Population Genetics of New Zealand *Pagrus auratus* and Genetic Variation of an Aquaculture Broodstock

By

David T. Ashton

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

Victoria University of Wellington

2013

This thesis was conducted under the supervision of:

Dr. Peter Ritchie

Victoria University of Wellington
Wellington, New Zealand

Abstract

Fisheries and aquaculture are major contributors of nutrition and animal protein worldwide. Understanding the genetic variation and differentiation within and between wild populations is important for both sustainable fisheries management and selection of aquaculture broodstock. This study determined the genetic variation and differentiation of New Zealand *Pagrus auratus* based on mitochondrial DNA control region sequencing and microsatellite DNA genotyping. Low but significant differentiation was measured between several sample sites, but otherwise the population was genetically panmictic. The M-ratio test and Fu's Fs statistics indicate that there may have been historical bottlenecks at all sample sites and a more recent bottleneck in the Tasman Bay. Two South Island sites were identified that had not been through recent bottlenecks and were not significantly differentiated from the Tasman Bay, which may provide a source of gene flow to aid its genetic recovery. Comparison of the broodstock and wild genetic variation indicate that the broodstock represented most of the genetic variation found in high frequency in wild populations, but further wild-caught individuals may be needed, based on the criteria used in several previous studies. Simulations indicate that adding approximately 20 and 48 wild-caught individuals from multiple populations to the current broodstock was needed to represent all genetic variation above a target frequency of 0.05 in the Tasman Bay and all sample sites, respectively.

Acknowledgements

This thesis would not have been possible without the support from each and every one of my family. A big thanks to Mum, Dad, and each of my brothers and sisters. I would especially like to thank my wife Ellie for supporting me with my study over the last 2 years.

I would also like to acknowledge the people I have met and worked with at VUW throughout this study; in particular everybody in the lab and Peter Ritchie's research group. Special thanks to the Hayden Smith, Henry Lane, and Sebastian Hernandez for all your help in the lab and with data analysis. Most importantly, I would like to thank my supervisor Dr. Peter Ritchie. Your guidance when selecting study material and many discussions throughout this thesis have really helped me gain a better understanding of the subject.

Finally, I would like to thank everybody at Plant and Food Research Nelson that has had input into this work in any way. Thank-you to all staff members who took time out of their busy schedules to arrange and collect samples; in particular Sharon Ford, your help with sample collection was indispensable. I would also like to thank Dr. Alistair Jerrett for his support for the project and of me pursuing further study.

Table of Contents

Chapter 1: General Introduction

1:1 Fisheries and Aquaculture Overview

Fisheries and aquaculture are major contributors of nutrition and animal protein worldwide. Globally 148 million tonnes of fish was produced in 2010 and production is projected to exceed that of beef, pork, and poultry in the next decade (FAO 2012). Wild-capture fisheries production has remain relatively stable since the late 1990's, but increased production from the aquaculture sector has driven the growth of global fisheries production at an average of 3.2% per year (FAO 2012). The static production of the wild-capture fisheries is primarily because stocks are at, or close to, their maximum yields. Increases in aquaculture production will be crucial for meeting the needs of human population size growth.

Commercial and recreational fisheries are an important source of food and income in New Zealand. The total allowable commercial catch (TACC) for all of New Zealand fisheries is approximately 600 thousand tonnes per year. Total export value for fisheries and aquaculture is between $1.2-1.8 billion dollars; of which aquaculture produces about $200 million dollars or $279 million dollars in 2009 (MFish 2012). A significant amount of growth in the aquaculture sector is predicted over the next few years and a target of a $1 billion dollar industry by 2025 has been set by the aquaculture industry.

Fish is unique among the major human food items because it is still largely harvested directly from wild populations. Sustainable use of this natural resource relies on careful management and maintenance of fisheries stocks. The importance of careful stock assessment and management has been highlighted by the crash of several major fisheries including the Newfoundland cod and Peruvian anchovy fisheries (Walters & Maquire 1996; Idyll 1973). The quota management system (QMS) was introduced in 1980's amidst concerns about the state of New Zealand fisheries. The QMS structures the exclusive economic zone (EEZ) into different quota management areas (QMA) for each fishery species. Each QMA has a total allowable catch (TAC) which is set to the specific management requirements of the stock (Lock & Leslie 2007).

The effective use of the QMS relies on correlation of the management areas with biological stocks, and setting the TAC at a level, which maintains the biological sustainability of the fisheries. Population genetic markers are an important tool for determining the boundaries and structure of a fishery stocks (Connor 2001; Reiss *et al.* 2009). Genetic differentiation between stocks arises largely due to a combination of restriction of migration, the spatial distances between stocks.

