have evolved. These approaches have been particularly informative when the sequences can be mapped onto a phylogeny (for example Messier *et al.* 1996; Orti *et al.* 1997; Angers and Bernatchez, 1997; Primmer and Ellegren, 1998; Harr *et al.* 2000). Studies such as these suggest that DNA microsatellites begin from a small number of repeats units and in more recently derived individuals in the phylogeny the loci tend to expand into larger sets of repeat units. While the findings reported in this thesis cannot be directly compared to the cited studies, potentially orthologous microsatellite sequences have been identified from many distantly related oysters that can be used to investigate whether this group has similar evolution patterns. Taylor *et al.* (1999) suggested that the ‘death’ of a microsatellite sequence occurs by two types of mutations, the first mutation interrupts the repeat, preventing slipped-strand miss-pairing stabilising the repeat, and the subsequent mutations involve large deletions of the repeat. My findings support the proposed mechanism as Figure 2.8 shows the presence of a substitution in the microsatellite region of the more recently derived *O. chilensis* individual, and this appears to be consistent with the first mutation in the process of ‘death’ of a microsatellite sequence. Although this type of sequence change could only be found in two species it would be

possible to extend this line of enquiry by using the primers designed to amplify that sequence and test for the presence of the repeat sequence in other oyster species – this avenue of research remains to be explored in future studies.

### **7.3 Limitations common to population genetic studies**

The problems presented in this thesis are common to many studies of closely related species and other studies around New Zealand.

#### ***(1) Problems with marker development and concerns of using RAPDs for populations genetics studies***

Microsatellite marker development has been shown to be particularly problematic in molluscs (e.g. McInerney *et al.* 2010). A large amount of time and resources were invested into the development of microsatellite markers, but due to several unresolvable technical problems I eventually had to develop an alternative genetic marker. However, further research into the reasons for difficulties of development marker in these taxa, and the evolutionary mechanisms acting on microsatellite loci have been identified, which are discussed in more detail in section 7.5.

Randomly amplified polymorphic DNA (RAPD) was used as an alternative marker. Various concerns have previously been raised about the usefulness of RAPD markers due to their dominant nature, possible lack of reproducibility, and the consistency of interpretation of the amplified products. However, I addressed these issues using a rigorous data quality checking procedure, with independent and replicate band scoring, as well as the use of software to aid with the consistent interpretation of the amplified products. Despite the difficulties involved with using this marker a number of projects involving New Zealand species have successfully used RAPD markers (e.g. Wood and Gardner, 2007; Jones *et al.* 2008; Reisser *et al.* 2011). However, it should be stated that there is a low level of polymorphism when compared to data sets using more numerous markers (such as SNPs) or more polymorphic markers (such as microsatellites), therefore, the findings using RAPDs should be treated with caution. A potential way to overcome the lack of polymorphism in the data set would be to increase the number of RAPD primers used.

## **(2) *Limited population sampling***

The study focused on eight sampled locations distributed around New Zealand's South Island and one from the Chatham Islands. The study would benefit from more extensive sampling particularly on the North Island of New Zealand. The sampling sites however, represent the current known distribution in New Zealand, because many of the previously sampled populations in the North Island are now thought to be functionally extinct (personal communication with Professor Andrew Jeffs). The Manukau harbour in the North Island of New Zealand has been used as a sampling site for previous studies, however, only two individuals were found in that area, the samples were not used in the study due to the small sample size.

The number of sampling sites was comparable to the number of sampling sites reported in the only other New Zealand wide population genetic study of a marine taxon using RAPDs (Jones *et al.* 2008). That study obtained a total of eight sampled locations distributed across the North and South Island of New Zealand – three of which were from the South Island. Therefore, it has been previously demonstrated that inferences can be made from this number of sampling of sites.

