Population genetic structure of New Zealand blue cod (*Parapercis colias*) based on mitochondrial and microsatellite DNA markers

Clare Louise Gebbie

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Master of Science

Victoria University of Wellington

2014

Abstract

Parapercis colias (blue cod) is an endemic temperate reef fish that supports an important commercial and recreational fishery in New Zealand. However, concerns have been raised about localized stock depletion, and multiple lines of evidence have suggested *P. colias* may form several biologically distinct populations within the New Zealand Exclusive Economic Zone. Mark and recapture studies along with otolith and stable isotope studies have indicated that individuals are sedentary with very limited movement beyond the scale of 10-20km. The primary goal of this research was to advance the current knowledge of *P. colias* population genetic structure. This information can be incorporated into stock assessment models with the aim of improving the management of the *P. colias* fishery. This study made use of 454 pyrosequencing technology to isolate and develop the first set of microsatellite DNA markers for *P. colias*. These seven microsatellite loci, along with mitochondrial control region sequences, were used to determine the levels of genetic variation and differentiation between sites around the New Zealand coastline, including the Chatham Islands.

Significant differentiation was observed between the Chatham Islands and mainland New Zealand sample sites, indicating that these two regions form distinct populations. Interpretation of the results for the mainland sites was more complex. Mitochondrial sequence data detected no significant pairwise differentiation between mainland sites, although a pattern of isolation-by-distance was observed. However, evidence for genetic differentiation among mainland sites was weak based on the microsatellite DNA analysis. Although pairwise G_{ST} levels were significant in some sites, this was not reflected in principal component analysis or Bayesian structure analysis. It is likely that through long range dispersal, migration is at or above the threshold for genetic connectivity, but below a level necessary for demographic connectivity. This is indicated by both the genetic structure reported here, along with previous studies showing limited dispersal of *P. colias*.

Acknowledgements

First and foremost, I would like to thank my supervisor Peter Ritchie for extending me the opportunity to take on this project. I would also like to thank Natalie, for her support and friendship. You made the lab work and uni-life much more enjoyable. The rest of my lab group also needs a mention- Balam, Rachael and Ange; you guys have been awesome support. A huge thank you needs to be extended to Dan McGibbon from NIWA who organised the collection of West Coast blue cod samples. You made the collection of samples far simpler than I thought it would be! Emily and Glenn, thank you for all the proof-reading you have both done over the past few weeks. Much appreciated.

Last but not least, I need to thank my wider group of friends, flatmates and family for tolerating my absence these past few months. I know I've missed a few important events, but I am now available to make up for it all! Thanks for being so understanding and supportive.

Table of Contents

Population genetic structure of New Zealand blue cod (*Parapercis colias*) based on microsatellite DNA markers

Abstract	1
Acknowledgements	2
List of Tables	5
List of Figures	6
Chapter One	
General Introduction	
New Zealand fisheries and the stock concept	7
Genetic variation and population structure	9
The use of molecular markers	10
Study species: Blue cod (Parapercis colias)	12
Thesis aims	15
Chapter Two	
The phylogeographic structure of New Zealand blue cod (<i>Parapa</i> mitochondrial DNA control region sequences	ercis colias) based on
2.1 Introduction	17
2.2 Methods	20
Sample collection and DNA sequencing	20
Genetic analyses	20
Demographic history	22
2.3 Results	24
Genetic diversity	24
Demographic history	
2.4 Discussion	34
Population structure	34
Demographic history	36
Chapter Three	
Development of novel microsatellite DNA markers for blue cod	(Parapercis colias)
3.1 Introduction	
3.2 Methods	

3.3 Results	45
3.4 Discussion	47
Chapter Four	
Population genetic structure of blue cod (<i>Parapercis colias</i>) based on seven microsatellite markers	
4.1 Introduction	49
4.2 Methods	52
Sample collection	52
Genotyping	53
Genetic diversity	54
Population structure	55
Mutation-drift Equilibrium and Bottleneck	56
4.3 Results	58
Genetic diversity	58
Population structure	64
Mutation-drift Equilibrium and Bottleneck	69
4.4 Discussion	70
Genetic diversity	70
Population structure	71
Population structure – mainland sites	71
Isolation-by-distance	72
Mutation-drift equilibrium and Bottleneck	73
Management implications	73
Chapter Five	
General Discussion and Implications for Fisheries Management	
General	75
Management overview	75
Management suggestions	76
Future work	78
Appendix	80
References	84

List of Tables

Chapter Two

2.1	Summary statistics for the different sites at which <i>P. colias</i> were sampled	23
2.2	Pairwise Φ_{ST} values for all sample sites	27
2.3	Results from Analysis of Molecular Variance (AMOVA) for mtDNA analysis	29
2.4	Mismatch distribution and neutrality test results	29
Chapter '	Three	
3.1	Characteristics of seven microsatellite loci	44
Chapter 1	Four	
4.1	Primer sequence and PCR condition of seven microsatellite loci	52
4.2	Summary statistics for seven microsatellite loci	57
4.3	Pairwise $F_{ST}(\theta)$ values	63
4.4	Pairwise G _{ST} and standardised G' _{ST} values	63
4.5	Pairwise R _{ST} and Jost's D _{est} values	63
4.6	Results from Analysis of Molecular Variance (AMOVA) for seven microsatellite loci	64
4.7	Results from the Wilcoxon bottleneck test and M-ratio analysis	67

List of Figures

Chapter One

1.1	Blue cod quota management areas defined by the New Zealand quota	13
	management system	

Chapter Two

2.1	Sample sites and sizes with haplotype frequency chats and minimum	24
	spanning network	
2.2	Rarefaction curves for each sample site	26
2.3	Isolation-by-distance between mainland sample sites	26
2.4	Mismatch distribution plots	30
2.5	Bayesian skyline plots for the Chatham Islands and mainland	31
	populations	

Chapter Four

4.1	Sample locations and sizes for seven sample sites	50
4.2	Rarefaction curve for the locus BCO_03	57
4.3	Allele frequency distribution plots for the BCO_12 and BCO_15 loci	58
4.4	Allele frequency distribution plots for the BCO_08 and BCO_17 loci	59
4.5	Allele frequency distribution plots for the BCO_13 and BCO_11 loci	60
4.6	Allele frequency distribution plots for the BCO_03 locus	61
4.7	Principal component analysis	65
4.8	Bayesian structure analysis	65
4.9	Factorial component analysis	66
Appendix 1	Sequence length distribution of 454 pyrosequencing reads	78
Appendix 2	Delta K values for Bayesian structure analysis	78
Appendix 3	Bayesian structure plots for K=2-7	79
Appendix 4	Rare allele frequency distributions	80
Appendix 5	Rarefaction curves for each microsatellite locus	83

Chapter One

General Introduction

New Zealand fisheries and the stock concept

The fishing industry is unique, because it harvests a major human food item directly from wild populations. Over the last few decades, over-exploitation of marine fish has become more common, and this has led to a global trend of declining commercial fish stocks. The sustainable management of fish populations is essential if productivity is to be maintained, and currently depleted stocks are to be recovered as the human population continues to increase in size. New Zealand has the fourth largest Exclusive Economic Zone (EEZ) in the world, covering 4.4 million km² (Ministry for Primary Industries), which generates 600,000 tonnes of seafood annually. Most of this seafood is exported, contributing significantly to the nation's economy. In 2013, for example, over NZ\$1.5 billion of fish was exported (www.stats.govt.nz). The sustainability of New Zealand's fisheries is thus important from the economic perspective, but also from the cultural point of view, with significant recreational and customary fishing practices around the country.

In response to the concern about decreasing catch rates, the New Zealand Quota Management System (QMS) was introduced in 1986. New Zealand has placed nearly 100 species under the QMS, which separates the EEZ into independently managed areas referred to as Quota Management Areas (QMA). For each species, a total allowable catch (TAC) is set annually for each QMA, on the basis of stock assessment data, with a proportion allocated to the total allowable commercial catch (TACC). The TACC is measured in tonnes, and is divided into Individual Transferable Quotas (ITQs), which can be traded by commercial fishers in exchange for the rights to catch a predefined volume of fish within a particular QMA.

The main purpose of the QMS is to define management areas and catch levels based on stock assessments, in order to maximise the sustainability and increase the economic efficiency of the fishing industry. However, an accurate stock assessment can be difficult to achieve, as incorporating detailed information on the biology and population dynamics of fishery species can be expensive and technically challenging (Waples et al., 2008). Because of this, a fishery stock is often defined arbitrarily in terms of various geographical locations or historical

fishing areas, while incorporating little or no information concerning the biological population (Ward, 2000). The term 'stock' itself can have different meanings, depending on the context (Carvalho and Hauser, 1994). Under the New Zealand Fisheries Act 1996, a stock is simply defined "in reference to units for the purpose of fisheries management" (Ministry of Primary Industries, 2014). Under this definition, a fish stock can be made up of multiple biologically distinct populations (or even different species), or it may incorporate a portion of one panmictic population. This can have important implications for fisheries management, because when genetically distinct populations are managed as a single stock, there is a significant risk of unequal harvesting pressure, which can lead to local depletion, loss of genetic diversity and selective genetic changes (Allendorf et al., 2008).

In order to limit overharvesting and aid the recovery of depleted stocks, a fisheries stock should be defined by the identification of demographically and genetically independent populations. This can help fisheries managers achieve a 'maximum sustainable yield' for a fishery, provided harvest rates do not exceed a level that can be sustained by recruitment within natural populations, whilst at the same time maximising stock productivity (Carvalho and Hauser, 1994). Stock assessment models are therefore crucial for the successful management of a fishery species.

Methods for attempting to define stocks based on biological information can be categorised as direct or indirect approaches. Direct methods include fishery surveys and mark and recapture studies, which can provide useful information such as spatial distribution, life history characteristics and dispersal patterns. However, data can often be difficult to obtain, and the interpretation of its biological relevance may be challenging. For example, mark and recapture studies can estimate the movements of a certain population of fish, but do not provide information on the reproductive behaviour of migrants, which could be biologically relevant, since natal homing is common in some fish species. Indirect genetic studies can provide information on genetic variation within populations, as well as population structure, which can then be used to define management boundaries. It has been proposed that a holistic approach, combining both direct and indirect methods, should be used for the estimation of a stock unit, and to achieve the maximum sustainable yield (Begg and Waldman, 1999).

Genetic variation and population structure

For the sustainable management of fisheries, stock assessment models should include available knowledge of the population's genetic structure. Populations evolve over time, and when gene flow between populations is limited, the forces of genetic drift and selection can lead to differentiation between populations, whereas migration can act to homogenise populations. Three basic types of population structure are described by Laikre (2005). The first is 'panmixia', which describes genetically homogenous populations with no genetic differentiation. The second category of population structure is called 'isolation-by-distance', in which the genetic composition of a population changes continuously over an area, with the genetic distance between groups of individuals increasing with geographic distance. This type of structure can be common, since individual dispersal capabilities are often limited relative to the species' geographical distribution. Thirdly, 'distinct populations' occur where gene flow is limited to the point where significant genetic differentiation is observed between populations. When populations are shown to be genetically differentiated, these separate stocks can form the basis for a management unit.

Due to their pelagic larval duration (PLD), adult mobility, and a lack of oceanic dispersal barriers, it has often been suggested that marine fish are likely to exhibit limited population structure, as a result of high levels of gene flow. This is reflected in the typically low levels of genetic differentiation found in marine fish, when compared to that of fresh water and terrestrial species (Ward and Elliott, 2001). However, teleost species can exhibit low demographic and genetic connectivity for a variety of reasons, including environmental and oceanographic conditions, biological traits such as territorial behaviour, natal homing or short PLD. For example, Henriques (2014) identified two highly divergent populations in the coastal fish species *Atractoscion aequidens*, and attributed this to a disruption of gene flow resulting from an upwelling cell in the Bengula Current in the south-eastern Atlantic.

Genetic markers can be used to address two fundamental questions concerning population connectivity. First, they can be used to estimate genetic variations within a population, and secondly, they can be used to estimate genetic differentiation between populations. Maintaining genetic variation within populations is important for the long-term persistence and adaptability of a species. Subpopulations can act as reservoirs of unique alleles and act as evolutionary significant units (ESU). When gene flow between populations is restricted, genetic differentiation can occur. However, population genetic theory suggests that even a small degree of migration between populations can be sufficient to maintain genetic homogeneity (Ovenden, 2013). This is because genetic connectivity between populations is influenced by the absolute number of migrants and, theoretically, only a few migrants per generation are required to maintain connectivity. In contrast, demographic connectivity between populations is a function of the proportion of migrants moving from one population to another (Ovenden, 2013). Therefore, if stocks are genetically undifferentiated, it can be difficult to draw conclusions on demographic connectivity, since the migration rate may not necessarily be high, or the stocks may in fact be decoupled, with insufficient time having passed for this to be reflected in their genetic structure (Lowe and Allendorf, 2010).

The use of molecular markers

There are several types of molecular markers available for the study of patterns of genetic differentiation. A variety of genetic markers can be applied to studies of population structure, such as allozymes, mitochondrial DNA (mtDNA), microsatellite DNA and single nucleotide polymorphisms (SNPs). Each molecular marker differs in its mode of inheritance and rate of evolution, providing insight into diverse aspects of population connectivity (Hoffman et al., 2009). A successful marker is one that will provide a level of variation that is sufficiently high to allow accurate investigation of differentiation, but not so high that all individuals are unique, as this would reduce the statistical power with which relatedness could be inferred (Hellberg, 2009). Population genetic studies have focused on selectively neutral markers, as they more accurately reflect patterns of genetic drift and migration, and are less likely to be confounded by selection processes. The most commonly used markers in contemporary studies are mitochondrial and microsatellite DNA, due to their higher resolving power when compared to allozymes, and their ease of use when compared to SNPs.

Several characteristics make mtDNA a valuable tool for detecting genetic variation and population structure. MtDNA markers are useful for obtaining a first look at population structure, as they are straightforward to develop, and the genetic data is straightforward to collect and analyse. As different regions of the mitochondrial genome evolve at different rates, they have become popular for inferring both genetic relationships between populations, as well as the phylogenies of closely related species. Using coalescence theory, the relationship between individual haplotypes can be used to infer the demographic history of populations. The maternal inheritance of mtDNA reduces the effective population size by a quarter, relative to the nuclear genome, which acts by increasing the sensitivity of mtDNA to

demographic change through genetic drift (Castro et al., 1998). Because mtDNA is haploid and does not undergo recombination, all regions of the genome are linked and therefore represent a single locus.

The mtDNA genome is comprised of highly variable regions as well as conserved gene sequences. Because of this, a wide range of PCR primers are available, and previously developed primers can often be applied to the study of new species. The control region (CR) is frequently used for population-level studies, as it is the major non-coding section of the genome and often more variable than other areas (Lee et al., 1995). The rate of evolution of the CR varies with a central domain region that remains relatively well conserved, flanked by hypervariable regions on either side. This elevated mutation rate has led to the control region often being used for the study of genetic diversity within a range of species (Rocha-Olivares et al., 2005). Other popular regions of the mtDNA include the cytochrome b gene and COX1; however this is more suitable for inter-specific applications. Advances in DNA sequencing technologies have now made it easier to study the entire mitochondrial genome of a species, since the genomic arrangement can differ between species (Miya et al., 2003).

Mitochondrial DNA is maternally inherited and it represents the evolutionary history of the female line only. Interpreting population structure can therefore be misleading if populations are heavily sex biased. In light of this, it is important to also use multiple unlinked loci such as microsatellite DNA genotyping, which represents the population history of both sexes, to provide a more complete picture of the population structure. Microsatellite DNA, also termed simple sequence repeats (SSR) or short tandem repeats (STR), are tandemly repeated DNA sequences that are abundantly distributed throughout the nuclear genome. They consist of 1-6 base pair repeat motifs, which mutate through the gain or loss of these motifs resulting from polymerase slippage during DNA replication. Microsatellites are co-dominant markers that can infer population structure, through alleles derived from the number of repeat units, and these can be determined through standard PCR methods. They are characterised by a rapid mutation rate which is estimated to be several orders of magnitude greater than non-repetitive nuclear DNA, thus making them suitable for the detection of recent structure at the population or subpopulation level (Chistiakov et al., 2006).

The mechanisms of microsatellite evolution are not fully understood, but it is thought that they mutate predominantly by polymerase slippage during DNA replication (Chistiakov et al., 2006). Slipped strand mispairing can occur during replication when the polymerase enzyme dissociates briefly with the template strand, with the subsequent reattachment causing the addition or subtraction of one or more repeat units. Additionally, microsatellites are thought to mutate via unequal crossing over during recombination (Chistiakov et al., 2006). This occurs during meiosis when sister chromosomes fail to pair up accurately, resulting in longer sequences being gained or lost, which explains larger changes in repeat units. These different methods of microsatellite evolution have led to the development of multiple mutation models. The step-wise mutation model (SMM) describes the equal probability of a single repeat unit being gained or lost (Kimura and Ohta, 1978). This allows the relationship between alleles to be incorporated into differentiation measurements. At the other end of the spectrum, the infinite alleles model (IAM) states that each mutation can result in any number of repeat increases (or decreases), leading to a novel allele in the population (Kimura and Crow, 1964). However, several studies have shown that the most realistic models are the two-phase model (TPM) and the generalized stepwise model, which are variations combining both SMM and IAM (Estoup et al., 2002).

Recent technological advances in DNA sequencing technologies have improved the efficiency of microsatellite isolation for primer development, and as a result microsatellite loci can now be applied more readily and at a lower cost (Guichoux et al., 2011). Next generation sequencing techniques, GS-FLX Titanium (Roche, 454 Life Sciences) in particular, have swiftly replaced traditional enrichment based methods since their first application for microsatellite development in 2009 (Abdelkrim et al., 2009; Gardner et al., 2011). The use of both mitochondrial and microsatellite DNA markers can provide a more comprehensive investigation into population structure, as these two methods investigate population history over different timescales. As mtDNA is a 'sequence' marker, it can be used to look at demographic history, although it is representative of a single locus and describes the history of only one sex. In contrast, microsatellite DNA 'frequency' markers are rapidly evolving and highly polymorphic. Although this can result in a higher resolving power, it implies that these markers are less useful for the study of long-term population structure (Hellberg, 2009).

Study species: blue cod (Parapercis colias)

Parapercis colias (commonly blue cod or rawaru) is a temperate reef fish and member of the Penguipedidae family (Forster, 1801). It is an endemic species that is widely distributed in rocky coastal waters throughout New Zealand, and is found in greater abundance south of

Cook Strait. Blue cod is fished both commercially and recreationally, and is popular for its medium texture and mild flavour. It is one of the most frequently caught recreational finfish in the coastal waters around the South Island and Chatham Islands. Major commercial fisheries also exist off Southland and the Chatham Islands, with smaller but significant fisheries in other areas around the South Island (Ministry for Primary Industries, 2014). Commercially fished blue cod are caught mainly in pots, with small landings reported as a bycatch from inshore trawl fisheries. Recreationally, they are fished using hook and line, and are occasionally landed by spearfishing.

Parapercis colias live down to a depth of 150 m, and are benthic opportunistic carnivores, feeding mainly on small fish, crabs and shellfish. Although little is known about their early life history (Cole et al., 2001; Rodgers and Wing, 2008), it has been estimated that eggs are pelagic for five days after spawning, before spending a further five days as pelagic larvae (Henderson, 2009). Blue cod are haremic and protogynous hermaphrodites, with most fish undergoing sex inversion from females to males, although some fish do begin life as males and never undergo sex inversion. Age and size at sexual maturity varies significantly from one location to another, with northern fish reaching maturity earlier and at smaller sizes than fish further south. For example, male fish reach sexual maturity at 10-19 cm (2-3 years of age) in Northland, in contrast to Southland fish that mature at 26-28 cm (4-6 years of age) (Blackwell, 2006). Blue cod are sedentary and the males are territorial, with observations suggesting that a typical male may defend a small group of 3-5 females (Carbines and McKenzie, 2001). Tagging studies have shown limited movements of these fish, with 60% traveling distances of less than 1 km. However, the maximum distance travelled by one individual was 156 km over a 20 month study period, suggesting that long range dispersal can occur (Carbines and McKenzie, 2001; Carbines and McKenzie, 2004).

To manage blue cod, eight separate fisheries or Fisheries Management Areas (FMA) have been established (Figure 1.1). The TACC has been set at 2,680 tonnes since 2001, and reported catches have been relatively stable since that date, varying from 2,161 to 2,501 tonnes between 2002 and 2010 (Ministry for Primary Industries, 2012). However, there has been some concern over the sustainability of individual fisheries areas, particularly around the Marlborough Sounds, which is now treated as a single management unit (Ministry for Primary Industries, 2012). The recreational fishery in this area was first closed in 2008, following a NIWA survey highlighting the substantial decline in blue cod. The Marlborough Sounds area is now subjected to annual closure, from September 1st to December 19th, thus coinciding with the blue cod spawning season. Catch limitations for the MSA have varied over the years, and are currently set at a limit of 2 per person at 30-35 cm.



Figure 1.1 Blue cod quota management areas defined by the New Zealand quota management system. Adapted from the Ministry for Primary Industries by Smith 2012.

Stable isotope and otolith elemental studies of blue cod populations in the Fiordland area of southern New Zealand indicated spatially structured populations between the inner and outer zones of the fiord system (Beer et al., 2011; Rodgers and Wing, 2008). The limited mixing among subpopulations at scales of tens of kilometres, together with evidence of highly resident adults, is consistent with the results of earlier mark-recapture studies (Carbines and McKenzie, 2004). Smith (2012) analysed the mitochondrial control region of 475 blue cod samples from 14 locations, in order to assess the population genetic structure. This author's analysis showed that within the mainland, populations showed low genetic differentiation. However, the Chatham Islands population was genetically distinct from all other populations.

Mainland populations also showed a correlation between geographic distance and the level of genetic differentiation, indicating an isolation-by-distance pattern. This suggests that distinct populations among mainland sites could exist, although this was not detected by the mitochondrial DNA marker. It is possible that distinct populations among the mainland sites could be revealed through further investigations into population structure, using a higher resolution marker such as microsatellites.

Thesis aims

The general objective of this study is to advance previous population studies of blue cod, through the use of mitochondrial and microsatellite DNA markers to determine the population genetic structure of blue cod samples from around coastal New Zealand. The first section of this study (Chapter Two) advances the previous work of Smith (2012), which used mitochondrial DNA markers to characterise the genetic variation and population structure of blue cod by including samples from an additional location from the top of the South Island's west coast. Using mtDNA markers developed for blue cod (Smith 2012), the DNA sequence of a portion of mtDNA control region was determined for individuals sampled from a site in the Golden Bay area. These data were then pooled with sequence data from the rest of New Zealand (Smith 2012), and it was used to re-analyse the genetic variation and population structure.

Chapter Three focuses on the development of a set of blue cod-specific microsatellite markers. Next generation genomic sequence data was used to select 20 potentially amplifiable microsatellite loci. These loci were tested for genotyping quality and allelic variation, resulting in the development and characterisation of seven novel microsatellite markers. Chapter Four investigates the fine-scale population structure of blue cod. Seven microsatellite loci were used to determine the levels of genetic variation and to determine whether genetically distinct populations exist around the mainland of New Zealand. Chapter Five concludes, with a general discussion of results and management implications.

Chapter Two

The phylogeographic structure of New Zealand blue cod (*Parapercis colias*) based on mitochondrial DNA control region sequences

2.1 Introduction

Evolutionary processes, such as natural selection and genetic drift, typically result in the structuring of species into groups of individuals that are, to a greater or lesser extent, isolated from one another (Laikre et al., 2005). Identifying these independent stocks and understanding their connectivity is important for the effective management of a fishery. To achieve a sustainable fishery, management units should be based on a detailed knowledge of the population structure. To avoid over-harvesting, fishing induced mortality should be balanced by within-stock recruitment, and this is most effectively achieved at a level where demographically isolated populations are managed separately. Genetic markers have played an important role in defining stocks, because when two stocks are genetically differentiated they are also likely to be demographically uncoupled. However, when stocks are found to be genetically undifferentiated it can be difficult to draw conclusions about their connectivity, as migration rates could be either high or low, or stocks could have only recently become uncoupled (Ward, 2000). Understanding the patterns of genetic variation within a species is also important, because over-fishing can deplete genetic variation, alter the population structure of stocks, and even lead to selective genetic changes (Allendorf et al., 2008).

Parapercis colias (blue cod) is a temperate reef fish and a member of the Penguipedidae family (Forster 1801). It is an endemic species that is widely distributed throughout New Zealand's rocky coastal waters, down to depths of 150 metres. The species is a popular recreational game fish, particularly in the Marlborough Sounds, Otago and Stewart Island areas. The commercial catch is also significant, with a Total Allowable Commercial Catch (TACC) of 2600 tonnes per annum. Little is known about the spawning behaviour of *P. colias*, but it has been suggested that the eggs and larvae are pelagic for approximately 10 days after spawning (Henderson, 2009; Rodgers and Wing, 2008; Smith, 2012). Adult individuals are sedentary and have been shown to exhibit limited dispersal, with mark and

recapture studies indicating that most fish travel distances of less than 1 km (Cole et al., 2000; Mace and Johnston, 1983). However, the maximum distance travelled by an individual has been measured to be 156 km over a 20 month study period, suggesting that long range dispersal does occur (Carbines and McKenzie, 2001; Carbines and McKenzie, 2004). *P. colias* is a conspicuous benthic predator and, as is typical for a member of the Penguipedidae family, is territorial, haremic and a protogynous hermaphrodite (Carbines, 2004). If a dominant male is removed from an area, nearby females may undergo sex inversion, and because of this, it has been suggested that fishing pressure could result in unnaturally skewed sex ratios (Beentjes and Carbines, 2005). Several authors have proposed a high level of localised stock structure due to the relatively short pelagic larval duration, high territoriality and moderate movements of individuals among sites (Carbines and McKenzie, 2004; Mace and Johnston, 1983).

Stable isotope and otolith elemental studies of P. colias populations in the Fiordland area of southern New Zealand indicated spatially structured populations between inner and outer locations of the fiord system (Beer et al., 2011; Rodgers and Wing, 2008). The limited mixing among subpopulations at scales of tens of kilometres, together with evidence of highly resident adults, is consistent with the results of earlier mark-recapture studies (Carbines and McKenzie, 2004). A subdivided stock structure could mean that P. colias is highly vulnerable to localised depletion. This may have been the case when declining catch rates were found in the Marlborough Sounds area, which was linked to high recreational fishing pressure (Ministry for Primary Industries). The Marlborough Sounds fishery was first closed in 2008. A subsequent 2011 survey of the adult biomass revealed that the stocks were recovering, and the fishery was opened shortly thereafter. This pattern of depletion and recovery suggests that P. colias could be spatially structured, but that there is also migration of the species into depleted areas. The use of genetic data will help to determine whether this pattern of population subdivision has prevailed over a longer evolutionary time scale, or whether migration rates are sufficiently high to have led to intermixing throughout multiple locations.

Fisheries managers have used several different types of genetic markers to describe stock structure, including allozymes, mitochondrial DNA (mtDNA) and microsatellite DNA. Mitochondrial DNA has been one of the most popular markers, since it is relatively straightforward to collect sequence data, and because gene content is conserved across a wide range of species, and is often highly variable within a given species. In addition, the

matrilinear inheritance of mtDNA reduces the effective population size by a quarter of a nuclear locus, therefore rendering mtDNA more sensitive to demographic change (Hedgecock et al., 2007). However, as protogynous hermaphroditic fish can reproduce as both females and later as males, this may be less applicable for *P. colias*. All of these features make mtDNA a good choice for an initial assessment of the genetic population structure of a fishery stock. In light of this, the aim of this study was to determine the level of population diversity and differentiation of *P. colias* individuals sampled from a range of sites along the New Zealand coastline, including the Chatham Islands. DNA sequences from the mitochondrial control region were used to test whether New Zealand stocks comprise a large panmictic population, or whether there are distinct genetic populations, which would be consistent with previous findings.

2.2 Methods

Sample collection and DNA sequencing

An inshore trawl survey conducted by the National Institute of Water and Atmospheric Research Ltd (NIWA) along the west coast of New Zealand's South Island during March-April 2013 collected a total of 57 blue cod tissue samples from the Golden Bay area, at the top of the west coast of the South Island. The sex and age of the fish was randomised. Tissue was frozen and then placed in 80% ethanol for storage.

Genomic DNA was extracted using a standard phenol-chloroform protocol (Sambrook et al., 1989). Mitochondrial control region primers specifically designed for *P. colias* were used (Smith 2012). The forward primer (H-tRNA-PRO-BCO 5'-GAG CGC CGG TCT TGT AAA CCG G-3) and reverse primer (H-CCD-CR-BCO 5'-AGA ACG CTC GGC ATG TTG GGT-3') were used to amplify a 600 bp portion of the control region from the tRNA-Pro site to the highly conserved central domain region.

All PCRs consisted of 67 mM Tris-HCL pH 8.8, 16 mM (NH₄)₂SO₄, 2.0 mM MgCl₂, 200 μ M of each dNTP, 0.5 μ M of each primer, 0.6 μ g/ μ L Bovine Serum Albumin (BSA), 1 unit of BIOTAQTM DNA Polymerase (Bioline), and 1 μ L of template DNA in 10 μ L volume reactions. Thermocycling conditions were 95°C for 5 minutes, 94°C for 30 seconds, 54°C for 30 seconds, for 35 cycles, followed by a final extension of 10 minutes at 72°C. PCR products were purified with ExoSAP-it and sequenced with the reverse primer using an ABI 3730 Genetic Analyzer (Macrogen Incorporated, Korea).

Genetic analyses

The 57 DNA sequences from the Golden Bay sample were pooled with DNA sequence data reported in an unpublished thesis study by Smith (2012). The final dataset consisted of 529 individuals from 14 locations, one of which was the Chatham Islands, and 13 of which were coastal sites from mainland New Zealand (North, South and Stewart Island) (Table 2.1; Figure 2.1). All sequences were then edited and aligned in GENEIOUS 6.1 (Biomatters). Summary statistics such as the number of segregating sites (S), haplotypes (H), nucleotide (π) and haplotype diversity (Hd), private haplotypes (H_P) and the average number of pairwise differences (k) were estimated using DNASP 5 (Librado and Rozas, 2009). Fu's F statistic, Tajima's D, and pairwise fixation index (Φ_{ST}) were measured in ARLEQUIN 3.5 (Excoffier

et al., 2005). A Minimum Spanning Network was generated in POPART (www.popart.otago.ac.nz), which shows the relationships among haplotypes and their frequency within the population. The network was constructed with 1,000 iterations of the data. Rarefaction curves were constructed for each population using the ANALYTIC RAREFACTION 1.3 software (Holland, 2003). This simulates the number of haplotypes that would be discovered with increasing sample sizes. After plotting the number of 'discovered' haplotypes against increasing sample size, the size at which a plateau is reached indicates the sample size that could accurately reflect the diversity of the wider population.

Due to the distance between the Chatham Islands sample site (CHA) and the New Zealand mainland, genetic analyses were performed for CHA separately from the remainder of the sample sites (which collectively have been labelled as 'mainland sites' (ML)), where appropriate. Other groupings are self-explanatory, such as the North Island (NI) and South Island (SI) sites, with the Stewart Island site included in the SI group.

Genetic differentiation was estimated among and within populations, using Analysis of Molecular Variance (AMOVA) implemented in ARLEQUIN 3.5. The following six population groupings were tested, each with 10,000 permutations: (1) two regions, ML and the CHA site; (2) three regions, NI and SI, and CHA sampling sites; (3) two regions, NI and CHA, excluding SI; (4) two regions, SI and CHA, excluding NI; (5) two regions, sites organised into North and South biogeographic regions relative to the 42° latitude line. The CHA site was excluded from this analysis; and (6) two regions, NI and SI, excluding the CHA site. These groupings were based on potential genetic breaks identified in New Zealand's marine environment (Apte and Gardner, 2002; Ross et al., 2009).

A test for a pattern of isolation-by-distance was conducted by plotting linearised Φ_{ST} values against the geographic distance (kilometres) among sampling sites. Geographical distances were estimated as the shortest path, by sea, between two sample site coordinates using GOOGLE EARTH 7.1. Two exceptions to this were the Auckland and Ahipara sampling sites, which were combined to form the Northern site (NOR), due to their low sample sizes. Coordinate measurements were based on a point equidistant to the two sample locations (35°50'45"S, 174° 4'35"E). To assess the significance of the regression analysis, a Mantel test was implemented in ARLEQUIN 3.5 using 10,000 permutations.

Demographic history

Three approaches were used to explore the demographic history of *P. colias* populations: (a) Neutrality tests calculating Fu's Fs (Fu, 1997) and Tajima's D (Tajima, 1989) statistics, (b) mismatch distribution analyses, and (c) Bayesian Skyline Plots. Firstly, Fu's Fs and Tajima's D statistics were calculated using 1,000 simulations in ARLEQUIN 3.5, to determine their statistical significance. Significant negative values (P < 0.02) indicate an excess number of segregating sites, given the observed nucleotide diversity, which can be expected for populations having undergone a recent population expansion (Fu, 1997; Tajima, 1989).

Mismatch distribution analysis was conducted using DNASP 5, by estimating the frequency of pair-wise differences between haplotypes. The shape of the distribution indicates different demographic history scenarios. A population that has undergone a recent demographic expansion will exhibit a unimodal distribution, whereas a multimodal and erratic distribution reflects long-term stability (Rogers and Harpending, 1992). The data was fitted to a sudden expansion model (1,000 replicates) with Harpending's raggedness index (Harpending, 1994), and the sum of squared deviations (SSD) calculated for all regions (North Island, South Island, Mainland populations, Chatham Islands and all populations) in ARLEQUIN 3.5. To examine changes in population size, the demographic parameters Tau (τ), θ_0 and θ_1 were estimated from the mismatch distribution in ARLEQUIN 3.5. The time since population expansion occurred could then be estimated using the equation $t = \tau/2\mu$ (Rogers and Harpending, 1992), where μ is the mutation rate of the sequence (not to be confused with the mutation rate per nucleotide). The value of μ can be measured using the formula $\mu = 2\mu k$, where μ is the mutation rate per nucleotide site and k is the number of analysed nucleotide base pairs. As the mutation rate was not known for the control region sequence in P. colias, time since divergence was estimated by allowing the mutation rate to vary between 2.0×10^{-8} substitutions/site/year (s/s/yr) through to a more conservative 5.0 x 10^{-8} s/s/yr (Bowen et al., 2006; Mach et al., 2011). The frequency of observed pairwise differences between haplotypes was plotted separately for the mainland sites and the Chatham Islands, in order to investigate the demographic history of both regions.

The demographic history was investigated further using the skyline plot method (Drummond et al., 2005; Ho and Shapiro, 2011). Bayesian skyline plots were constructed in BEAST 1.8.0 (Drummond et al., 2012) for two population groupings; the Chatham Islands population and the mainland New Zealand populations. The Marko Chain Monte Carlo simulations were run

for 10^7 iterations, with the GTR substitution model, and a strict molecular clock. Two mutation rates were used; 2.0×10^{-8} and 2.0×10^{-7} (Bowen et al. 2006). The results were then inspected, and skyline plots constructed, in TRACER (Rambaut et al., 2014).

2.3 Results

Genetic diversity

DNA sequences from the mtDNA control region were determined for 529 P. colias individuals sampled from 14 locations. The sequences were aligned and trimmed to 493 bp long and showed asymmetric base frequencies of $\pi_A = 0.330$, $\pi_C = 0.155$, $\pi_G = 0.183$ and $\pi_T =$ 0.332. The alignment contained 70 polymorphic sites with two indels; 48 of these were parsimony-informative sites with an overall nucleotide diversity (π) of 0.0057. A total of 97 haplotypes were recorded with a total haplotype diversity (Hd) of 0.782, and the within-site haplotypic diversity ranged from 0.693 at Puysegur Point, to 0.923 in the Northern site. All sites are dominated by two main haplotypes, with 40% of individuals characterized by Haplotype 1 and a further 20% of individuals exhibiting Haplotype 2. Haplotype 1 was also the only haplotype to appear in every site. The Chatham Islands population had the highest number of haplotypes, equal to 23, of which 12 (private) are not found in any other population. The haplotype distribution appears to be different in the Chatham Islands, with only 6% of individuals showing Haplotype 1, and 40% showing Haplotype 2. The Chatham Islands Fu's F statistic (-16.47) is also different from that determined for the other New Zealand sites, which ranged from -0.099 to -6.003. The rarefaction curves do not reach a plateau, indicating that if larger sample sizes were used, additional haplotypes could be discovered (Figure 2.2). The rarefaction analysis also showed that after accounting for differences in sample size, the sites do differ in terms of the number of haplotypes present.