## **(4) *A priori assumptions in the seascape genetics analysis***

The main limitation of the seascapes genetic analysis was that there are no *a priori* expectations that there is an influence of environmental variables on the distribution of genetic variation (e.g. Wei *et al.* 2013a). The environmental variables tested in the analysis may not directly influence the apparent genetic structure and be the key mechanism shaping variation, indeed the structure could be influenced by one or more un-sampled or unknown variables. Also, a large amount of environmental variables were used – this has the problem of creating a very complicated relationship – this is particularly true in the GLZ analysis where most to the environmental variables were used to explain the genetic structure, this has the problem of possibly masking the true environmental signal. However, there is a growing body of research that have used these methods; the strength of the approach is the ability for comparison with a number of other studies (e.g. Wei 2013b, Constable, 2014 and Hannan, 2014).

## 7.4 Recommendations for further study

### ***(1) Further development of microsatellite DNA markers***

Microsatellite DNA markers have presented difficulties in some taxa (Lepidoptera: Meglecz *et al.* 2004; Coleoptera: Arthofer *et al.* 2007; Galatheids: Bailie *et al.* 2010). Molluscs are a group that have been particularly problematic for studies using microsatellite DNA (Reece *et al.* 2004; Weetman *et al.* 2001, 2005). Multi-copy microsatellite families with similar flanking regions are characteristic of taxa that have difficulties associated with microsatellite DNA marker development (Meglecz *et al.* 2004). Zhang (2004) suggested a mechanism for genomic dispersal of new microsatellites was via transposition of mobile elements, followed by accumulation of point mutations in the flanking regions. Therefore, some of the problems I found could be either the flanking regions being too similar between multi-copy loci, or the flanking regions being too different between individuals. Future research projects could seek to determine the copy number of microsatellite loci, with an aim to identifying single copy loci. I suggest to firstly identify microsatellite families (those with highly similar flanking regions), and secondly identify the microsatellite containing sequences that are associated with transposon sequences (this could be achieved using the genomic data developed in Chapter Two). Further PCR primer development should be conducted using sequences that do not have highly similar flanking regions or a known association with transposons. An initial search in Chapter Two identified orthologous transposon sequences, these sequences could be used to test the phylogenetic relationship between taxa (following the work of Lopez-Flores *et al.* 2004). Also, the phylogenetic relationship of the transposons within the *O. chilensis* genome could be tested, which could be used to test the theory of Zhang (2004).

### ***(2) Temporal sampling and Genetic monitoring of oyster populations***

One of the key findings of the thesis is that the population genetic structure would in *O. chilensis* populations are likely to be the result of local recruitment, with limited gene flow between populations. To confirm this finding a temporal sampling strategy should be conducted. If the population structure is under the influence of chaotic processes, then one would expect a different population structure to be found than the one found in this thesis.

Genetic monitoring is the systematic, temporal study of genetic variation, within a particular species or population, with the aim to detect changes or losses that could compromise the

genetic diversity of a species or population (Laikre *et al.* 2008). While it is impossible to determine the extent of historic translocations of *O. chilensis* around New Zealand this thesis showed the current genetic structure is not one of panmixia, it cannot be assumed that the populations are not locally adapted, and such the levels of genetic diversity in local populations should be maintained. The findings of population genetics research are often overlooked in conservation management plans (e.g. Laikre *et al.* 2010). Laikre *et al.* (2008) highlighted four main threats to gene level biodiversity: (1) loss of genetically distinct populations, (2) reduction of genetically effective population sizes caused by increased rates of loss of genetic variation through drift, (3) change of genetic composition and loss of genetic variation through human mediated selection, and (4) inflow of alien genes. Although currently not a threatened species, continued genetic monitoring is important in the understanding of the human impacts on a commercially exploited organism. Human exploitation of natural populations has the potential to change allele frequencies in populations – this could alter the effects of local adaptation and long term evolutionary potential of a population – the effects of these impacts are important to monitor (Allendorf *et al.* 2008). Such impacts on gene level diversity could result from stock enhancement measures (Ryman and Laikre, 1991), therefore enhancement measures need to consider the current genetic structure of a population. Future impacts of climate change on level of genetic variation and the genetic structure of populations could also be monitored in this way, as well as the effects of future outbreaks of bonamiosis. It is important to note that not all genetic changes are anthropogenic in nature, an assessment of the population genetic structure over time could also reveal the effects of chaotic mechanisms (such as Hedgecocks sweepstakes hypothesis) on the stability of genetic structure. The population genetics research reported in this thesis can inform monitoring efforts and it establishes a base line for future comparisons.