Haplotype frequencies within each site are shown in Figure 2.1, with haplotypes within a site having a frequency of less than 0.05 grouped together, and private haplotypes also being grouped separately. The Minimum Spanning Network revealed the presence of two main haplotypes, with the majority of other haplotypes branching off these in a star-like pattern (Figure 2.1). There are four mutational differences between these two haplotypes, and the maximum number of mutational steps between two haplotypes is 11. Black circles represent mutational steps not represented by a haplotype.

Location		Lat	Long	N	h	Hd	S	K	$H_{\rm P}$	π	Fu's Fs	Tajima's D
Northern	NOR	35	174 [°] 4'35.40"E	14	10	0.9230	20	4.637	5	0.0094	-2.739	-0.877
Bay of Plenty	BOP	37	176 ⁰ 10'5.41"E	22	7	0.7489	10	2.701	2	0.0055	-0.099	-0.052
D'Urville Island	DUI	40	173 ⁰ 47'1.78"E	32	8	0.7339	11	2.631	2	0.0054	-0.238	-0.117
Pelorus Sound	PEL	40	173 ⁰ 54'54.70"E	32	10	0.7540	12	2.675	6	0.0054	-1.740	-0.329
Queen Charlotte Sound	QCS	41	174 ⁰ 8'39.00''E	33	14	0.8390	17	2.890	7	0.0059	-5.151	-1.041
Wellington	WEL	41	174 ⁰ 42'46.35"E	41	12	0.7439	18	2.788	5	0.0057	-2.539	-1.097
Kaikoura	KAI	42	173 ⁰ 39'40.94"E	42	15	0.7666	19	2.834	3	0.0058	-5.238	-1.168
Akaroa	AKA	43	172 ⁰ 59'43.05"E	39	12	0.7446	15	2.516	7	0.0051	-3.208	-0.932
Chatham Islands	CHA	43	176 ⁰ 31'6.81"W	64	23	0.8423	22	2.264	12	0.0046	-16.275	-1.597
Fiordland	FIO	45	166 ⁰ 42'25.41"E	26	11	0.7969	12	2.677	5	0.0055	-3.302	-0.326
Otago	OTA	45	170 ⁰ 47'41.50"E	49	13	0.7279	17	2.527	4	0.0051	-3.387	-0.840
Puysegur Point	PUY	46	166 ⁰ 35'31.34"E	40	12	0.6936	14	2.364	4	0.0048	-3.456	-0.889
Stewart Island	STW	46	168 ⁰ 13'42.66"E	38	12	0.7056	17	2.663	3	0.0054	-2.997	-1.119
Golden Bay	GDB	40	173 [°] 1"3.14"E	57	17	0.7187	24	2.823	8	0.0058	-6.003	-1.461
Total				529	97	0.7818	73	2.837	73	0.0057	-123.429	-2.110

Table 2.1 Summary statistics including sample size (N), number of haplotypes (h), haplotype diversity (Hd), segregating sites (S), pairwise differences (K), private haplotypes (H_P) and nucleotide diversity (π).



Figure 2.1 (A) Sample sites and sizes with haplotype frequency charts. Haplotypes with a frequency below 0.05 were grouped as 'rare', and haplotypes seen in one population only were grouped as 'private'. (B) The minimum spanning network displaying the relationships between all 97 haplotypes. The circle size is indicative of the number of individuals with each haplotype. Each branch represents a mutational step, with additional mutational steps indicated by a black circle.

Population structure

Pairwise Φ_{ST} indicated a significant level of genetic differentiation between the Chatham Islands and all of the mainland New Zealand sites (Table 2.2). The difference is most prominent between the Chatham Islands and Golden Bay (Φ_{ST} =0.156 *P*<0.0001), followed by Stewart Island (Φ_{ST} =0.141 *P*<0.0001). The site that is the least differentiated from the Chatham Islands is Pelorus Sound (Φ_{ST} =0.056 *P*<0.001). Very little differentiation exists between the mainland New Zealand sites. Differentiation was found when the Pelorus Sound site was paired with the Northern and Golden Bay sites (Φ_{ST} =0.055, 0.036 *P*<0.05), however this is no longer significant after Bonferroni correction. A pattern of isolation-by-distance was observed for the New Zealand sites, with a positive correlation between Φ_{ST} and geographic distance (R²=0.117 *P*<0.05). This pattern is stronger when the Chatham Islands are excluded (Figure 2.3, R²=0.198 *P*<0.005).

The AMOVA showed that a significant proportion of genetic differentiation can be attributed to the regional separation between the Chatham Islands and the rest of New Zealand. This pattern is most evident between the Chatham Islands and the South Island, with this separation accounting for 24% of the genetic variation ($\Phi_{CT}=0.248 P=0.089$), followed by the Chatham Islands and mainland New Zealand grouping with 23.51% of the variation ($\Phi_{CT}=0.235 P=0.074$). The test for three regions (NI, SI and CHA) shows that 13.34% of the variation can be explained by this comparison ($\Phi_{CT}=0.133 P=0.003$). The remainder of the variation is observed within the populations. When the Chatham Islands site is excluded, there is no evidence of further genetic differentiation between the mainland sites, with nearly all genetic variation present within sites. This was seen under both the North Island/South Island split, as well as the separation of sites along the 42nd parallel.



Figure 2.2 Rarefaction curves for each sample site.



Figure 2.3 Linearized pairwise F_{ST} plotted against geographic distance (km) between all mainland sites, with the Chatham Islands excluded (n=78, R²=0.198 *P*<0.005).

	AKA	NOR	BOP	CHA	DUI	FIO	KAI	OTA	PUY	PEL	QCS	STW	WEL	GDB
AKA		0.117	0.432	0.000	0.811	0.496	0.883	0.658	0.712	0.405	0.676	0.586	0.649	0.126
NOR	0.034		0.081	0.000	0.162	0.081	0.126	0.081	0.036	0.036	0.198	0.063	0.162	0.126
BOP	-0.003	0.043		0.000	0.739	0.874	0.496	0.423	0.423	0.901	0.405	0.270	0.153	0.153
CHA	0.108*	0.102*	0.062*		0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000
DUI	-0.016	0.024	-0.015	0.107*		0.459	0.838	0.955	0.811	0.432	0.541	0.739	0.766	0.496
FIO	-0.006	0.037	-0.024	0.060*	-0.007		0.568	0.486	0.351	0.757	0.676	0.252	0.144	0.108
KAI	-0.013	0.028	-0.006	0.104*	-0.014	-0.010		0.802	0.766	0.333	0.613	0.757	0.586	0.288
OTA	-0.008	0.032	-0.003	0.123*	-0.019	-0.003	-0.010		0.964	0.216	0.252	0.919	0.757	0.667
PUY	-0.010	0.048	-0.002	0.134*	-0.015	0.000	-0.012	-0.015		0.324	0.216	0.964	0.595	0.496
PEL	0.003	0.055	-0.023	0.056	-0.005	-0.016	0.002	0.008	0.009		0.261	0.126	0.054	0.009
QCS	-0.010	0.012	-0.002	0.068*	-0.007	-0.010	-0.009	0.002	0.006	0.002		0.207	0.360	0.072
STW	-0.007	0.037	0.006	0.141*	-0.013	0.003	-0.011	-0.015	-0.016	0.020	0.006		0.793	0.667
WEL	-0.009	0.015	0.016	0.138*	-0.014	0.011	-0.008	-0.010	-0.007	0.024	0.001	-0.011		0.532
GDB	0.012	0.023	0.024	0.156*	-0.004	0.023	0.003	-0.007	-0.003	0.036	0.020	-0.009	-0.006	

Table 2.2 Pairwise Φ_{ST} values below the diagonal, with P < 0.05 in bold, and values remaining significant after Bonferroni correction marked with an asterisk, *P*-values are provided above the diagonal.

Demographic history

The mismatch distributions of pairwise differences were fitted to a sudden expansion model for the Chatham Islands, North Island, South Island, mainland sites, and then all sites. Distribution plots were drawn for the Chatham Islands site and the mainland region (Figure 2.4). The observed pairwise differences between haplotypes do not differ significantly from the smooth unimodal distribution that could be expected under a sudden expansion model. This is seen in all population groupings, as Harpending's raggedness index ranged from 0.0639-0.0814 (P=0.137-0.345) and the sum of square deviations ranged from 0.0188-0.0367 (P=0.149-0.367) (Table 2.3). Fu's Fs in all sites is -25.889 (P<0.001) and SI and ML are very similar to this. The North Island and Chatham Islands are closer to zero, with Fs=-7.908 (P=0.006) and -16.275 (P<0.001), respectively. Fu's Fs statistic is negative and significant in all population groups. Tajima's D statistic was also negative and significant in all populations apart from the North Island region. The estimated time since expansion in the Chatham Islands population ranges between 6-16 thousand years, depending on the mutation rate. This is more recent than the estimation for the other New Zealand sites, which lie in the range of between 50 and 130 thousand years (Table 2.3). Bayesian skyline plots indicated that the mainland New Zealand region underwent a population expansion but this was much less prominent in the Chatham Islands site (Figure 2.5).

Table 2.2 Results from Analysis of Molecular Variance (AMOVA) of *P. colias* with six a priori geographic groupings: (1) mainland New Zealand separated from Chatham Islands, (2) North Island, South Island and the Chatham Islands all grouped separately, (3) North Island separated from the Chatham Islands (excluding the South Island), (4) the South Island separated from the Chatham Islands (excluding the North Island), (5) regional separated of the mainland sites along the 42° latitude line, and (6) the North Island separated from the South Island. The d.f. is the degrees of freedom, Var is variance component, % Var is variance component in percentage of total variation.

Regions	Among groups						ong populat	ions within	groups		Within populations				
	d.f.	Var	% Var	$\Phi_{\rm CT}$	P-value	d.f.	Var	% Var	$\Phi_{\rm SC}$	P-value	d.f.	Var	% Var	$\Phi_{\rm ST}$	<i>P</i> -value
ML/CHA	1	0.409	23.51	0.235	0.074	12	-0.006	-0.35	-0.005	0.724	515	1.337	76.85	0.232	0.000
NI/SI/CHA	2	0.20480	13.34	0.133	0.003	11	-0.006	-0.42	-0.005	0.736	515	1.337	87.08	0.129	0.000
NI/CHA	1	0.31672	18.59	0.186	0.246	2	0.032	1.89	0.023	0.202	137	1.355	79.51	0.205	0.000
SI/CHA	1	0.425	24.76	0.248	0.089	9	-0.011	-0.63	-0.008	0.874	441	1.302	75.87	0.241	0.000
N42/S42	1	0.003	0.24	0.002	0.102	11	-0.009	-0.64	-0.006	0.784	452	1.365	100.40	-0.004	0.729
NI/SI	1	0.001	0.08	0.001	0.229	11	-0.007	-0.53	-0.005	0.735	452	1.365	100.45	-0.005	0.731

Table 2.3 Estimation of population expansion using Fu's Fs, and mismatch distribution analyses: sum of squared deviation (SSD), Harpending's raggedness index, tau (τ), population size before expansion (θ_0), population size after expansion (θ_1) and time since population expansion occurred (t) with *P*-values and confidence intervals in brackets.

	Fu's Fs (P)	Tajima's D (P)	SSD (P)	Raggedness (P)	τ (95% CI)	θ ₀ (95% CI)	θ ₁ (95% CI)	t
CHA	-16.275 (<0.001)	-1.597 (0.029)	0.0188 (0.367)	0.0639 (0.245)	0.654 (0.00-5.01)	1.920 (0.00-5.20)	99999 (6.02-99999)	16,582-6,633
NI	-7.908 (0.006)	-1.454 (0.052)	0.0367 (0.149)	0.0814 (0.137)	5.105 (0.67-9.70)	0.000 (0.00-1.85)	6.094 (2.99-99999)	129,437-51,775
SI	-26.188 (<0.001)	-2.089 (0.001)	0.0351 (0.247)	0.0663 (0.345)	5.125 (0.27-10.29)	0.007 (0.00-1.11)	3.975 (1.49-99999)	129,944-51,978
ML	-25.991 (<0.001)	-2.065 (0.00)	0.0347 (0.275)	0.0662 (0.337)	5.125 (0.39-10.89)	0.002 (0.00-1.44)	4.239 (1.78-99999)	129,944-51,978
All sites	-25.889 (<0.001)	-2.066 (0.001)	0.0351 (0.176)	0.0646 (0.240)	4.969 (0.00-29.97)	0.002 (0.00-2.79)	5.027 (1.05-99999)	125,989-50,396



Figure 2.4 Mismatch distributions of the frequency of pairwise differences in sequences from (A) mainland New Zealand and (B) the Chatham Islands. The solid line indicates the observed pairwise differences; the dashed lines correspond to the expected distribution of a constant population size, and the dotted line is the expected distribution under a growth-decline population model. The blue area represents the 95% confidence interval of observed pairwise differences.



Figure 2.5 Bayesian Skyline plots. (A) Chatham Islands with a mutation rate of 2.0×10^{-7} substitution/sites/year (s/s/y) (B) Chatham Islands with 2.0×10^{-8} s/s/y (C) Mainland New Zealand sites with 2.0×10^{-7} s/s/y (D) Mainland New Zealand sites with 2.0×10^{-8} s/s/y.

2.4 Discussion

The DNA sequence analyses reported in this study revealed the genetic structure of *P. colias* on two levels: overall, there is significant genetic differentiation between the Chatham Islands and the mainland New Zealand sites, and secondly, within the mainland sites a pattern of isolation-by-distance was detected. These findings are somewhat in concordance with previous studies, which have suggested that *P. colias* populations could be highly structured, due to limited adult dispersal and a relatively short larval duration (Beer et al., 2011; Cole et al., 2000).

Population structure

Evidence of a genetic break between the Chatham Islands and mainland New Zealand is supported by significant pairwise Φ_{ST} between all sites, with the exception of PEL and CHA (Φ_{ST} =0.056 *P*=0.009), which is no longer significant after Bonferroni correction. The composition of the haplotypes is also distinctive in each region; the predominant haplotype in mainland sites is Haplotype 1. In contrast, that haplotype occurred at a low frequency in the Chatham Islands site, and Haplotype 2 was most common (Figure 2.1). Given the geographical distance between the Chatham Islands and the main islands of New Zealand, the level of genetic differentiation between these locations was unsurprising, as *P. colias* is a coastal demersal species and it is unlikely that juveniles and adults would disperse over deeper water. Previous studies suggested that they are generally found down to a depth of not more than 150 m (Beentjes and Carbines, 2005).

Although the early life history of *P. colias* has received little attention, it has been estimated that the pelagic larval duration (PLD) is approximately 10 days (Henderson, 2009). Using simulations from satellite data, Chiswell (2009) projected that it would take 30 to 50 days for the passive dispersal of the larvae of a species to reach the Chatham Islands from the mainland coast. This indicates that larval dispersal of *P. colias* between the Chatham Islands and mainland New Zealand could be infrequent. The oceanographic current patterns around New Zealand are complex. The general trend of major current systems is an eastward flow, with the currents splitting when they reach the continental shelf and continuing around the bottom of the South Island and through Cook Strait, out towards the Chatham Rise (Heath, 1985). The use of PLD as a proxy for dispersal capabilities is not straightforward, since many factors such as ocean currents and larval behaviour can influence dispersal (Cowen and
Sponaugle, 2009; Waples, 1998). A study of New Zealand's coastal benthos (Ross et al., 2009) showed that the correlation between larval duration and genetic differentiation was more likely in taxa with shorter larval durations, such as *P. colias*.

Very little genetic differentiation is observed between the mainland New Zealand *P. colias* sample sites, and no evidence was found of significant differentiation on either side of the Cook Strait. At its narrowest point, the Cook Strait is only 22 km wide and, although multiple mark and recapture studies have shown that most fish remain within 1 km, long range dispersal of *P. colias* can occur, with one individual dispersing 156 km over a 20 month study period (Carbines and McKenzie, 2001). While many New Zealand marine species have shown differentiation between North and South Island populations, these have been predominantly marine invertebrates and plants (Ross et al., 2009). A study comparing the population structure of eight different endemic triplefin species (family Tripterygiidae) found that, even in species showing strong population structuring, the Cook Strait was not an obvious barrier to gene flow (Hickey et al., 2009). This also appears to be the case for *P. colias*.

A lack of differentiation between mainland sites indicates a high enough level of gene flow between sites to homogenize variation; however, this does not necessarily equate to high levels of demographic connectivity (Ward, 2000). The population structure of mainland sites showed an isolation-by-distance pattern, as the pairwise Φ_{ST} is positively correlated with geographic distance (Figure 2.3). Although this pattern is weak ($R^2=0.198 P < 0.005$), it indicates that gene flow may be restricted as geographical distance increases. Demographic connectivity is dependent on the proportion of migrants, in contrast to genetic connectivity which is maintained by a certain number of migrants, regardless of population size (Lowe and Allendorf, 2010). Therefore, even small departures of F_{ST} from zero could theoretically indicate demographically uncoupled populations (Ovenden, 2013; Waples and Gaggiotti, 2006). Consequently, it could be misleading to assume demographic connectivity among mainland P. colias populations, particularly as tagging and microchemistry studies have demonstrated the sedentary behaviour of P. colias. It is likely that through long range dispersal, migration is above the threshold for genetic connectivity, but below the level for demographic connectivity, as indicated by the pattern of isolation-by-distance along with estimates of movement.

A gradient of change in mainland *P. colias* populations can also be observed at an ecological level. Not only do *P. colias* occur less frequently in northern latitudes, but their size and age at sexual maturity also differ between locations. Fish in northern New Zealand mature earlier and at smaller sizes compared to fish in the south (Ministry for Primary Industries, 2014). This suggests that the isolation-by-distance pattern may not be driven primarily by limited dispersal, due to geographical distance, but instead by influences that could be driving biological differences, such as habitat and environmental parameters (i.e. temperature), or competition from other marine species.

It is important to consider the distributions of sample locations and the number of samples when interpreting the results (Waples, 1998). As *P. colias* is less abundant to the north of Cook Strait compared to the south, there were a smaller number of samples and the fewer sampled locations in the northern areas of their distribution. The isolation-by-distance analysis was repeated with the Northern sample site (NOR) excluded in order to reduce the possibility of a statistical bias. However, that exclusion of the northern area did not change in the overall result. Fishing areas around the Cook Strait are of particular concern to fisheries managers, due to heavy recreational fishing in the area, particularly in the Marlborough sounds. It is therefore important to note the relatively high pairwise Φ_{ST} observed between the Golden Bay and Pelorus Sound sites (Φ =0.036 *P*=0.009), which appears as an outlier amongst the very low levels of Φ_{ST} observed between other mainland population pairs. This result is particularly interesting because of the proximity of the two sampling locations. While not statistically significant (after Bonferroni correction), this could suggest a departure from panmixia.

Demographic history

Measures of genetic variation can provide insight into the demographic history of *P. colias*. Multiple lines of evidence suggest that *P. colias* populations have undergone a recent population expansion. High levels of haplotypic diversity with low levels of nucleotide diversity, as well as a star-like haplotype network, are characteristic of populations that have undergone rapid expansion (Aboim et al., 2005). Mismatch distribution analyses further supports a population expansion as neither mainland New Zealand nor Chatham Islands significantly differed from a pairwise distribution expected under an expansion model. Bayesian skyline plots indicated size expansion in the mainland population, but this was less evident in the Chatham Islands site (Figure 2.5). This was also reflected in the Fu's Fs which

was consistent with skyline plot analyses in that the evidence of a population expansion was stronger in the mainland groupings compared with the Chatham Islands site (Table 2.4).

Inevitably, high levels of sustained harvesting pressure will, at some level, lead to genetic changes of a fisheries stock (Allendorf et al., 2008). Fishing pressure is often size selective, which could have unknown genetic consequences in protogynous hermaphroditic fish species, where the larger fish are generally male. A naturally occurring bias in sex ratios could be exacerbated by males being continually removed from a population, promoting further sex inversion (Beentjes and Carbines, 2005). This could have important management implications as a departure from a 1:1 sex ratio could significantly reduce the effective population size of P. colias populations (Waples, 1998). For hermaphroditic fish species it is possible for individuals to pass on alleles to offspring, both as mature females, and then again as males after undergoing sex inversion (Armsworth, 2001). This makes it more difficult to evaluate the extent to which current population genetic theories can correctly model genetic trends in hermaphroditic fish. For gonochoristic species, the effective population size (Ne) of maternally inherited mtDNA is considered to be a quarter of the size of autosomal loci. However, this rule of thumb may not be appropriate for hermaphroditic species and, consequently, could affect the sensitivity of mtDNA markers to fluctuations in population demography. How this relates to P. colias is uncertain, as some studies have found a female bias in the sex ratio (Carbines and Beentjes, 2009), while others have found a strong male bias (Blackwell, 2006), which the authors suggest could be the result of over-compensatory sex inversion due to harvesting pressure (Beentjes and Carbines, 2005; Henderson, 2009).

One limiting factor of this study, which was highlighted by the rarefaction curve analysis, was the small sample sizes of some sites. The curves did not reach a plateau, indicating that the sample sizes used in this study are not sufficiently large to capture all of the diversity present in *P. colias* populations (Figure 2.2). This is also reflected by the high numbers of unique haplotypes represented by single individuals, which could have affected the power of the tests (Waples, 1998; Waples and Gaggiotti, 2006).

From a management perspective, an isolation-by-distance pattern can be the most difficult type of structure to incorporate into fisheries management plans. An isolation-by-distance pattern indicates that populations are not panmictic, and therefore should not be treated as one stock. However, unlike when population structure analyses detect distinct populations, this type of pattern provides no evidence of genetic breaks between populations which can then be used to define management boundaries (Lowe and Allendorf, 2010; Manni et al., 2004). The pattern of isolation-by-distance could also indicate that genetically distinct populations around the mainland do exist, but that the pattern could not be detected by mtDNA. As mtDNA does not undergo recombination, the entire mitochondrial genome acts as a single locus, which could limit its statistical power when compared to other markers. The use of nuclear markers such as microsatellites or single nucleotide polymorphisms (SNPs), which sample multiple points along the genome, could be necessary to further deduce population structure (Waples, 1998).

Chapter Three

Development of novel microsatellite DNA markers for Blue Cod (*Parapercis colias*)

3.1 Introduction

Microsatellite DNA, or simple sequence repeats (SSR), are among the most extensively used DNA markers in fisheries genetics. They are made up of tandemly repeated DNA sequences of 1-6 base pair motifs. The alleles are defined based on the number of repeats (allelic diversity) and these can be determined using PCR amplification and the standard size separation technique of electrophoresis. The co-dominant inheritance of alleles and their highly polymorphic nature make microsatellite DNA ideal for population genetic studies, in particular investigating fisheries structure, species identification, and the analysis of parentage (Ward, 2000).

Microsatellite-type repeats are ubiquitous and widely distributed in the genome (Ellegren, 2004). Each microsatellite type is classified according to the nucleotide composition of the repeat sequence and the number of repeats, e.g. $(AT)_{15}$ describes a dinucleotide microsatellite sequence of AT, repeated 15 times. Microsatellites can be 'perfect' when repeat units are uninterrupted, 'imperfect' e.g. $(AT)_{10}GC(AT)_5$, or 'compound' e.g. $(AT)_{10}(CG)_{15}$. Microsatellite loci are often assumed to be selectively neutral because they are less likely to occur within a coding sequence, although several studies showed that they can be associated with biological functions, such as chromosome structuring (Chistiakov et al., 2006).

Although some aspects of microsatellite evolution are still unclear, it is thought that size variation predominantly occurs through polymerase slippage during DNA replication (Chistiakov et al., 2006). Polymerase slippage can occur during replication when the polymerase enzyme dissociates briefly from the template strand. Due to the repetitive nature of the DNA, subsequent reattachment of the enzyme can align the strands incorrectly, resulting in a different number of repeats on each strand. If this is not corrected by DNA mismatch repair mechanisms, the result is either a gain or a loss of one or more repeat units. This process can allow the same allele to arise multiple times, generating size homoplasy. Microsatellite size homoplasy describes a situation in which electromorphs (DNA fragments

of the same size detected using electrophoresis) may be identical in state, although not identical by descent (Estoup et al., 2002). Many population genetic analyses assume the same size allele means it was inherited from a common ancestor when inferring relatedness. However, Estoup (2002) argues that the levels of homoplasy are typically not high enough to pose a significant problem for most applications, unless loci show strong allele size constraints coupled with a high mutation rate.

Several mutation models have been proposed to account for uncertainty about the mechanism that generates microsatellite polymorphism. The step-wise mutation model (SMM) states that a mutation only involves the loss or gain of a single repeat unit (Kimura and Ohta, 1978). This allows information about the relationship of alleles to be incorporated into estimates of differentiation. The infinite alleles model (IAM) states that a mutation can result in any number of repeat changes and every mutation leads to a novel allele (Kimura and Crow, 1964). Studies have shown that the most realistic models are the two-phase model (TPM) and the generalized stepwise model, both of which are relaxed variations of the SMM (Estoup et al., 2002). The rate of mutation may differ, depending on the loci, with the length of the microsatellite also influencing the mutation rate (Chistiakov et al., 2006).

A major consideration to be assessed before undertaking studies that use microsatellite DNA markers, is the time and cost involved in developing and optimising a new set of loci. Species-specific primers often need to be developed for new study species (Zane et al., 2002). A range of microsatellite DNA isolation protocols have been reported, such as hybridisation selection (Armour et al., 1994) or the linker-ligation enrichment protocol (Hamilton et al., 1999). Most of these protocols are variations on the traditional approach, involving the production of size-selected genomic libraries and the screening of clones with repeat sequence specific probes (Chistiakov et al., 2006).

Until recently, clone-based protocols have been technically challenging and time consuming. However, advances in DNA sequencing technology, in particular the advent of massively paralleled sequencing methods, have enabled the cost-effective and rapid isolation of potential amplifiable loci (PALs). The non-cloning methods of shotgun DNA sequencing, coupled with improved bioinformatics searching tools, has now eliminated the need for DNA probe-by-probe screening of a library (Abdelkrim et al., 2009). This direct DNA sequencing approach to microsatellite development, termed 'seq-to-SSR', has numerous advantages including the fact that it does not require short-read sequence assembly, and that all microsatellite repeat types can be identified simultaneously.

The Roche 454 pyrosequencer (GS-FLX Titanium, 454 Life Sciences) has been the most commonly used sequence platform for microsatellite isolation, because it generates long average sequence reads (400-600bp) (Gardner et al., 2011). The DNA sequence reads are only useful if they contain a sufficiently long sequence region flanking the repeat sequence, which can accommodate PCR primer sites. The 454 procedure involves nebulizing the purified genomic DNA sample into small fragments, which are then attached to beads via adapter sequences to facilitate emulsion PCR, allowing sequence reads to be determined using pyrosequencing.

The development of fluorescent labelling via a third universal primer can also reduce the cost of microsatellite allele detection. Rather than direct labelling of the primer pairs with a fluorescent dye, a universal primer sequence can be incorporated into one primer that corresponds to a fluorescently labelled third primer (Schuelke, 2000). Loci can then be combined at the PCR stage (multiplexing) or genotyping stage (pool-plexing), as long as loci with overlapping allele size ranges are labelled with a different fluorescent dye.

The aim of the research presented in this chapter was to develop species-specific microsatellite DNA markers for *P. colias* using massively parallel 454 pyrosequencing technology. These markers can potentially be used for a range of studies such as genetic population structure studies or parentage analysis. Chapter four reports on how these markers were used to estimate levels of genetic diversity and fine scale population structure of New Zealand *P. colias*. The overall goal of the research is to provide relevant information for stock assessment models for use in the management of the fishery.

3.2 Methods

Genomic DNA was isolated from one blue cod individual using a standard phenolchloroform extraction protocol (Sambrook et al., 1989). Purified genomic DNA was sent in 2012 to the High Throughput DNA Sequencing Unit at Otago University, and run on 1/8th of a plate on the Roche 454 Genome Sequencer FLX system. The resulting library contained over 280,000 DNA sequence reads. MSATCOMMANDER was used to identify the number of repeat sequences (Faircloth, 2008). DNA sequence reads containing repetitive DNA were detected using the QDD 2 software, and primers were developed for the candidate loci using the Primer3 function with default settings (Meglecz et al., 2010). This program returned an excel file containing approximately 35,000 potential microsatellite marker loci (PALs) with corresponding primer pairs.

PALs with perfect repeat sequences were favoured in order to maintain the equivalence between the electromorph and amplicon sequences (Guichoux et al., 2011). It has been well documented that microsatellite DNA with higher numbers of repeat units typically show higher levels of polymorphism; in order to maximise the chances of sampling high allelic variation, PALs with more than 11 repeat units were favoured (Guichoux et al., 2011). Motif type was also considered, as shorter motif types (such as dinucleotides) allow for smaller allele size ranges, which can be beneficial when using pool-plexing during the allele sizing stage. However, dinucleotide motifs are also more likely to cause a stutter effect during PCR amplification. The stutter effect is the result of polymerase slippage during the DNA synthesis stage of PCR, resulting in a range of amplicon sizes around the actual allele amplicon. This causes multiple fluorescent peaks on a chromatogram, which can hinder the identification of the allele peak and allele calling (Guichoux et al., 2011). The 35,000 candidate primer sets were ranked, and 20 PALs were selected on the basis the multiple factors previously mentioned such as PCR product length, motif unit type and size, as well as the primer penalty score that was implemented in Primer3, which incorporates factors such as GC content. The 20 PAL primer sequences were also mapped back onto the original sequence read using GENEIOUS 7.1 (Biomatters) to consider the primers' position relative to the microsatellite repeat sequence, and to check for repeat sequences within primers. The Primer3 function implemented in GENEIOUS was also used to determine whether other primer sequences were more appropriate. The 20 selected primer pairs were labelled BCO_01 to BCO_20 and were synthesized with the addition of an M13 sequence (5'-TGT AAA ACG

ACG GCC AGT) on the 5' end of the forward primer to enable M13-tail fluorescent labelling (Schuelke, 2000).

Locus primer pairs were assessed by PCR amplifying 15 individuals from 4 or 5 locations to optimise the PCR conditions, and to assess the genotyping quality and levels of allelic variation. Primers were trialled in three stages; first, the loci were amplified separately without the addition of M13 fluorescent labelling, and the PCR conditions were optimised. Secondly, the loci were amplified with the addition of an M13 primer labelled with FAM, VIC or PET fluorescent dyes, and genotyped by the Massey University Genome Service (using an ABI 3730 Genetic Analyzer) to determine the range of allele sizes and levels of polymorphism. Lastly, the loci were optimised for pool-plex by combining PCR products into two groups that minimised the overlap of allele size ranges among loci.

Microsatellite loci were amplified separately in PCR 10 µl volumes comprising 67 mM Tris-HCL pH 8.8, 16 mM (NH₄)₂SO₄, 2.0 mM MgCl₂, 200 µM of each dNTP, 0.1 µM of forward primer, 0.5 µM of reverse primer, 0.5 µM of M13 primer (FAM/VIC/PET), 0.6 µg/µL of Bovine Serum Albumin (BSA), 1 unit of BIOTAQTM DNA Polymerase (Bioline), and 1 µL of template DNA. The thermocycling conditions were: 95 °C for 5 minutes, 94 °C for 30 seconds, 54 °C for 40 seconds, 72 °C for 40 seconds for 30 cycles, followed by a second set of 10 cycles: 94 °C for 30 seconds, and 53 °C for 40 seconds and 72 °C for 40 seconds. These sequences were followed by a final extension of 72 °C for 10 minutes. The success of amplification was checked using electrophoresis in a 2% agarose gel and ethidium bromide staining.

PCR products were pool-plexed into two combinations of loci, based on an analysis conducted using MULTIPLEX MANAGER 1.2 (Holleley and Geerts, 2009). Amplicons were then sent for genotyping by capillary separation at the Massey University Genome Service, which used an ABI3730 Genetic Analyzer (Applied Biosystems Inc). The GeneScanTM 500 LIZTM Size Standard was used as a relative size marker for determining PCR fragment sizes. PEAK SCANNER v2 (Applied Biosystems) was used for allele identification and size calling, and the Excel macro AUTOBIN was used to facilitate allele size binning.

The presence of null alleles, scoring errors and large allelic drop-out was assessed using MICRO-CHECKER 2.2.3 (van Oosterhout et al., 2004). Deviation from the Hardy Weinberg Equilibrium (HWE) and the inbreeding coefficient (F_{IS}) were measured in GENEPOP 1.2

(Raymond and Rousset, 1995). Diversity estimates, such as number of alleles, observed (H_o) and expected heterozygosities (H_e) and tests for linkage disequilibrium, were estimated using ARLEQUIN 3.1 (Excoffier et al., 2005).

3.3 Results

The DNA sequence library generated from 454 sequencing contained 288,946 reads ranging from 21 to 1319 bp in length, with a mean of 281.7, SD 138.9 (Appendix 1). Approximately 68,000 of the reads were at least 400 bp long, the suggested read length to enable primer sites to be identified either side of the repeated sequences. A search of sequence reads using MSATCOMMANDER identified 41,351 dinucleotide, 15,236 trinucleotide, 10,605 tetranucleotide, and 1098 pentanucleotide microsatellite-type DNA sequences. Not all of these were appropriate for the development of primer pairs. QDD returned approximately 35,000 PAL primer pairs for further assessment.

The 20 PALs chosen for primer synthesis consisted of five dinucleotides, 11 trinucleotides, two tetranucleotides, one pentanucleotide and one tetra-di complex, and these were labelled from BCO_01 to BCO_20. Of these 20 PALs, four failed to amplify correctly and a further eight were either difficult to genotype correctly, or not polymorphic. Eight loci appeared to be acceptable and were selected as the final marker set for population genetic analysis. The final eight loci consisted of one dinucleotide, one tetranucleotide and six trinucleotides. After conducting a thorough PCR optimisation process, the optimum annealing temperatures were found to be: 58 °C for BCO_11, BCO_13, BCO_15 and BCO_17; 56 °C for BCO_03, BCO_10 and BCO_12; and 54 °C for BCO_08. During the development and optimisation stages the eight loci were found to be highly polymorphic, with the number of alleles ranging from 13 to 28.

Because an excess of homozygotes was detected at the BCO_10 locus, this suggested there were null alleles and this locus was discarded from further analyses. None of the other loci showed evidence of null alleles at any sample locations. The observed heterozygosity ranged from 0.727 to 0.894, and the expected heterozygosity ranged from 0.707 to 0.919 (Table 3.1). The loci BCO_08, BCO_13 and BCO_15 deviated from Hardy-Weinberg Equilibrium (HWE). None of the loci showed a significant level of linkage disequilibrium.

Table 3.4 Characteristics of seven microsatellite loci. N = sample size, A = number of alleles, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{IS} = inbreeding coefficient, AT = PCR annealing temperature, P_{HWE} = HWE *P*-value

Locus	Motif	Forward primer 5'-3'	Reverse primer 5'-3'	Size (bp)	AT	N	А	P _{HWE}	H _o	H _e	F _{IS}
BCO_03	(GT) ₁₆	TTG TCT GAT CTG GAG CGG AT	GCT GTG AGG AAA TCA AAC GC	217-281	56	268	28	0.393	0.894	0.919	0.022
BCO_08	(AGG) ₁₅	TTG GCA GAG CGG AGA TAA AC	TCT CTG CGT GGA CAG TTG AG	87-132	54	265	16	0.021	0.820	0.834	0.023
BCO_11	(ATG) ₁₆	CAG GAC GCC AGT ACA CCT TC	CCA GAG GAC ACG TGA GGA GA	218-266	58	268	13	0.485	0.727	0.707	-0.024
BCO_12	(CTG) ₂₁	ACG TCA ACA GCT CCG TCT CT	AGA GGT AAC ACG ACC CGT CA	92-149	56	268	16	0.808	0.826	0.829	-0.006
BCO_13	(GAG) ₁₄	CTG TTC GGG TTA ACG AGG TG	GCA TCT GGG TAC GGT TTC TG	178-223	58	261	15	0.021	0.787	0.806	0.030
BCO_15	(GCA) ₁₅	AGA ACG ACC AGG AGG ACC AT	CCT GAC CCT GAA CTC ACC AC	88-151	58	261	22	0.036	0.874	0.884	0.008
BCO_17	(ACAG) ₁₂	ATG GCA TCG TAC AGC TGG TT	TGG TCA CTA AAG GTC GTC CG	105-165	58	269	16	0.126	0.778	0.838	0.091

3.4 Discussion

This study reports the first microsatellite loci to be developed for *P. colias*, with the description of seven highly polymorphic markers. The 454 pyrosequencing system generated a pool of thousands of PALs, of which 20 were selected for further development. Of these 20 loci, 12 were discontinued during the PCR optimisation stage, and one indicated the presence of null alleles. The remaining seven loci were all highly polymorphic, with the number of alleles ranging from 13-28.