### **(3) Incorporate new genomic technologies**

There are many new areas of genomic research that could be investigated. The technologies used for genomics are making rapid advances, the costs continue to come down and quality of genomic sequencing is increasing. There will also be improvements in genotyping technologies and the associated statistical and computational framework as technologies advance. Additional ‘spill over’ benefits as advances will occur in the ‘genomic reference’ species such as *Ostrea edulis* and *Crassostrea gigas*.

The developments made with the flat oyster *O. edulis* (a closely related species to *O. chilensis*), is of particular relevance to this work since it is an important aquaculture species that is similarly affected by a *Bonamia* parasite. Therefore, developments in this species may serve as a useful guideline. Genetic linkage maps have also been developed using 20 microsatellite markers and 246 AFLP markers (Lallias *et al.* 2007). The production of a genetic linkage map is the next step towards identifying quantitative trait loci (QTL), and then the subsequent use of these QTL for selective breeding through marker assisted selection (MAS). QTL have been found that appear to have a higher level of resistance to the parasite *Bonamia ostreae* in *O. edulis* (Lallias *et al.* 2009). These QTL will be valuable for supporting a number of selective breeding programmes, with the aim of developing resistance to bonamiosis in *O. edulis*; some progress in developing resistance has been made in France (Naciri-Graven *et al.* 1998) and in Ireland (Culloty *et al.* 2004).

Candidate genes that are associated with levels of resistance have been identified (Morga *et al.* 2012). Elevated levels of seven expressed sequence tags (EST) showed a significant increase in Bonamiosis resistant oysters, the authors followed up with investigating two of the genes: *OeFas-ligand* and *OelAP* that were completely sequenced, these are involved in apoptosis and are up-regulated in resistant haemocytes. While I could not find homologs of these genes using a BLAST search procedure on the *O. chilensis* genomic data, it is likely there would be a similar pathway of immune response to Bonamiosis.

The most recent development in the aqua-cultural genetics of *O. edulis* is the development of SNP arrays. In the work of (Lapegue *et al.* 2014) they used the large number of expressed sequence tags (EST) available in public data bases to ‘*in silico*’ discover and develop SNPs. These SNP arrays are intended for use in the characterisation of genetic diversity and structure of oyster populations, and to potentially identify discontinuities in the distribution of allele frequencies and signatures of selection. It will also be used to better characterise QTL associated with disease.

#### ***(4) Complete mitochondrial genome and further analyse codon usage***

The work presented in Chapter Five reconstructed a partial mitochondrial genome for *O. chilensis*, primers have been designed to enable the complete sequencing of the genome by Sanger sequencing. Using a full genome sequence is unlikely to alter the presented phylogenetic position of the species, but it will be useful for comparisons of genome

evolution. Unlike other molluscs the gene order in Oysters seems to be conserved (Danic-Tchaleu *et al.* 2011), the full sequence of the *O. chilensis* genome will be useful in order to compare gene order – this is currently impossible since the genome is in fragments that do not contain more than one gene. The complete sequence will also reveal the number and type of tRNAs that are present in the mitochondrial genome – this will allow the testing of the hypothesis that the codon bias found in Chapter Six is related to the type of tRNA present in the mitogenome.