The genetic diversity revealed in this study was similar to that reported for another Penguipedidae species (*Parapercis cylindrica*), in which the authors described eight microsatellite DNA markers with 3-18 alleles per locus and $H_E 0.375-0.895$ (Smith-Keune et al., 2006). Smith-Keune et al. (2006) used a linker-ligated enrichment protocol (Hamilton et al., 1999) to isolate the *P. cylindrica* microsatellite loci, which requires a priori selection of microsatellite repeat sequences. The method used in the present study was not limited by that approach. The 454 sequence reads generated a library of thousands of PALs and future studies could use the same data set to discover and develop additional loci for *P. colias*.

In 2011, a review of contemporary microsatellite-genotyping practices (Guichoux et al., 2011) indicated that the multiplexing of loci, in which two or more loci are amplified in a single PCR, was still not common practise. Although a multiplex approach can significantly increase the efficiency of amplifying loci, particularly in studies with large datasets, this must be balanced with the time spent further optimising the PCR protocol for multiple primer pairs. In this thesis study the pool-plexing approach was implemented, in which loci were combined post-PCR for a more cost-effective and rapid genotyping method. Although they were not optimised for multiplex, due to time constraints, the annealing temperatures of these seven loci could be further optimised to allow multiplexing. This approach would increase the efficiency of the procedure for future studies.

Although the type and frequency of microsatellites distributed throughout genomes varies among taxa, dinucleotides are generally the most common type of locus, followed by tetraand then trinucleotides (Chistiakov et al., 2006). In *P. colias*, dinucleotides are also the most common type of microsatellite, representing 60.6% of identified repeat sequences, followed by trinucleotides 22.3%, tetranucleotides 15.5%, and pentanucleotides 1.6%. This pattern is consistent with observations reported from most other fish species (Chistiakov et al., 2006). Advances in high throughput sequencing have now eliminated the need for clone-based microsatellite isolation, and have significantly reduced the cost and time needed for the development of microsatellite loci for non-model species. Despite the recent popularity of 454 pyrosequencing for microsatellite loci isolation, in October 2013 Roche announced that it would discontinue the Life Sciences 454 sequencers by mid-2016. Although 454 pyrosequencing continues to be popular, due to its long average read lengths, considerable advances have been made in massively parallel sequencing technology, and many authors have suggested that other platforms could be equally or even more suitable for microsatellite isolation. For example, Castoe et al. (2012) used Illumina paired-end sequence data to identify potentially amplifiable loci in the Burmese python (Python molurus bivittatus), and compared this with 454 sequence data. The authors concluded that the greater number of reads obtained with Illumina, together with the significantly lower cost of sequencing, outweighed the benefits of longer average read lengths achieved with 454 pyrosequencing (Castoe et al., 2012). Recent studies using Illumina paired-end sequencing to develop loci for marine fish include the isolation of 84 loci for red snapper (Norrell et al., 2014) and the development of 16 microsatellite loci for the yellowfin tuna (Antoni et al., 2014). A study by Elliott (2014) comes to similar conclusions when comparing 454 sequencing to the Ion Torrent PGM semiconductor sequencer, and points out that PGM has the advantage of generating large quantities of data at a lower cost than 454 sequencing.

The development of seven highly polymorphic microsatellite loci provides a valuable tool for the investigation of the genetic diversity and fine-scale population structure of *P. colias*. The use of high throughput sequencing technologies, allowing a seq-to-SSR development approach, which allowed microsatellite DNA markers to be developed in a rapid and cost-effective manner, which otherwise could not have been achieved within the scope of this study.

Chapter Four

Population genetic structure of Blue cod (*Parapercis colias*) based on seven microsatellite markers

4.1 Introduction

Understanding the genetic structure of populations is important for the effective management of fishery stocks. By separately managing biologically distinct populations, fishery managers have a greater chance of attaining a maximum sustainable yield (MSY), in which harvesting pressure is balanced by within stock recruitment. Furthermore, through the incorporation of information related to the population structure of a fishery, managers can promote the recovery of depleted stocks and minimise the deleterious consequences of over-harvesting. However, estimating levels of population connectivity can be complex because it is driven by numerous factors, including the characteristics of the environment and life history traits. Genetic markers are a valuable tool for defining stocks, because gene flow represents the ongoing patterns of successful dispersal and reproduction between populations. When gene flow is restricted between two populations, evolutionary processes can work independently in each group, resulting in genetic differences. The differences can often be detected with genetic markers that sample multiple loci. If these markers are selectively neutral, it is reasonable to conclude that significant genetic differences between the two sites are the consequence of demographic uncoupling and isolation. However, if no genetic differences are detected, a firm conclusion about the levels of connectivity cannot be reached, because the number of migrants needed to genetically homogenise a population may be much lower than the level necessary for demographic connectivity (Ward, 2000).

Parapercis colias (blue cod) is a temperate reef fish and a member of the Penguipedidae family. It is an endemic species that is widely distributed throughout New Zealand, in rocky coastal waters down to depths of 150 metres. The species is more abundant south of the Cook Strait where it is a popular recreational game fish, particularly in the Marlborough Sounds, Otago and Stewart Island areas. Commercial catches of *P. colias* are predominantly taken from the area around Stewart Island, and in the vicinity of the Chatham Islands. Little is known about the early life history of *P. colias*, but it has been suggested that the eggs and

larvae are pelagic for approximately 10 days after spawning (Henderson, 2009; Rodgers and Wing, 2008; Smith, 2012). Adult individuals are generally sedentary and have been shown to exhibit limited dispersal, with mark and recapture studies showing that 60% of fish move less than 1 km (Cole et al., 2000). However, long range dispersal of adults can occur, with 156 km being the greatest distance an individual has been observed to move (Carbines and McKenzie, 2001).

Parapercis colias is currently managed within eight different quota management areas (QMA) (Chapter One, Figure 1.1), but over 95% of reported landings (2013) are taken from just three of these management areas (Ministry for Primary Industries, 2014). Considerable concern has been voiced over the localised depletion of some areas, the Marlborough Sounds region in particular, where declining stocks have been attributed to intense recreational fishing pressure. The Marlborough Sounds fishery was first closed in 2008 to assist the recovery of stocks, and since 2011 seasonal closures have been implemented to coincide with the spawning season.

A variety of different molecular markers are used to detect population structure. These include allozymes, mitochondrial DNA (mtDNA), microsatellite DNA and single nucleotide polymorphisms (SNPs). Mitochondrial and microsatellite DNA markers are the most popular markers, and are often useful in combination as they provide complementary information. Microsatellite loci have a high mutation rate and are highly polymorphic, which makes them more powerful for the detection of recent population structure, when compared to mtDNA. Microsatellite DNA markers have been shown to be beneficial in demonstrating fine-scale genetic structure that remained undetected by mtDNA markers in the case of Atlantic salmon (*Salmo salar L.*) (Tessier et al., 1995) and Atlantic cod (*Gadus morhua*) (Bentzen et al., 1996). As mtDNA is maternally inherited, and microsatellite DNA is bi-parentally inherited, significant differences in population structure between the two could result from differences between male and female life histories, such as sex bias dispersal (Patton et al., 1997).

One of the main goals of this research is to investigate whether New Zealand's mainland sites are genetically differentiated, indicating the presence of distinct biological populations. However, due to the high levels of heterozygosity at microsatellite loci, and the uncertainties surrounding some aspects of microsatellite evolution, the interpretation of differentiation through the use of microsatellites is highly debated, with no clear consensus on which differentiation statistic is best suited for microsatellite data (Balloux and Lugon-Moulin, 2002; Jost, 2008; Ryman et al., 2006; Verity and Nichols, 2014). Differences in the levels and biological significance of differentiation statistics need to be considered in the context of how each statistical model is applied to the data and what assumptions it makes. F_{ST} was first developed by Wright for the analysis of bi-allelic data, and can be used to measure heterozygote deficiency due to population subdivision. Nei's G_{ST} (1973), and the 'theta' parameter proposed by Weir and Cockerham (1984) (used here as F_{ST}), were developed for multi-allelic data and are commonly used for the analysis of microsatellite data. Both are generally considered to be equivalent (Whitlock, 2011). However, they have been criticised as being 'constrained', since F_{ST} should theoretically range between zero and one, whereas with highly variable loci the maximum value of F_{ST} (and G_{ST}) can be much lower than one. This has led to countless other statistics including R_{ST} (Slatkin 1995), Jost's D (Jost 2008) and G'_{ST} (Hendrick 2005), although each have their own disadvantages (Meirmans and Hedrick, 2011; Whitlock, 2011).

The aim of the present research is to determine the levels of genetic diversity and fine scale population structure of *P. colias* in New Zealand. Data here has been compared with results from the previous mtDNA study (Chapter Two). This makes it possible to determine whether or not gene flow has been sufficiently restricted to create differentiated populations, which were not detected by mtDNA analysis. From the management perspective, this will help to further elucidate the isolation-by-distance pattern, allowing knowledge of the genetic structure of *P. colias* to be implemented into the management program.

4.2 Methods

Sample collection

Sample collection took place during February 2010 and June 2011, in seven locations around New Zealand (Figure 4.1): Bay of Plenty (BOP), Chatham Islands (CHA), Wellington (WEL), Pelorus sound (PEL), Kaikoura (KAI), Otago (OTA) and Puysegur Point (PUY). A total of 274 fin clips were collected during fishing operations on commercial vessels, fishery surveys, or using hook and line (Hayden Smith). These were stored in 85% ethanol. The sex and age of the fish in the sample were randomised. Sample sizes for each location are shown in Figure 4.1. Sampling sites north of the Cook Strait were less common and only smaller sizes were obtained due to the patchy distribution of *P. colias* in northern areas, rather than a lack of sampling effort.



Figure 4.1 Sampling locations and sizes (n) for the seven sample sites throughout New Zealand.

Genotyping

Genomic DNA was extracted from samples using a standard phenol-chloroform protocol (Sambrook et al., 1989). The PCR primers for the seven microsatellite DNA loci were described in Chapter Three. All loci were amplified separately in PCRs consisting of 67 mM Tris-HCL pH 8.8, 16 mM (NH₄)₂SO₄, 2.0 mM MgCl₂, 200 μ M of each dNTP, 0.1 μ M forward primer, 0.5 μ M of reverse primer, 0.5 μ M fluorescently-labelled M13 primer, 0.6 μ g/ μ L Bovine Serum Albumin (BSA), 1 unit of BIOTAQTM DNA Polymerase (Bioline), and 1 μ L of template DNA in 10 μ L volumes. Thermocycling conditions were 95 °C for 5 minutes, 94 °C for 30 seconds, annealing temperature (see Table 4.1) for 40 seconds, 72 °C for 40 seconds for 30 cycles, then a second set of 10 cycles at 94 °C for 30 seconds, 53 °C for 40 seconds, and 72 °C for 40 seconds, followed by a final extension at 72 °C for 10 minutes. The second set of cycling conditions was performed to facilitate annealing of the M13 primer with either a PET, VIC or FAM fluorescent label, according to the method outlined in Schuelke (2000). The amplified products were confirmed by electrophoresis in 2% agarose gel, and visualisation using ethidum bromide under a UV-light source. Table 4.1 lists the PCR primer sequence and annealing temperature for each locus.

The PCR products were pool-plexed into two combinations of loci, based on the recommendation provided by MULTIPLEX MANAGER 1.2 (Holleley and Geerts, 2009). The pooled amplicons were then genotyped using an ABI3730 Genetic Analyzer (Massey University Genome Service). The GeneScanTM 500 LIZTM Size Standard was used for fragment sizing, and the PEAK SCANNER Software v2 (Applied Biosystems) was used to determine the allele sizes. The Excel macro AUTOBIN was used to determine the 'bin' size for each allele.

Locus	Repeat unit	Size (bp)	5' Forward Primer 3' 5' Reverse Primer 3'	Annealing temp (⁰ C)	M13 primer
BCO_03	(GT) ₁₆	217-281	TTG TCT GAT CTG GAG CGG AT	56	PET
			GCT GTG AGG AAA TCA AAC GC		
BCO_08	(AGG) ₁₅	87-132	TTG GCA GAG CGG AGA TAA AC	54	FAM
			TCT CTG CGT GGA CAG TTG AG		
BCO_11	$(ATG)_{16}$	218-266	CAG GAC GCC AGT ACA CCT TC	58	PET
			CCA GAG GAC ACG TGA GGA GA		
BCO_12	(CTG) ₂₁	92-149	ACG TCA ACA GCT CCG TCT CT	56	VIC
			AGA GGT AAC ACG ACC CGT CA		
BCO_13	$(GAG)_{14}$	178-223	CTG TTC GGG TTA ACG AGG TG	58	VIC
			GCA TCT GGG TAC GGT TTC TG		
BCO_15	(GCA) ₁₅	88-151	AGA ACG ACC AGG AGG ACC AT	58	PET
			CCT GAC CCT GAA CTC ACC AC		
BCO_17	$(ACAG)_{12}$	105-165	ATG GCA TCG TAC AGC TGG TT	58	VIC
			TGG TCA CTA AAG GTC GTC CG		

Table 4.5 Primer sequence and PCR conditions of seven microsatellite loci

Genetic diversity

The presence of null alleles and score errors was assessed using MICRO-CHECKER 2.2.3 (van Oosterhout et al., 2004) with 1,000 iterations/permutations. Deviations from the Hardy Weinberg equilibrium (HWE) were tested for, across loci and populations, using the exact test with a 10^6 Markov chain length and 10^5 de-memorization steps implemented in ARLEQUIN 3.5 (Excoffier et al., 2005). A pairwise test of linkage disequilibrium (LD) with 10,000 permutations (likelihood ratio test) was also carried out in ARLEQUIN 3.5. Significance values were estimated at the 0.05 level after Sequential Bonferroni corrections, for the HWE and LD tests.

The diversity indices calculated in ARLEQUIN 3.5 included observed and expected heterozygosities and numbers of alleles per locus, by population. The inbreeding coefficient (F_{IS}), allelic richness, and number of private alleles were estimated using FSTAT 2.9.3.2 (Goudet, 1995). Allele frequency distributions were generated for each population and loci using GENALEX 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012).

Allele rarefaction curves were calculated for each locus using the PopGenKit package in R 3.1.1 (Ihaka and Gentleman, 1996; Rioux Paquette, 2011). A separate curve was made for each population and locus. The 'discovered' alleles were plotted against cumulative sample sizes and used to determine whether the sample sizes were sufficiently large to accurately

estimate allelic diversity. To test the assumption of loci neutrality, the F_{ST} outlier detection method was performed using LOSITAN (Antao et al., 2008). This analysis was used to identify candidates for selection by plotting the relationship between F_{ST} and heterozygosity (Beaumont and Nichols, 1996). Loci under selection, or linked to genes under selection, appear as outliers when compared to loci influenced by drift and gene flow. The outlier analysis was performed with 100,000 simulations at a confidence interval of 0.995, using the stepwise (SMM) and infinite allele (IAM) mutation models.

Population Structure

The levels of genetic differentiation between sites were estimated using several different approaches. (1) Weir and Cockerham's (1984) $F_{ST}(\theta)$ was used, because it is designed for the analysis of multi-allelic and multi-loci data. F_{ST} values were estimated without assuming Hardy Weinberg Equilibrium, with 1,000 iterations and 95% confidence intervals, with Bonferroni correction in FSTAT 2.9.3.2 (Goudet, 1995). The use of F-statistics for the analysis of microsatellite data has been criticized, because although F_{ST} should theoretically range between zero and one, in the case of highly variable loci the maximum value of F_{ST} is much lower than one (Jost, 2008; Meirmans and Hedrick, 2011; Verity and Nichols, 2014). Hedrick (2005) proposed a standardized analogue G'ST, adapted from Nei's (1973) GST, which remains proportional to the maximum possible level of differentiation. (2) Pairwise G_{ST} and G'_{ST} were estimated using the population genetic analysis programs FSTAT 2.9.3.2 and GENODIVE (Meirmans and Van Tienderen, 2004). (3) Slatkin's R_{ST} is an F_{ST} analogue, which has been adapted for the analysis of microsatellite DNA loci, and assumes that loci evolve according to a step-wise mutation model (Slatkin, 1995). The significance of the R_{ST} values was estimated using a permutation test with 10,000 replicates and a 0.05 significance level in ARELQUIN 3.5.1. (4) For comparison, Jost's D estimator, D_{ST} (Jost, 2008) was measured with 1,000 bootstrap replicates using SMOGD 1.2.5 (Crawford, 2010). In addition to accounting for limited maximum differentiation, Jost's D is measured using the effective number of alleles rather than heterozygosity (Jost, 2008).

A test for a pattern of isolation-by-distance (IBD) was conducted using linearised F_{ST} values plotted against the shortest geographical distance (kilometres) by sea between all pairs of sample locations, excluding the Chatham Islands (CHA). The level of significance for this pattern was determined using a Mantel test with 10,000 permutations in ARELQUIN 3.5.1.

The levels of genetic diversity were estimated among and within populations using an Analysis of Molecular Variance (AMOVA) implemented in ARLEQUIN 3.5. The following six different population groupings were tested, each with 10,000 permutations: (1) two regions, mainland New Zealand (NZ) sites separated from the Chatham Islands (CHA); (2) three regions, North Island (NI) and South Island (SI) of New Zealand separated by the Cook Strait, and the Chatham Islands sampling sites; (3) two regions, North Island and Chatham Islands, excluding the South Island; (4) two regions, South Island and Chatham Islands, excluding the North Island; (5) two regions, sites separated into North and South biogeographic regions with respect to the 42° latitude line (the Chatham Islands were excluded from this analysis); and (6) two regions, the North Island and the South Island, excluding the Chatham Islands. These groupings were based on potential genetic breaks identified in New Zealand's marine environment (Apte and Gardner, 2002; Ross et al., 2009).

A Bayesian clustering analysis was performed using STRUCTURE 2.3.4 (Pritchard et al., 2000). This analysis infers population structure by probabilistically assigning individuals to one or more population clusters (K), in which each cluster is characterised by a set of allele frequencies. The simulations were repeated 10 times for each value of K, representing the number of populations, which was allowed to range from 1 to 7. Markov chain Monte Carlo (MCMC) simulations were run for a total of 10^{10} iterations after a burn-in period of 10^5 steps, under both admixture and non-admixture models with the 'locprior' function implemented. Delta K (Δ K) measured in STRUCTURE HARVESTER 0.6.94 (Earl and vonHoldt, 2012) was used to estimate the value of K most likely to reflect the population's structure (Evanno et al., 2005).

Principal Component Analysis (PCA) was performed using PCA-GEN 1.2 (Goudet, 1999), to examine the relationships among populations based on pairwise F_{ST} values. This analysis was run with 15,000 randomisations using: A) all sites, and B) all sites excluding CHA. A Factorial Component Analysis (FCA) implemented in GENETIX (Belkhir et al., 2000) was conducted to examine differences among all individuals. Two graphs were constructed to examine three main axes.

Mutation-drift Equilibrium and Bottleneck

Three approaches were used to determine whether the populations had undergone a recent reduction in effective population size (i.e. a genetic bottleneck). When a population size is reduced, the strength of genetic drift increases and alleles at low frequencies are typically eliminated earlier than alleles at higher frequencies. The first method compared the proportion of alleles at low frequencies to the expected L-shaped distribution of allele frequencies predicted in the case of a mutation-drift equilibrium (Garza and Williamson, 2001). Allele frequency distribution plots were constructed for each population using the excel add-on GENALEX (Peakall and Smouse, 2006; 2012). Each population was tested for a deviation from an L-shaped distribution using a mode-shift test implemented in BOTTLENECK 1.2.02 (Piry et al., 1999). Secondly, a Wilcoxon sign-rank test was used to assess heterozygosity excess in BOTTLENECK 1.2.02. The distribution of expected heterozygosity, based on the observed number of alleles (k), was simulated using three different mutation models: (i) the Stepwise Mutation Model (SMM), (ii) the Infinite Alleles Model (IAM), and (iii) the Two-Phase Model (TPM). Analyses were performed with 1000 replications. The TPM was set to 95% single-step mutations and 5% multi-step mutations (Piry et al., 1999). Thirdly, M-ratios were calculated for each population using the software M_P_VAL (Garza and Williamson, 2001). This analysis is designed to compare the observed values of M (the ratio of the total number of alleles to the range in allele size) with the critical M values, (M_C), calculated for simulated populations at mutation-drift equilibrium. The input parameters were chosen in accordance with the recommendations of Garza and Williamson (2001), and theta $\theta = 4N_{ef}\mu$ was calculated for each population in ARLEQUIN 3.5.

4.3 Results

Genetic diversity

Seven microsatellite loci were used to genotype 274 individuals from seven New Zealand sample locations. Three individuals were removed from the data set because they had missing data at more than three loci. The final data set comprised a total sample of 271 individuals. The test for homozygote excess showed no evidence of null alleles, large allelic drop out, or scoring errors at the 95% confidence interval. The loci data set had less than 5% missing data. The summary statistics for each population are presented in Table 4.2. All of the loci were highly polymorphic, ranging from 13 alleles at the BCO_11 locus to 28 alleles at the BCO03 locus. The allelic richness per site ranged between 7.38 for BCO_11 and 16.46 for BCO_03. See Figure 4.3-4.6 for Allele frequency distributions.

Only one of the 49 location-by-locus tests (CH at BCO_12) shows a significant departure from HWE after correction for multiple tests using sequential Bonferroni correction (P < 0.007). The Chatham Islands is also the only population to show significant LD between the BCO_12 and BCO_13 loci after correction for multiple tests (P<0.0024). As none of the loci are found to be candidates for either positive or balancing selection, the assumption of neutrality was not rejected.

The allele rarefaction curves for each locus showed that the loci appear to reach a plateau when cumulative allelic richness was plotted against the number of genotyped individuals (Figure 4.2). This trend was observed for the population with the smallest sample size (BOP n=22), suggesting that the sample size at all of the sites was sufficiently large to represent the actual levels of genetic diversity in the total population. The observed heterozygosity (H_o), averaged across loci, ranged from 0.775 (BOP) to 0.869 (WEL). The expected heterozygosity ranged from 0.804 (CHA) to 0.851 (WEL). Private alleles were present at all sites, with six private alleles at the BCO_08 locus.

Sample site		Ν	H _o	H _e	A _r	F _{IS}	PA
Bay of Plenty	BOP	22	0.775	0.826	13.27	0.064	2
Chatham Islands	CHA	45	0.815	0.804	9.32	-0.013	3
Kaikoura	KAI	45	0.815	0.838	10.61	0.028	3
Puysegur Point	PUY	39	0.805	0.833	10.67	0.034	2
Pelorus Sound	PEL	30	0.836	0.833	10.64	-0.004	4
Otago	OTA	48	0.795	0.831	10.01	0.044	3
Wellington	WEL	42	0.869	0.851	10.76	-0.021	4
Total/Average		271	0.816	0.831	10.55	0.019	3

Table 4.6. Summary statistics for seven microsatellite loci. Sample size (N), Observed heterozygosity (H_o), expected heterozygosity (H_e), allelic richness (A_r), fixation index (Fis) and number of private alleles (PA).

Allele discovery curve for locus BCO_03 30 25 - BOP Number of alleles 20 ----- CHA 15 KAI **---** PUY 10 $- \cdot - \cdot PEL$ - OTA 5 -- Wel 0 10 20 50 0 30 40 Number of individuals genotyped

Figure 4.2 Rarefaction Curve for the most polymorphic locus BCO_03. See appendix 5 for the six other loci.



Allele size (base pair)

Figure 4.3 Allelic frequency distribution for the microsatellite loci BCO_12 and BCO_15 for all sample sites. Allele sizes are given in base pairs of PCR products.



Figure 4.4 Allelic frequency distribution for the microsatellite loci BCO_08 and BCO_17 for all sample sites. Allele sizes are given in base pairs of PCR products.



Figure 4.5 Allelic frequency distribution for the microsatellite loci BCO_13 and BCO_11 for all sample sites. Allele sizes are given in base pairs of PCR products.



Figure 3. Allelic frequency distribution for the microsatellite locus BCO_03 for all sample sites. Allele sizes are given in base pairs of PCR products.

Population structure

The pairwise F_{ST} shows that there was a significant level of differentiation between CHA and all other sites, apart from BOP, but there was no significant differentiation observed among the other mainland sites (Table 4.3). The pairwise R_{ST} values showed a significant level of differentiation between the same pairs of sample locations as those observed in the F_{ST} analysis. The only exception in the R_{ST} analyses was the significant level of differentiation between WEL, when compared to BOP and OTA (Table 4.5). A similar pattern is observed in Jost's D values, with higher values for the CHA pairs when compared to the other Fstatistics. G_{ST} analysis indicated the highest number of significantly differentiated population pairs, but also has the lowest values. The G_{ST} values revealed a significant level of differentiation between CHA and all other sites, between BOP and all sites apart from PUY and OTA, between KAI and OTA and WEL, as well as differentiation between OTA and WEL. The AMOVA analysis showed little evidence of population differentiation among the regional groupings, with the highest variance found in the SI/CHA grouping (Var=0.024, %Var=0.85 F_{CT}=0.007 *P*=0.199) followed by NZ/CHA (Var=0.018, %Var = 0.65, F_{CT}=0.007 P=0.142). For each grouping, more than 98% of the genetic variation was found to be within individuals (Table 4.6). The Mantel test used to examine the presence of an isolation-bydistance pattern, which did not include samples from the Chatham Islands, did not show a significant correlation between genetic and geographic distance (R^2 =0.002, *P*=0.496).

The Bayesian cluster analysis implemented in STRUCTURE 2.3.4 revealed that two genetic clusters K=2 best describe the allele frequency distribution, based on delta K (admixture Δ K=7.98, non-admixture Δ K=79.50, Appendix 2). Under the assumption of K=2, individuals from the Chatham Islands were assigned mainly to one genetic cluster, with individuals from the rest of New Zealand assigned mainly to the other cluster (Figure 4.8). Due to the possible presence of a population split, STRUCTURE was re-run without the Chatham Islands sample; the results of this sequential analysis revealed no further evidence of population structuring. The PCA results suggested that CHA was differentiated from the other NZ sample sites along the first axis (PCA 1: F_{ST} =0.006, 34.01% inertia, *P*=0.017, Figure 4.7). The second axis (PCA 2) revealed separation of the other sites, although this difference is not significant (PCA 2: F_{ST} =0.004, 21.94% inertia, *P*=0.339). The FCA revealed one main cluster with a few outliers, and no discernible differences between populations (Figure 4.9). Overall, the three main FCA axes explain only 7.15% of the observed variations.

	BOP	CHA	KAI	PEL	PUY	OTA	WEL
BOP		0.005	0.062	0.252	0.671	0.276	0.057
CHA	0.008		0.002	0.002	0.002	0.002	0.002
KAI	0.007	0.014		0.031	0.462	0.131	0.219
PEL	0.003	0.019	0.002		0.302	0.012	0.340
PUY	-0.001	0.013	0.000	-0.001		0.238	0.481
OTA	-0.002	0.017	0.005	0.003	0.000		0.005
WEL	0.008	0.012	0.003	0.002	0.001	0.006	

Table 4.7 Pairwise F_{ST} (θ) values below the diagonal, *P*-values above the diagonal with 5% significance after Bonferroni correction P < 0.003.

Table 4.8 Pairwise G_{ST} with significant values (P < 0.001) in bold, Hendricks G'_{ST} above the diagonal

	BOP	CHA	KAI	PEL	PUY	OTA	WEL
BOP		0.039	0.040	0.018	-0.007	-0.012	0.046
CHA	0.004		0.078	0.104	0.070	0.096	0.071
KAI	0.003	0.007		0.010	0.003	0.032	0.022
PEL	0.002	0.009	0.001		-0.004	0.017	0.011
PUY	-0.001	0.006	0.000	0.000		0.002	0.006
OTA	-0.001	0.009	0.003	0.001	0.000		0.040
WEL	0.004	0.006	0.002	0.001	0.000	0.003	

Table 4.9 Pairwise R_{ST} values below the diagonal, significant values P < 0.05 in bold, P < 0.01 with an asterisk. Jost's D_{est} values above the diagonal.

	BOP	CHA	KAI	PEL	PUY	OTA	WEL
BOP		0.016	0.007	0.000	0.000	0.000	0.011
CHA	0.006		0.018	0.044	0.036	0.039	0.022
KAI	0.007	0.010*		0.001	0.001	0.007	0.003
PEL	0.003	0.015*	0.002		0.000	0.001	0.000
PUY	-0.001	0.008*	0.001	-0.001		0.000	0.000
OTA	-0.001	0.014*	0.005	0.003	0.000		0.009
WEL	0.007	0.008*	0.003	0.001	0.000	0.006	

Table 4.10 Analysis of Molecular Variance (AMOVA) results for 6 different variations: (1) two regions, mainland New Zealand (NZ) sites separated from the Chatham Islands (CHA); (2) three regions, North Island (NI) and South Island (SI) New Zealand separated at the Cook Strait, and Chatham Islands sampling sites; (3) two regions, North Island and Chatham Islands, excluding the South Island; (4) two regions, South Island and Chatham Islands; (5) two regions, sites organised into North and South biogeographic regions with respect to the 42° latitude line. The Chatham Islands are excluded from this analysis; and (6) two regions, North Island and South Island, excluding the Chatham Islands.

Regions		NZ/CHA	NI/SI/CHA	NI/CHA	SI/CHA	N42/S42	NI/SI
Among groups	Df	1	2	1	1	1	1
	Var	0.018	0.008	-0.004	0.024	-0.001	-0.002
	%Var	0.65	0.27	-0.16	0.85	-0.05	-0.08
	F _{CT}	0.007	0.003	-0.002	0.008	-0.001	-0.001
	Р	0.142	0.151	1.00	0.199	0.503	0.736
Among	Df	5	4	1	3	4	4
populations within	Var	0.008	0.009	0.024	0.006	0.008	0.008
groups	%Var	0.28	0.32	0.88	0.22	0.26	0.27
	F _{SC}	0.003	0.003	0.009	0.002	0.003	0.003
	Р	0.049	0.041	0.041	0.154	0.048	0.044
Among	Df	264	264	106	202	220	220
individuals within	Var	-0.005	-0.006	-0.091	0.001	0.029	0.029
populations	%Var	-0.21	-0.21	-3.33	0.05	1.00	1.00
	F _{IS}	-0.002	-0.002	-0.034	0.001	0.010	0.010
	Р	0.590	0.588	0.982	0.500	0.194	0.191
Within individuals	Df	271	271	109	207	226	226
	Var	2.806	2.806	2.803	2.790	2.827	2.827
	%Var	99.27	99.62	102.61	98.88	98.79	98.81
	F _{IT}	0.007	0.004	-0.026	0.011	0.012	0.012
	Р	0.420	0.437	0.968	0.336	0.140	0.140



Figure 4.7 Principal component analysis: (A) all sites, (B) sites excluding the Chatham Islands



Figure 4.8 Results for Bayesian structure analysis of P. colias assuming k=2 genetic clusters. Each line represents an individual and plot is ordered by population.



Figure 4.9 Factorial component analysis (FCA) showing Axis 1, 2 and 3

Mutation-drift Equilibrium and Bottleneck

The mode-shift test showed that the allele frequency spectrum did not deviate significantly from an L-shaped distribution, which is the expected pattern for a population in mutationdrift equilibrium (see Appendix 4 for rare allele distribution graphs). In addition, the M-ratio test did not detect any reduction in effective population size, since the M-ratio values for each population were higher than the M critical value ($M_C = 0.733$) (Table 4.7). However, the Wilcoxon sign-rank test suggests that the BOP, PUY and PEL populations are characterized by heterozygosity excess under the TPM, which may be the consequence of a recent contraction in population size (Table 4.7).

	Wilcoxon sign-rank test			M-ratio		
	IAM	TPM	SMM	М	%	$\Theta_{ m (H)}$
BOP	0.375	0.023	0.023	0.781	28.52	2.979
CHA	0.039	0.078	0.016	0.830	43.54	2.677
KAI	0.016	0.688	0.016	0.793	28.41	3.188
PUY	0.039	0.039	0.008	0.835	49.84	3.095
PEL	0.016	0.039	0.016	0.789	30.20	3.090
OTA	0.008	0.078	0.008	0.774	21.39	3.068
WEL	0.008	0.469	0.078	0.773	22.60	3.439
Total				0.907	82.52	3.077

Table 4.11 Results from the Wilcoxon bottleneck test and M-ratio analysis for each population.

4.4 Discussion

The results of this study showed that there was genetic differentiation between the Chatham Islands and mainland sites, which is consistent with the findings of the mtDNA analysis (Chapter Two). Interpretation of the results for the mainland sites is more complex, since the pairwise F-statistics (F_{ST} , G_{ST} and R_{ST}) between sites was generally low, yet there is evidence of significant differentiation between some sites. However, PCA and Bayesian Structure analysis did not support the suggestion of significant genetic differences between the mainland sites. The pattern of weak population structure along the mainland is consistent with the previous study using mtDNA control region sequences, which found no evidence of genetic differentiation, although a pattern of isolation-by-distance was detected.

Genetic diversity

The observed levels of genetic variation were high and similar across all sites, with a mean expected heterozygosity (H_e) of 0.83, mean number of alleles per locus (A) of 18.0, and an average allelic richness per site (A_r) of 10.55. Marine species typically show higher levels of heterozygosity and alleles per locus (He=0.79 and A=20.6) than freshwater (He=0.46 and A=7.5) and non-piscine animals (H=0.58 and A=7.1) (DeWoody and Avise, 2000). This has been attributed to higher levels of fecundity and a more 'open' environment, leading to larger population sizes (DeWoody and Avise, 2000). Population decline due to harvesting can reduce genetic variation within populations. Maintenance of genetic variation is important, because a loss of variation not only increases the risk of reduced individual fitness, which can suppress the productivity of populations, but it also reduces a population's ability to evolve in a changing environment (Allendorf et al., 2008).

Due to the large population sizes of many marine species, fishery stocks considered to be "collapsed" may still comprise several million individuals (Hauser et al., 2002). It is possible that even heavily overfished populations would fail to show a significant loss in genetic diversity. However, due to factors such as variable reproductive success and selective sweeps, the effective population size of marine fish may be disproportionally smaller than their census size. An example of this difference has been observed in New Zealand snapper (*Pagrus auratus*), where temporal measurements indicated a loss of microsatellite diversity over time, reflecting low effective population size (Hauser et al., 2002). As stock numbers may not reflect reproductive success, measurements of genetic diversity can be extremely important for the identification of populations in decline.
Population structure

The findings of this thesis study support the hypothesis that there are two distinct genetic groupings of P. colias; the Chatham Islands and the sampled locations around mainland New Zealand. The finding of differentiation is consistent with the previous mtDNA analysis, indicating a significant genetic break between the island groups. Microsatellite pairwise Fstatistics (FST, RST and GST), Bayesian Structure analysis, and Principal Component Analysis all show that P. colias sampled from the Chatham Islands are genetically differentiated from mainland New Zealand sites. The only exception was between the Bay of Plenty and the Chatham Island sites, where G_{ST} does support differentiation, whereas F_{ST} and R_{ST} do not. The Bay of Plenty appears to be the most genetically similar site to the Chatham Islands, although the site has a smaller sample size than the other mainland sites, which could have caused bias in the analysis. Oceanographic currents can influence the patterns of connectivity of populations. However, the complexity of New Zealand's coastal current system makes it difficult to identify the expected locations that might be more similar to the Chatham's site. Furthermore, mtDNA analysis suggested that the Pelorus sound site was the least genetically distinct from the Chatham Islands site. Given the geographical distance between the Chatham Islands and the mainland islands of New Zealand, this genetic differentiation is unsurprising. Adult P. colias are sedentary, benthic, and are predominantly found down to a depth of only 150m, such that the mtDNA and microsatellite DNA results support the hypothesis that adult migration between the Chatham Islands and the mainland islands is highly infrequent.

Population structure – mainland sites

Differentiation estimates for mainland sites were not always consistent, and there did appear to be some difference that arose depending on which statistical model is applied. Whereas F_{ST} (θ) indicated no pairwise differentiation, R_{ST} and G_{ST} indicated a low but significant differentiation between Wellington and the Bay of Plenty, Kaikoura and Otago, and between Otago and Wellington. G_{ST} indicated differentiation between Kaikoura and Wellington, and between the Bay of Plenty and Kaikoura and Pelorus sound. Although differentiation estimates were low, this was probably due to the 'constrained' nature of the test statistics, which have been shown to underestimate differentiation when the mutation rate is high and migration is low (Whitlock, 2011), and when variation within subpopulations is high (Hedrick, 1999). This is supported by the high levels of G'_{ST} and Jost's D which are less constrained than G_{ST} . A finding of low genetic differentiation is common among marine species, and often at greater geographical scales than those used in this study (Waples, 1998; Ward et al., 1994). For example, low but significant differentiation was found across the southern hemisphere in the deep-sea teleost *Hoplostethus atlanticus* (orange roughy), with no divergence found between Australia and New Zealand, despite otolith analyses previously indicating differentiation (Varela et al., 2013). More relevant to the results reported here for *P. colias*, is the inshore teleost species *Pagrus auratus* (Australasian snapper) exhibited low levels of significant differentiation in some sample sites but was otherwise largely panmictic (Ashton, 2013).