## 7.5 Implications

### *(1) Can the genetic data identify stocks for use in fisheries management/aquaculture?*

Although not the aim of the thesis research, identification of fisheries management units and genetic stock for use in aquaculture breeding programmes could be aided by the results of this study. Typically the goal is to identify potential reproductive independent units - in a genetic analysis of population structure there can be a range of genetic structures from minor divergence in allele frequencies to reciprocal monophyly at mtDNA haplotypes. Moritz (1994) gave a definition that allows the delimitation of populations at these two levels; these reflect the difference between historical processes, and more contemporary gene flow. Reflecting the influence of historical factors on genetic structure is the evolutionary significant unit (ESU), for a population to be defined as an ESU it should be reciprocally monophyletic for mtDNA haplotypes and show a significant level of divergence of allele frequencies at nuclear loci. Reflecting the influence of more contemporary processes are management units (MUs), MUs are populations with significant divergence of allele frequencies at nuclear or mitochondrial loci, regardless of phylogenetic distinctiveness of the alleles. MUs are therefore the logical unit for population monitoring and demographic study, since they are more applicable to short term management than ESUs. Genetic connectivity between populations is also part of the concept of stock. Theoretically, one migrant per generation is thought to be enough to maintain genetic connectivity (Spieth, 1974). However, this concept is based upon many simplifying assumptions that may not hold in the natural world (Mills and Allendorf, 1996). The authors stated circumstances where more than one migrant per generation could be necessary. Of relevance to a fisheries population is the fact that an increase above one migrant per generation could be desirable in populations where inbreeding is thought to be a problem. It is also important to seek a balance between the

effects of local adaptation and outbreeding depression – with that in mind the authors' state that ten migrants per generation is unlikely to disrupt that balance.

All of the *O. chilensis* populations except the two Foveaux Strait populations showed significant divergence between populations. Therefore they should all be considered separate management units according to the MUs definition, with the exception of the two Foveaux strait populations that should be considered as one management unit under this definition. There are very low levels of genetic connectivity between populations: migration rates between populations are all well below 10 migrants per generation. With the exception of the Foveaux Strait populations that have higher levels of connectivity, which further supports the suggestion that they should be considered separate management units. Therefore with the low levels of connectivity found in this study, it is important to monitor the population for signs of inbreeding depression.

## ***(2) Integration of population genetics and Seascapes genetic knowledge into the design of Marine Protected Areas***

Recently some Marine Protected Areas (MPAs) are being designated with the aim to provide ecologically coherent networks (Johnson *et al.* 2008). An ecologically coherent network is one where each individual protected area supports the other, interacting with the habitat of the surrounding area. This study aimed to elucidate the processes that were generating population genetic structure of *O. chilensis*, and then test the influence of environmental variables on the apparent genetic structure - this can inform the design of ecological coherent MPAs. Marine organisms have wide variation in dispersal potential due to different life history strategies hence there may not be a suitable MPA strategy that is ideal for all species. In this sense *O. chilensis* lends itself as a model organism with limited dispersal due to its much reduced PLD so findings here can reflect the needs of other organisms with reduced gene flow.

Overall, this study has revealed three main processes that have influenced the genetic structure in *O. chilensis*. These patterns and processes should be considered in the design of Marine Protected Areas (MPAs). As previously discussed, genetic structure of *O. chilensis* has been revealed to be influenced by two main factors: (1) Life history traits, *Ostrea chilensis* is a species with short PLD, which implies it has limited dispersal capabilities. This is supported by the results of Chapter Three that showed there is reduced connectivity between , suggesting that populations are essentially self-recruiting; (2) Correlation with

environmental variables, the seascape analysis highlighted the importance of a sediment variable and a metric of sea surface temperature on the genetic structure.

The similar findings reported in this study with other studies of New Zealand marine taxa can be integrated to make management decisions. The Seascapes genetics analysis from the studies of Constable (2014), Hannan (2014) and Wei *et al.* (2013b) have all pointed towards the influence of metrics associated with sea surface temperature – therefore in order for MPAs to fulfil the objective of an ecologically coherent network they should aim to capture the full scale of sea surface temperatures across New Zealand waters.

## 7.6 Summary

1. This study has made a number of new contributions to science. Previously unknown information concerning the genetic stock of *O. chilensis* in New Zealand has been identified. This information can be used to inform fisheries and aquaculture management. This information also fits into a body of research that allows for the comparison of patterns of genetic structure for a range of marine taxa around New Zealand.
2. The study provides novel genetic tools to study evolution at the molecular level. Some progress has been made to describe these patterns and processes, but many more questions have been raised. I have made recommendations for future studies based on the findings of this thesis, and the current progress being made in other commercially important oyster species.
3. The work presented in this thesis focused on a commercially important and iconic species. The information generated here is of value to the future economic growth of the species in New Zealand. The tools and knowledge generated during this thesis research will support future economic growth of the species in terms of selective breeding and a greater understanding of fishery stock.