The question of how to interpret the low but statistically significant levels of genetic differentiation has been investigated by Knutsen et al. (2011), in a case study of *Gadus morhua* (Atlantic cod). This study used temporally replicated sampling over a 10 year period, coupled with extensive mark recapture studies, to conclude that low but statistically significant F_{ST} (average of 0.0037) was biologically relevant for Atlantic cod (Knutsen et al., 2011). Although it would be misleading to draw direct comparisons between the Atlantic cod study and the results reported here for *P. colias*, the finding highlights the value of genetic differentiation estimates for microsatellite loci, and the fact that weak levels of differentiation are still likely to represent separate biological populations (Knutsen et al., 2011).

Isolation-by-distance

A pattern of isolation-by-distance is a common type of population structure, since individual dispersal capabilities are often limited relative to the overall species' distribution range. This often results in a positive correlation between genetic distance and geographic distance. However, the strength of the association may be strong in some areas and weak in others, due to the influence of local conditions. Mitochondrial control region sequence data suggested that there was a pattern of isolation-by-distance along the mainland New Zealand sites (Chapter Two). This initial finding suggested that a higher resolution genetic marker would find more substantial levels of population structure and this was the predominant motivation for investigation into population structure using nuclear microsatellite loci. No similar pattern was found using seven microsatellite loci. It is difficult to determine whether this finding reflected a biological pattern or the statistical limitation of the seven loci and few sample locations used in this study.

One explanation could be that the isolation-by-distance pattern is driven by only some of the sites, and that those sites were not sampled with microsatellite loci. To test for this, using the mtDNA data from chapter two, genetic distance was plotted against geographic distance using only those sample sites that were also used for microsatellite analysis, thus allowing a more direct comparison to be made between the two markers. The resulting correlation between mtDNA genetic distance and geographic distance between the six mainland sites (Bay of Plenty, Wellington, Pelorus Sound, Kaikoura, Puysegur Point and Stewart Island) is found to be even stronger (R^2 =0.408), suggesting that the discrepancy between marker types is not caused by the use of different sampled sites.

Mutation-drift equilibrium and Bottleneck

A large reduction in effective population size, termed a bottleneck effect, can lead to a loss of genetic variation. As the strength of genetic drift is stronger in smaller populations, rare alleles are more readily eliminated by genetic drift than alleles at higher frequencies. This can lead to various changes, such as an increased rate of inbreeding depression, increased expression of deleterious alleles, as well as a reduction in the adaptive potential of populations (Allendorf and Luikart, 2007). The mode-shift test and the M-ratio test, both of which are based on allele frequencies, do not show any evidence of a population size reduction, indicating that the populations are likely to be in mutation-drift equilibrium. This is consistent with the mtDNA analysis, which detected that *P. colias* underwent population expansion in the past. Heterozygote excess was detected at a number of sites, which might indicate recent population contractions. However, heterozygosity levels are less sensitive to a bottleneck effect than allelic diversity (Allendorf and Luikart, 2007).

Management implications

Genetic markers reflect patterns of gene flow over thousands of years, a time scale that has little relevance for contemporary management issues. Therefore, demographic history inferred from genetic markers is unlikely to reflect any recent population decline due to overfishing or other recent environmental factors. Low levels of significant genetic differentiation between sites should be carefully considered, as even low levels of demographic connectivity can lead to genetic homogeneity. This has important management implications, because the level of migration necessary to genetically homogenise stock structure could be much smaller than the level that would be beneficial for the rebuilding of depleted stocks (Waples, 1998).

The genetic results reported here reject the null hypothesis of panmictic population structure of *P. colias* populations in the New Zealand EEZ. *P. colias* is currently managed as eight separate QMAs, and there is no evidence to suggest that this precautionary approach should be disregarded. The Chatham Islands site is significantly differentiated from the rest of mainland New Zealand and should continue to be managed independently. The QMA BCO3 encompasses an area covering the east coast of the South Island, including the Otago and Kaikoura sample sites. As low but significant differentiation was found between these two sites, this QMA is likely to contain biologically distinct populations. Management boundaries may need to be re-evaluated to reflect these findings. However, the observed levels of differentiation are relatively weak, and it could be advantageous to gather more information about the species' life history characteristics, and conduct a genetic study that involved a larger sample of microsatellite loci with more sample sizes, before considering any revisions of the *P. colias* management boundaries.

Chapter Five

General Discussion and Implications for Fisheries Management

General

The primary goal of this research was to advance current knowledge of the *Parapercis colias* population structure using samples collected from sites around the New Zealand coastline, including the Chatham Islands. Microsatellite DNA and mtDNA markers were used to determine the levels of genetic variation within sites, as well as the level of differentiation between sites. The finding of genetically differentiated sites most likely reflects demographically uncoupled populations, and that information can be incorporated into stock assessment models and improve the management of the *P. colias* fishery. This study made use of 454 pyrosequencing technology to develop the first set of microsatellite DNA markers for *P. colias*. It is expected that the new loci will become an important resource for future studies of the fishery.

Management overview

Parapercis colias is an important inshore fishery, with significant commercial and recreational catches being taken from around the South Island and the Chatham Islands. This species is more abundant south of the Cook Strait and is considered to be the most widely sought-after finfish in the South Island (Ministry for Primary Industries, 2014). The total allowable commercial catch (TACC) was set at 2,680 tonnes per annum in 2001, and was lowered to 2,332 tonnes in 2012. The major commercial fisheries lie off Southland and the Chatham Islands, whereas extensive recreational fishing is practiced in the Marlborough Sounds. *P. colias* is currently managed in eight distinct quota management areas (QMA), by the Ministry for Primary Industries. Catch rates are low, and are considered to be sustainable in the three most northern management areas (BCO1, BCO2 and BCO10). Management concerns have focused on the remaining five areas, which are considered to comprise the main *P. colias* stocks (Figure 1.1; Ministry for Primary Industries, 2014). BCO5, the largest and southernmost management area, is allocated over half of the total allowable commercial catch (TACC) and reports the highest volumes of commercial landings. Concern has been

expressed over the sustainability of this fishery, as the biomass and catch per unit effort (CPUE) have recently been in decline (Ministry for Primary Industries, 2014).

The Marlborough Sounds are also of particular concern for fishery managers. As a consequence of past declines in stock attributed to recreational fishing pressure, the area is now managed as a separate area within BCO7. Recreational harvests have been managed predominantly through size and daily bag limits, with the Marlborough Sounds currently implementing a 30-35 cm size restriction and a bag limit of 2 fish per person per day (Ministry for Primary Industries, 2014).

Management suggestions

For the sustainable management of fishery stocks, biologically distinct populations should be managed separately (Carvalho and Hauser, 1994). To avoid over-harvesting, fishing induced mortality should be balanced by within-stock recruitment, and this is most effectively achieved at a level where demographically isolated populations are managed separately. If two distinct populations are treated as one management unit, unequal harvesting pressure can exacerbate stock depletion and increase the deleterious effects of harvesting pressure. The latter can include a loss of genetic variation, an altered population substructure, and may possibly lead to selective genetic changes (Allendorf et al., 2008).

Previous studies have suggested that *P. colias* could consist of several biologically distinct populations and which could leave them vulnerable to localised depletion if they were managed incorrectly. Tagging studies have consistently indicated that *P. colias* is sedentary, with 60-65% of individuals moving less than 1 km (Carbines, 1999; Carbines and McKenzie, 2001; Cole et al., 2000). However, long-range dispersal of adults can occur, with 156 km being the greatest distance an individual has been observed to move (Carbines and McKenzie, 2001). Stable isotope and otolith elemental studies of *P. colias* in Fiordland have both suggested limited movements of *P. colias*, consistent with tagging studies (Beer et al., 2011; Rodgers and Wing, 2008). Using mtDNA analysis, Smith (2012) found that the Chatham Islands site is genetically distinct from mainland New Zealand sites. In Chapter Two, this outcome is re-analysed with the inclusion of an additional sample site, and reveals the same pattern. Microsatellite marker analysis (Chapter Four) has also been successful in identifying two distinct populations, consisting in those located in the Chatham Islands and mainland New Zealand. The results from this study support the current allocation of the Chatham Islands to a specific management area, i.e. BCO4 (Figure 1.1).

The genetic connectivity along mainland New Zealand has been more difficult to interpret, because the evidence for genetic differentiation is weak. Mitochondrial DNA analysis (Chapter Two) found no significant pairwise differentiation between mainland sites, although a pattern of isolation-by-distance was observed. Microsatellite analyses, based on Bayesian structure and principal component analysis (Chapter Four), found little evidence for the existence of distinct populations along the mainland. However, pairwise differentiation between some mainland New Zealand sites.

Without the evidence for distinct population structuring, it can be more difficult to draw conclusions on population connectivity, since only a few migrants per generation may be needed for a population to become genetically homogenous. This low level of migration needed for genetic connectivity may not be sufficiently high for populations to be demographically coupled (Lowe and Allendorf, 2010). Such low migration levels will thus be expected to be insufficient for the dispersal process to support the recovery of depleted stocks (Waples, 1998). Because of this, the low levels of differentiation observed from microsatellite pairwise G_{ST} and R_{ST} should be considered carefully, as many factors can reduce the likelihood of detecting differentiation with microsatellite loci (Balloux and Lugon-Moulin, 2002; Larmuseau et al., 2010).

Although microsatellite DNA markers can be a powerful tool for the detection of fine-scale population structure, many factors have been identified which can lower their resolving power (Balloux et al., 2000; Estoup et al., 2002). In the case of high levels of heterozygosity within subpopulations, F-statistics have been shown to be constrained, as the maximum F_{ST} and G_{ST} values may never approach unity (Verity and Nichols, 2014). This was observed in the microsatellite results in the present study, since Jost's D and Hendricks G'_{ST} levels, which were developed to be less constrained, were approximately 10 fold higher than G_{ST} and R_{ST} . The level of genetic differentiation observed in a population is not only dependent on demographic connectivity (migration), but is also influenced by the population size and history (Hauser and Carvalho, 2008). The mtDNA analysis described in Chapter Two indicates that *P. colias* have undergone population expansion, which could reduce the effects of genetic drift, and consequently the level of genetic differentiation observed is a low levels of microsatellite F_{ST} relative to mtDNA F_{ST} , was attributed to a recent demographic expansion resulting in a slower rate of genetic differentiation (Larmuseau et al., 2010).

The results of this study further support evidence from previous microchemistry and markrecapture studies, indicating that *P. colias* is likely to form several distinct populations within current management areas. This is particularly evident in management area BCO3, which covers the east coast of the South Island, including the Otago and Kaikoura sample sites used in this study (Figure 1.1, Figure 4.1). The majority of the BCO3 commercial catch is taken from the lower half of the management area, near Otago, and this section of BCO3 has also seen a decline in CPUE since 2002-03 (Ministry for Primary Industries, 2014). Low, but significant differentiation was found between these two sample sites (G_{ST} =0.003 *P*<0.001, R_{ST} =0.005 *P*<0.05), indicating that this management area may contain more than one biologically distinct population. In light of this, it is recommended that fishery managers reexamine management boundaries within BCO3.

Future work

This study reports on the first microsatellite DNA markers developed for *P. colias*. These seven loci, as well as the 454 data library of potentially amplifiable loci (PALs), provide an important resource for future studies of *P. colias*. Next-generation sequencing technology made it possible to implement the direct seq-to-SSR microsatellite development method. This approach is more rapid and cost-effective than the traditional microsatellite isolation which relies on clone-based isolation protocols (Guichoux et al., 2011). The aforementioned data library can be used to develop additional loci for future studies involving *P. colias*. In general, new microsatellite loci need to be developed for the study of a new species, although it has been shown that loci can sometimes be adopted from marker sets developed for a closely related species. These microsatellite loci can also be applied to *P. colias* and could potentially be applied to other species for a range of applications such as parentage analysis, broodstock development and species identification.

Most population genetic studies, including the present study, focus on selectively neutral genetic markers for the detection of an unbiased indication of population connectivity. However, genetic markers that detect both neutral and selective variation can provide a more comprehensive insight into population dynamics (Schlotterer, 2004). The detection of adaptive variation could help to determine the plasticity of biological differences observed in populations from different locations, such as differences in size at sexual maturity (Kirk and Freeland, 2011; Nielsen et al., 2009).

Although the microsatellite DNA markers did detect low levels of differentiation between some mainland sites, this study was significantly hampered by its low sample sizes, particularly in the northern sites. Further studies using additional loci could also reduce the uncertainty surrounding the statistical significance of the differentiation estimates. However, considering the low number of loci, sample locations and sample sizes, genetic differentiation between mainland sites has nevertheless been detected in some cases, thus supporting previous inferences concerning the limited dispersal of *P. colias* and the potentially low demographic connectivity within current management sites.

Appendix



Appendix 1 Sequence length distribution of 288,947 reads from 454 pyrosequencing



Appendix 2 Delta K values for genetic clusters (K) from 2 to 7 for *P. colias* microsatellite loci



Appendix 3 Structure plots for K=2-7



Appendix 4 Rare allele frequency distributions for each sample site. The black bars represent the proportion of alleles observed in each allele frequency class. An L-shaped distribution is expected for a population at mutation-drift equilibrium.



Appendix 5 Rarefaction curves for each microsatellite locus

References

- Abdelkrim, J., B. C. Robertson, J. A. L. Stanton, and N. J. Gemmell, 2009, Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing: Biotechniques, v. 46, p. 185-+.
- Aboim, M. A., G. M. Menezes, T. Schlitt, and A. D. Rogers, 2005, Genetic structure and history of populations of the deep-sea fish Helicolenus dactylopterus (Delaroche, 1809) inferred from mtDNA sequence analysis: Molecular Ecology, v. 14, p. 1343-1354.
- Allendorf, F., P. England, G. Luikart, and P. Ritchie, 2008, Genetic effects of harvest on wild animal populations, Trends in Ecology and Evolution, p. 327-337.
- Allendorf, F., and G. Luikart, 2007, Conservation and the Genetics of Populations, Blackwell publishing.
- Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart, 2008, LOSITAN: A workbench to detect molecular adaptation based on a F(st)-outlier method: Bmc Bioinformatics, v. 9, p. 5.
- Antoni, L., P. L. Luque, K. Naghshpour, and E. A. Saillant, 2014, Polymorphic microsatellite markers for the Yellowfin tuna (Thunnus albacares): Conservation Genetics Resources, v. 6, p. 609-611.
- Apte, S., and J. P. A. Gardner, 2002, Population genetic subdivision in the New Zealand greenshell mussel (Perna canaliculus) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA: Molecular Ecology, v. 11, p. 1617-1628.
- Armour, J. A. L., R. Neumann, S. Gobert, and A. J. Jeffreys, 1994, ISOLATION OF HUMAN SIMPLE REPEAT LOCI BY HYBRIDIZATION SELECTION: Human Molecular Genetics, v. 3, p. 599-605.
- Armsworth, P. R., 2001, Effects of fishing on a protogynous hermaphrodite: Canadian Journal of Fisheries and Aquatic Sciences, v. 58, p. 568-578.
- Ashton, D., 2013, Population Genetics of New Zealand Pagrus auratus and Genetic Variation of an Aquaculture Broodstock, Victoria University of Wellington.
- Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser, and J. Goudet, 2000, Microsatellites can be misleading: An empirical and simulation study: Evolution, v. 54, p. 1414-1422.
- Balloux, F., and N. Lugon-Moulin, 2002, The estimation of population differentiation with microsatellite markers: Molecular Ecology, v. 11, p. 155-165.
- Beaumont, M. A., and R. A. Nichols, 1996, Evaluating loci for use in the genetic analysis of population structure: Proceedings of the Royal Society B-Biological Sciences, v. 263, p. 1619-1626.
- Beentjes, M. P., and G. D. Carbines, 2005, Population structure and relative abundance of blue cod (Parapercis colias) off Banks Peninsula and in Dusky Sound, New Zealand: New Zealand Journal of Marine and Freshwater Research, v. 39, p. 77-90.
- Beer, N. A., S. R. Wing, and S. E. Swearer, 2011, Otolith elemental evidence for spatial structuring in a temperate reef fish population: Marine Ecology Progress Series, v. 442, p. 217-227.
- Begg, G. A., and J. R. Waldman, 1999, An holistic approach to fish stock identification: Fisheries Research, v. 43.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme, 2000, GENETIX 4.02, logiciel sous Windows TM pour la genetique des populations., Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Université de Montpellier II. Montpellier, France.
- Bentzen, P., C. T. Taggart, D. E. Ruzzante, and D. Cook, 1996, Microsatellite polymorphism and the population structure of Atlantic cod (Gadus morhua) in the northwest Atlantic: Canadian Journal of Fisheries and Aquatic Sciences, v. 53, p. 2706-2721.
- Blackwell, R. G., 2006, Abundance and size composition of blue cod in the Marlborough Sounds and Tasman Bay, September-October 2004., Final Research Report for Ministry of Fisheries Research Project BCO2003-01 Objectives 1-3, National Institute of Water and Atmospheric Research.

- Bowen, B. W., A. Muss, L. A. Rocha, and W. S. Grant, 2006, Shallow mtDNA coalescence in atlantic pygmy angelfishes (Genus Centropyge) indicates a recent invasion from the Indian Ocean: Journal of Heredity, v. 97, p. 1-12.
- Carbines, G., 1999, Determination of movement of blue cod in Southland, Final Research Report for Ministry of Fisheries, Wellington, New Zealand.
- Carbines, G., 2004, Age, growth, movement and reproductive biology of blue cod (Parapercis colias-Pinguipedidae): implications for fisheries management in the South Island of New Zealand.: PhD thesis thesis, University of Otago, Dunedin, New Zealand.
- Carbines, G., and J. McKenzie, 2001, Movement patterns and stock mixing of blue cod in Southland (BCO5), Final Research Report for Ministry of Fisheries Research Project BCO9702, National Institute of Water and Atmospheric Research.
- Carbines, G. D., and M. P. Beentjes, 2009, Relative abundance, size and age structure, and mortality of blue cod off north Canterbury (BCO3) in 2007-08, New Zealand Fisheries Assessment Report 2009/37, Ministry of Fisheries, Wellington, p. 56 p.
- Carbines, G. D., and J. R. McKenzie, 2004, Movement patterns and stock mixing of blue cod in Dusky Sound in 2002, New Zealand Fisheries Assessment Report 2004/34, Ministry of Fisheries.
- Carvalho, G. R., and L. Hauser, 1994, Molecular genetics and the stock concept in fisheries, v. 4, p. 326-350.
- Castoe, T. A., A. W. Poole, A. P. J. de Koning, K. L. Jones, D. F. Tomback, S. J. Oyler-McCance, J. A. Fike, S. L. Lance, J. W. Streicher, E. N. Smith, and D. D. Pollock, 2012, Rapid Microsatellite Identification from Illumina Paired-End Genomic Sequencing in Two Birds and a Snake: Plos One, v. 7, p. 10.
- Castro, J. A., A. Picornell, and M. Ramon, 1998, Mitochondrial DNA: a tool for populational genetics studies: International microbiology : the official journal of the Spanish Society for Microbiology, v. 1, p. 327-32.
- Chistiakov, D. A., B. Hellemans, and F. A. M. Volckaert, 2006, Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics: Aquaculture, v. 255, p. 1-29.
- Chiswell, S. M., 2009, Colonisation and connectivity by intertidal limpets among New Zealand, Chatham and Sub-Antarctic Islands. II. Oceanographic connections: Marine Ecology Progress Series, v. 388, p. 121-135.
- Cole, R. G., D. S. Tindale, and R. G. Blackwell, 2001, A comparison of diver and pot sampling for blue cod (Parapercis colias : Pinguipedidae): Fisheries Research, v. 52, p. 191-201.
- Cole, R. G., E. Villouta, and R. J. Davidson, 2000, Direct evidence of limited dispersal of the reef fish Parapercis colias (Pinguipedidae) within a marine reserve and adjacent fished areas: Aquatic Conservation-Marine and Freshwater Ecosystems, v. 10, p. 421-436.
- Cowen, R. K., and S. Sponaugle, 2009, Larval Dispersal and Marine Population Connectivity, Annual Review of Marine Science: Annual Review of Marine Science, v. 1: Palo Alto, Annual Reviews, p. 443-466.
- Crawford, N. G., 2010, smogd: software for the measurement of genetic diversity: Molecular Ecology Resources, v. 10, p. 556-557.
- DeWoody, J. A., and J. C. Avise, 2000, Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals: Journal of Fish Biology, v. 56, p. 461-473.
- Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus, 2005, Bayesian coalescent inference of past population dynamics from molecular sequences: Molecular Biology and Evolution, v. 22, p. 1185-1192.
- Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut, 2012, Bayesian Phylogenetics with BEAUti and the BEAST 1.7: Molecular Biology and Evolution, v. 29, p. 1969-1973.
- Earl, D. A., and B. M. vonHoldt, 2012, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method: Conservation Genetics Resources, v. 4, p. 359-361.

- Ellegren, H., 2004, Microsatellites: Simple sequences with complex evolution: Nature Reviews Genetics, v. 5, p. 435-445.
- Elliott, C. P., N. J. Enright, R. J. N. Allcock, M. G. Gardner, E. Meglecz, J. Anthony, and S. L. Krauss, 2014, Microsatellite markers from the Ion Torrent: a multi-species contrast to 454 shotgun sequencing: Molecular Ecology Resources, v. 14, p. 554-568.
- Estoup, A., P. Jarne, and J. M. Cornuet, 2002, Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis: Molecular Ecology, v. 11, p. 1591-1604.
- Evanno, G., S. Regnaut, and J. Goudet, 2005, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study: Molecular Ecology, v. 14, p. 2611-2620.
- Excoffier, L., G. Laval, and S. Schneider, 2005, Arlequin (version 3.0): An integrated software package for population genetics data analysis: Evolutionary Bioinformatics, v. 1, p. 47-50.
- Faircloth, B. C., 2008, MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design: Molecular Ecology Resources, v. 8, p. 92-94.
- Fu, Y. X., 1997, Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection: Genetics, v. 147, p. 915-925.
- Gardner, M. G., A. J. Fitch, T. Bertozzi, and A. J. Lowe, 2011, Rise of the machines recommendations for ecologists when using next generation sequencing for microsatellite development: Molecular Ecology Resources, v. 11, p. 1093-1101.
- Garza, J. C., and E. G. Williamson, 2001, Detection of reduction in population size using data from microsatellite loci: Molecular Ecology, v. 10, p. 305-318.
- Goudet, J., 1995, FSTAT (Version 1.2): A computer program to calculate F-statistics: Journal of Heredity, v. 86, p. 485-486.
- Goudet, J., 1999, PCAGEN software for windows V.1.2.1 Available from http://www2.unil.ch/popgen/softwares/pcagen.htm.
- Guichoux, E., L. Lagache, S. Wagner, P. Chaumel, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E.
 Revardel, F. Salin, and R. J. Petit, 2011, Current trends in microsatellite genotyping:
 Molecular Ecology Resourses.
- Hamilton, M. B., E. L. Pincus, A. Di Fiore, and R. C. Fleischer, 1999, Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites: Biotechniques, v. 27, p. 500-+.
- Harpending, H. C., 1994, SIGNATURE OF ANCIENT POPULATION-GROWTH IN A LOW-RESOLUTION MITOCHONDRIAL-DNA MISMATCH DISTRIBUTION: Human Biology, v. 66, p. 591-600.
- Hauser, L., G. J. Adcock, P. J. Smith, J. H. B. Ramirez, and G. R. Carvalho, 2002, Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus): Proceedings of the National Academy of Sciences of the United States of America, v. 99, p. 11742-11747.
- Hauser, L., and G. R. Carvalho, 2008, Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts: Fish and Fisheries, v. 9, p. 333-362.
- Heath, R. A., 1985, A review of the physical oceanography of the seas around New Zealand 1982: New Zealand Journal of Marine and Freshwater Research, v. 19, p. 79-124.
- Hedgecock, D., P. H. Barber, and S. Edmands, 2007, Genetic Approaches to Measuring Connectivity: Oceanography, v. 20, p. 70-79.
- Hedrick, P. W., 1999, Perspective: Highly variable loci and their interpretation in evolution and conservation: Evolution, v. 53, p. 313-318.
- Hedrick, P. W., 2005, A standardized genetic differentiation measure: Evolution, v. 59, p. 1633-1638.
- Hellberg, M. E., 2009, Gene Flow and Isolation among Populations of Marine Animals, Annual Review of Ecology Evolution and Systematics: Annual Review of Ecology Evolution and Systematics, v. 40: Palo Alto, Annual Reviews, p. 291-310.
- Henderson, I. M., 2009, Optimising recreational harvests of blue cod: The effects of catch-andrelease mortality and size selectivity: Fisheries Research, v. 99, p. 184-195.

 Henriques, R., W. M. Potts, C. V. Santos, W. H. H. Sauer, and P. W. Shaw, 2014, Population
 Connectivity and Phylogeography of a Coastal Fish, Atractoscion aequidens (Sciaenidae), across the Benguela Current Region: Evidence of an Ancient Vicariant Event: Plos One, v. 9.

- Hickey, A. J. R., S. D. Lavery, D. A. Hannan, C. S. Baker, and K. D. Clements, 2009, New Zealand triplefin fishes (family Tripterygiidae): contrasting population structure and mtDNA diversity within a marine species flock: Molecular Ecology, v. 18, p. 680-696.
- Ho, S. Y. W., and B. Shapiro, 2011, Skyline-plot methods for estimating demographic history from nucleotide sequences: Molecular Ecology Resources, v. 11, p. 423-434.

Hoffman, J. I., K. K. Dasmahapatra, W. Amos, C. D. Phillips, T. S. Gelatt, and J. W. Bickham, 2009, Contrasting patterns of genetic diversity at three different genetic markers in a marine mammal metapopulation: Molecular Ecology, v. 18, p. 2961-2978.

- Holland, S. M., 2003, Analytic Rarefaction 1.3, <u>http://strata.uga.edu/software/index.html</u>.
- Holleley, C. E., and P. G. Geerts, 2009, Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR: Biotechniques, v. 46, p. 511-+.
- Ihaka, R., and R. Gentleman, 1996, R: a language for data analysis and graphics: Journal of Computational and Graphical Statistics, v. 5, p. 299-314.
- Jost, L., 2008, *G*_{ST} and its relatives do not measure differentiation: Molecular Ecology, v. 17, p. 4015-4026.
- Kimura, M., and J. F. Crow, 1964, NUMBER OF ALLELES THAT CAN BE MAINTAINED IN FINITE POPULATION: Genetics, v. 49, p. 725-&.
- Kimura, M., and T. Ohta, 1978, STEPWISE MUTATION MODEL AND DISTRIBUTION OF ALLELIC FREQUENCIES IN A FINITE POPULATION: Proceedings of the National Academy of Sciences of the United States of America, v. 75, p. 2868-2872.
- Kirk, H., and J. R. Freeland, 2011, Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology: International Journal of Molecular Sciences, v. 12, p. 3966-3988.
- Knutsen, H., E. M. Olsen, P. E. Jorde, S. H. Espeland, C. Andre, and N. C. Stenseth, 2011, Are low but statistically significant levels of genetic differentiation in marine fishes "biologically meaningful'? A case study of coastal Atlantic cod: Molecular Ecology, v. 20, p. 768-783.
- Laikre, L., S. Palm, and N. Ryman, 2005, Genetic Population Structure of Fishes: Implications for Coastal Zone Management: A Journal of the Human Environment, v. 34, p. 111-119.
- Larmuseau, M. H. D., J. A. M. Raeymaekers, B. Hellemans, J. K. J. Van Houdt, and F. A. M. Volckaert, 2010, Mito-nuclear discordance in the degree of population differentiation in a marine goby: Heredity, v. 105, p. 532-542.
- Lee, W. J., J. Conroy, W. H. Howell, and T. D. Kocher, 1995, STRUCTURE AND EVOLUTION OF TELEOST MITOCHONDRIAL CONTROL REGIONS: Journal of Molecular Evolution, v. 41, p. 54-66.
- Librado, P., and J. Rozas, 2009, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data: Bioinformatics, v. 25, p. 1451-1452.
- Lowe, W. H., and F. W. Allendorf, 2010, What can genetics tell us about population connectivity?: Molecular Ecology, v. 19, p. 3038-3051.
- Mace, J. T., and A. D. Johnston, 1983, Tagging experiments on Blue Cod (Parapercis colias) in the Marlborough Sounds, New Zealand: New Zealand Journal of Marine and Freshwater Research, v. 17, p. 207-211.
- Mach, M. E., E. J. Sbrocco, L. A. Hice, T. A. Duffy, D. O. Conover, and P. H. Barber, 2011, Regional differentiation and post-glacial expansion of the Atlantic silverside, Menidia menidia, an annual fish with high dispersal potential: Marine Biology, v. 158, p. 515-530.
- Manni, F., E. Guerard, and E. Heyer, 2004, Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm: Human Biology, v. 76, p. 173-190.

- Meglecz, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, and J. F. Martin, 2010, QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects: Bioinformatics, v. 26, p. 403-404.
- Meirmans, P. G., and P. W. Hedrick, 2011, Assessing population structure: F-ST and related measures: Molecular Ecology Resources, v. 11, p. 5-18.
- Meirmans, P. G., and P. H. Van Tienderen, 2004, GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms: Molecular Ecology Notes, v. 4, p. 792-794.
- Miya, M., H. Takeshima, H. Endo, N. B. Ishiguro, J. G. Inoue, T. Mukai, T. P. Satoh, M. Yamaguchi, A. Kawaguchi, K. Mabuchi, S. M. Shirai, and M. Nishida, 2003, Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences: Molecular Phylogenetics and Evolution, v. 26, p. 121-138.
- Nei, M., 1973, ANALYSIS OF GENE DIVERSITY IN SUBDIVIDED POPULATIONS: Proceedings of the National Academy of Sciences of the United States of America, v. 70, p. 3321-3323.
- Nielsen, E. E., P. J. Wright, J. Hemmer-Hansen, N. A. Poulsen, L. M. Gibb, and D. Meldrup, 2009, Micro geographical population structure of cod Gadus morhua in the North Sea and west of Scotland: the role of sampling loci and individuals: Marine Ecology Progress Series, v. 376, p. 213-225.
- Norrell, A. E., D. Crawley, K. L. Jones, and E. A. Saillant, 2014, Development and characterization of eighty-four microsatellite markers for the red snapper (Lutjanus campechanus) using Illumina paired-end sequencing: Aquaculture, v. 430, p. 128-132.
- Ovenden, J., 2013, Crinkles in connectivity: combining genetics and othertypes of biological data to estimate movementand interbreeding between populations: Marine and Freshwater Research, v. 64, p. 201-107.
- Patton, J. C., B. J. Gallaway, R. G. Fechhelm, and M. A. Cronin, 1997, Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (Coregonus nasus) in the Colville and Sagavanirktok rivers in northern Alaska: Canadian Journal of Fisheries and Aquatic Sciences, v. 54, p. 1548-1556.
- Peakall, R., and P. E. Smouse, 2006, GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research: Molecular Ecology Notes, v. 6, p. 288-295.
- Peakall, R., and P. E. Smouse, 2012, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update: Bioinformatics, v. 28, p. 2537-2539.
- Piry, S., G. Luikart, and J. M. Cornuet, 1999, BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data: Journal of Heredity, v. 90, p. 502-503.
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000, Inference of population structure using multilocus genotype data: Genetics, v. 155, p. 945-959.
- Rambaut, A., M. Suchard, D. Xie, and A. Drummond, 2014, Tracer v1.6, Available from <u>http://beast.bio.ed.ac.uk/Tracer</u>.
- Raymond, M., and F. Rousset, 1995, GENEPOP (VERSION-1.2) POPULATION-GENETICS SOFTWARE FOR EXACT TESTS AND ECUMENICISM: Journal of Heredity, v. 86, p. 248-249.
- Rioux Paquette, S., 2011, PopGenKit: useful functions for (batch file conversion and data resampling in mircosatellite datasets, p. R Package.
- Rocha-Olivares, A., N. M. Garber, A. F. Garber, and K. C. Stuck, 2005, Structure of the mitochondrial control region and flanking tRNA genes of Mugil cephalus: Hidrobiologica, v. 15, p. 139-149.
- Rodgers, K. L., and S. R. Wing, 2008, Spatial structure and movement of blue cod Parapercis colias in Doubtful Sound, New Zealand, inferred from delta C-13 and delta N-15: Marine Ecology Progress Series, v. 359, p. 239-248.
- Rogers, A. R., and H. Harpending, 1992, POPULATION-GROWTH MAKES WAVES IN THE DISTRIBUTION OF PAIRWISE GENETIC-DIFFERENCES: Molecular Biology and Evolution, v. 9, p. 552-569.

- Ross, P. M., I. D. Hogg, C. A. Pilditch, and C. J. Lundquist, 2009, Phylogeography of New Zealand's coastal benthos: New Zealand Journal of Marine and Freshwater Research, v. 43, p. 1009-1027.
- Ryman, N., S. Palm, C. Andre, G. R. Carvalho, T. G. Dahlgren, P. E. Jorde, L. Laikre, L. C. Larsson, A. Palme, and D. E. Ruzzante, 2006, Power for detecting genetic divergence: differences between statistical methods and marker loci: Molecular Ecology, v. 15, p. 2031-2045.
- Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989, Molecular cloning: a lab manual, v. 3: New York, Cold Spring Harbour Laboratory Press.
- Schlotterer, C., 2004, The evolution of molecular markers just a matter of fashion?: Nature Reviews Genetics, v. 5, p. 63-69.
- Schuelke, M., 2000, An economic method for the fluorescent labeling of PCR fragments: Nature Biotechnology, v. 18, p. 233-234.
- Slatkin, M., 1995, A MEASURE OF POPULATION SUBDIVISION BASED ON MICROSATELLITE ALLELE FREQUENCIES: Genetics, v. 139, p. 457-462.
- Smith, H., 2012, Characterisation of the Mitochondrial Genome and the Phylogeographic Structure of Blue Cod (*Parapercis colias*), Victoria University of Wellington.
- Smith-Keune, C., D. R. Jerry, B. S. Evans, M. I. McCormick, P. L. Munday, and G. P. Jones, 2006, Development and characterization of eight new microsatellite markers for the haremic sandperch, Parapercis cylindrica (family Pinguipedidae): Molecular Ecology Notes, v. 6, p. 1036-1038.
- Tajima, F., 1989, STATISTICAL-METHOD FOR TESTING THE NEUTRAL MUTATION HYPOTHESIS BY DNA POLYMORPHISM: Genetics, v. 123, p. 585-595.
- Tessier, N., L. Bernatchez, P. Presa, and B. Angers, 1995, Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon: Journal of Fish Biology, v. 47, p. 156-163.
- van Oosterhout, C., W. F. Hutchinson, D. P. M. Willis, and P. Shipley, 2004, MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data: Molecular Ecology Notes, v. 4, p. 535-538.
- Varela, A. I., P. A. Ritchie, and P. J. Smith, 2013, Global genetic population structure in the commercially exploited deep-sea teleost orange roughy (Hoplostethus atlanticus) based on microsatellite DNA analyses: Fisheries Research, v. 140, p. 83-90.
- Verity, R., and R. A. Nichols, 2014, What is genetic differentiation, and how should we measure it- G_{ST} , D, neither or both?: Molecular Ecology.
- Waples, R. S., 1998, Separating the Wheat From the Chaff: Patterns of Genetic Differentiation in High Gene Flow Species: The Journal of Heredity, v. 89, p. 438-450.
- Waples, R. S., and O. Gaggiotti, 2006, What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity: Molecular Ecology, v. 15, p. 1419-1439.
- Waples, R. S., A. E. Punt, and J. M. Cope, 2008, Integrating genetic data into management of marine resources: how can we do it better?: Fish and Fisheries, v. 9, p. 423-449.
- Ward, R. D., 2000, Genetics in fisheries management: Hydrobiologica, v. 420, p. 191-201.
- Ward, R. D., and N. G. Elliott, 2001, Genetic population structure of species in the South East Fishery of Australia: Marine and Freshwater Research, v. 52, p. 563-573.
- Ward, R. D., M. Woodwark, and D. O. F. Skibinski, 1994, A COMPARISON OF GENETIC DIVERSITY LEVELS IN MARINE, FRESH-WATER, AND ANADROMOUS FISHES: Journal of Fish Biology, v. 44, p. 213-232.
- Weir, B. S., and C. C. Cockerham, 1984, ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION-STRUCTURE: Evolution, v. 38, p. 1358-1370.
- Whitlock, M. C., 2011, G '(ST) and D do not replace F-ST: Molecular Ecology, v. 20, p. 1083-1091.
- Zane, L., L. Bargelloni, and T. Patarnello, 2002, Strategies for microsatellite isolation: a review: Molecular Ecology, v. 11, p. 1-16.