## 7.7 References

- Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N. Genetic effects of harvest on wild animal populations. *Trends in Ecology and Evolution*. 2008; 23:327–337.
- Angers B, Bernatchez L (1997). Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution* 14, 230–238
- Apte S, Gardner JPA. (2001). Absence of population genetic differentiation in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791) as assessed by allozyme variation. *Journal of experimental marine biology and ecology*, 258(2), 173-194.
- Arthofer W, Schlick-Steiner BC, Steiner FM, Avtzis DN, Crozier RH, Stauffer C. (2007). Lessons from a beetle and an ant: coping with taxon-dependent differences in microsatellite development success. *Journal of molecular evolution*, 65(3), 304-307.
- Bailie DA, Fletcher H, Prodöhl PA. (2010). High incidence of cryptic repeated elements in microsatellite flanking regions of galatheid genomes and its practical implications for molecular marker development. *Journal of Crustacean Biology*, 30(4), 664-672.
- 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(1), 53-58.
- Constable HB. (2014). Population structure, temporal stability and seascape genetics of two endemic New Zealand Pleuronectiformes, *Rhombosolea plebeia* (sand flounder) and *R. leporina* (yellowbelly flounder). Ph. D Thesis. Victoria University of Wellington. NZ
- Culloty SC, Cronin MA, Mulcahy MF. (2004). Potential resistance of a number of populations of the oyster *Ostrea edulis* to the parasite *Bonamia ostreae*. *Aquaculture*, 237(1), 41-58.
- Danic-Tchaleu G, Heurtebise S, Morga B, Lapègue, S. (2011). Complete mitochondrial DNA sequence of the European flat oyster *Ostrea edulis* confirms Ostreidae classification. *BMC research notes*, 4(1), 400.

Foighil DÓ, Marshall BA, Hilbish TJ, Pino MA. (1999). Trans-Pacific range extension by rafting is inferred for the flat oyster *Ostrea chilensis*. *Biological Bulletin*, 122-126.

Gardner JPA, Bell JJ, Constable HB, Hannan D, Ritchie PA, Zuccarello GC. (2010). Multi-species coastal marine connectivity: a literature review with recommendations for further research: Victoria University of Wellington report (ZBD2009-10) to the New Zealand Ministry of Fisheries. Wellington: New Zealand Ministry of Fisheries. p.50.

Hannan, D. A. (2014). Population genetics and connectivity in *Paphies subtriangulata* and *Paphies australis* (Bivalvia: Mesodesmatidae). Ph. D Thesis. Victoria University of Wellington. NZ

Harr B, Zangerl B, Schlotterer C (2000). Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. *Molecular Biology and Evolution* **17**, 1001–1009

Johnson MP, Crowe TP, McAllen R, Allcock AL. (2008). Characterizing the marine Natura 2000 network for the Atlantic region. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 18(1), 86-97.

Jones TC, Gemmill CE, Pilditch CA. (2008). Genetic variability of New Zealand seagrass *Zostera muelleri* assessed at multiple spatial scales. *Aquatic Botany*, 88(1), 39-46.

Laikre L, Larsson LC, Palmé A, Charlier J, Josefsson M, Ryman N. (2008). Potentials for monitoring gene level biodiversity: using Sweden as an example. *Biodiversity and Conservation*, 17(4), 893-910.

Laikre L. (2010). Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics*, 11(2), 349-354.

Lallias D, Beaumont AR, Haley CS, Boudry P, Heurtebise S, Lapegue S. (2007). A first-generation genetic linkage map of the European flat oyster *Ostrea edulis* (L.) based on AFLP and microsatellite markers. *Animal genetics*, 38(6), 560-568.

Lapegue S, Harrang E, Heurtebise S, Flahauw E, Donnadiou C, Gayral P, Klopp C. (2014). Development of SNP-genotyping arrays in two shellfish species. *Molecular ecology resources*. 14: 820–830.

- Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y. (2002). Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism. *Journal of Heredity*, 93(5), 331-351.
- López-Flores I, de la Herrán R, Garrido-Ramos MA, Boudry P, Ruiz-Rejón C, Ruiz-Rejón, M. (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene*, 339, 181-188.
- McInerney CE, Allcock AL, Johnson MP, Bailie DA, Prodöhl PA. (2010). Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in molluscs. *Heredity*, 106(1), 78-87.
- Meglec E, Petenian F, Danchin E, D'Acier AC, Rasplus JY, Faure E. (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology*, 13(6), 1693-1700.
- Messier W, Li SH, Stewart CB (1996). The birth of microsatellites. *Nature* 381, 483
- Mills LS, Allendorf FW. (1996). The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 10(6), 1509-1518.
- Morga B, Renault T, Faury N, Arzul I. (2012). New insights in flat oyster *Ostrea edulis* resistance against the parasite *Bonamia ostreae*. *Fish & shellfish immunology*, 32(6), 958-968.
- Moritz C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, 9(10), 373-375.
- Naciri-Graven Y, Martin AG, Baud JP, Renault T, Gerard A. (1998). Selecting the flat oyster *Ostrea edulis* (L.) for survival when infected with the parasite *Bonamia ostreae*. *Journal of Experimental Marine Biology and Ecology*, 224(1), 91-107.
- Orti G, Pearse DE, Avise JC (1997). Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Science USA* **94**, 10745–10749
- 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(2), 319-326.

Primmer CR, Ellegren H (1998). Patterns of molecular evolution in avian microsatellites.

*Molecular Biology and Evolution* **15**, 997–1008

Reece KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity*, 95(4), 346-352.

Reece KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity*, 95(4), 346-352.

Reisser CM, Wood AR, Bell JJ, Gardner JPA. (2011). Connectivity, small islands and large distances: the *Cellana strigilis* limpet complex in the Southern Ocean. *Molecular Ecology*, 20(16), 3399-3413.

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(5), 1009–1027.

Ryman N, Laikre L. (1991). Effects of supportive breeding on the genetically effective population size. *Conservation Biology*, 5(3), 325-329.

Smith PJ, McKoy JL, Machin PJ. (1980). Genetic variation in the rock lobsters *Jasus edwardsii* and *Jasus novaehollandiae*. *New Zealand journal of marine and freshwater research*, 14(1), 55-63.

Spieth PT. (1974). Gene flow and genetic differentiation. *Genetics*, 78(3), 961-965.

Taylor JS, Durkin JM, Breden F. The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Molecular Biology and Evolution*. 16, 567–572 (1999).

Thomas L, Bell JJ. (2013). Testing the consistency of connectivity patterns for a widely dispersing marine species. *Heredity*, 111(4), 345-354.

Videla JA, Chaparro OR, Thompson RJ, Concha II. (1998). Role of biochemical energy reserves in the metamorphosis and early juvenile development of the oyster *Ostrea chilensis*. *Marine Biology*, 132(4), 635-640.

Wang J. (2004). Application of the one-migrant-per-generation rule to conservation and management. *Conservation Biology*, 18(2), 332-343.

Waters JM, Roy MS. (2003). Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography*, 30(12), 1787-1796.

Weetman D, Hauser L, Carvalho GR. (2001). Isolation and characterization of di- and trinucleotide microsatellites in the freshwater snail *Potamopyrgus antipodarum*. *Molecular Ecology Notes*, 1(3), 185-187.

Weetman D, Hauser L, Shaw PW, Bayes MK. (2005). Microsatellite markers for the whelk *Buccinum undatum*. *Molecular Ecology Notes*, 5(2), 361-362.

Wei K, Wood AR, Gardner JPA (2013a). Population genetic variation in the New Zealand greenshell mussel: locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. *Marine biology*, 160(4), 931-949.

Wei K, Wood AR, Gardner JPA. (2013b). Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. *Marine Ecology Progress Series*, 477, 107-121.

Wood AR, Gardner JPA (2007). Small spatial scale population genetic structure in two limpet species endemic to the Kermadec Islands, New Zealand. *Marine Ecology Progress Series*, 349, 159-170.

Zhang DX. (2004). Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology and Evolution*, 19(10), 507-509.