1.2: Genetic Population Structure

1.2.1: Factors That Influence Genetic Population Structure

The genetic structure of a population is determined by a combination of biological and environmental factors; including species mobility during and after the larval stage, distance between populations, breeding habits, required environmental conditions for growth and survival, and environmental barriers to migration (Shanks 2009; Larson & Julian 1999; Wright 1943; Olsson *et al.* 2011; Thorrold *et al.* 2000; Ruzzante *et al.* 2005). The relative importance of each of these parameters will vary among species and the range over which the populations are distributed. The interaction of some factors may also be involved or required in the formation of population structure. For example passive dispersal during the pelagic larval stage is influenced in part by pelagic larval duration (PLD), but also the particular water currents that occur at that location and time (Shanks 2009); so a species with long PLD may not travel very far if there are limited water currents to disperse it (Siegel *et al.* 2003; Siegel *et al.* 2008; Selkoe & Toonen 2011; Weersing & Toonen 2009).

1.2.2: Types of Population Structures

Laikre *et al.* (2005) identified three general types of genetic population structure; panmixia, isolation by distance (IBD), and distinct stocks (Figure 1.1.0). Firstly, a panmictic population is when there is no genetic differentiation across the distribution range of a species. Secondly, isolation by distance describes a population where the level of differentiation increases as a function of the geographic distance among the locations. Thirdly, distinct stocks describe populations that are different because of barriers to gene flow.

Figure 1.1.0: Three General Types of Genetic Population Structure (Laikre *et al.* 2005).

Wild populations can be comprised of a mixture of these three types of structure (Laikre *et al.* 2005). A population might exhibit isolation by distance across most of the population's geographic range, but also have some distinct populations in areas where barriers to gene flow occur. Indeed, even the distinct stocks structure is a combination of panmictic subpopulations with distinct barriers to gene flow at the end of each subpopulation's distribution.

1.2.3: Genetics and Fisheries Management

One of the primary goals of population genetic studies is to estimate the genetically effective population size and the rate of gene flow between fishery stocks. Understanding the genetic structure and variation within a fishery is important for management because maintaining the genetic diversity of a population affects the long-term persistence and adaptability of a species when confronted with environmental challenges (Franklin 1980). If a population is highly structured with very little or no gene flow between fishery stocks but is managed as a single stock, then each distinct stock is susceptible to over-exploitation because fishing pressures are typically unevenly distributed and concentrated in particular areas (Carvalho & Hauser 1994). The intense exploitation of a small discrete stock that is considered to be part of a larger stock can cause it to be depleted, because the expectation was that it would be replenished. On the other hand, if a species has no or low levels of population structure (e.g. panmixia, or weak IBD) but is managed as a group of separate

stock, then if one management stock is over-exploited it will most likely be replenished, because it is part of a larger reproductive group.

There are a number of direct and indirect methods that can be used to estimate migration rates and stock structure; including individual tagging, otolith microchemistry, and genetic markers (Willis *et al.* 2001; Morrison 2008; Bernal-Ramı´rez *et al.* 2003). Each of these has particular areas of strength when used as tools for fisheries management. Tagging studies are able to directly measure the movement of adults. Otolith microchemistry is able to directly measure the movement of larvae and developing adults. Genetic markers can be used as an indirect population marker to estimate population connectivity. In addition genetic markers are important tools for fisheries management because maintaining the genetic diversity of a population is important for the adaptability of a population when it experiences an environmental challenge (Franklin 1980).

1.2.4: The Genetic Effects of Industrial Scale Fishing

Industrial fishing can affect both the structure and levels of genetic variation of fishery stocks. Significant changes can occur in the stock structure if fishing pressure interrupts gene flow between areas or when subpopulations are displaced from their natural location (Allendorf *et al.* 2008). Overfishing can cause a reduction in population size, which increases the strength of genetic drift and causes a loss of genetic variation. The loss of genetic variation in populations with a small population size can result in inbreeding depression (which may depress the reproductive output of the population) and a reduction in evolutionary potential. The loss of evolutionary potential might not be important in the short-term, but will inevitably limit the adaptive response of the population when faced with a new environmental challenge (Jamieson & Allendorf 2012). Fishing techniques may also introduce artificial selection pressures such as selecting for fish that are small and breed early because the larger breeders are typically targeted; eventually resulting in a smaller biomass (Allendorf & Hard 2009).

1.2.5: Temporal Changes in Population Structure

A number of studies have highlighted the importance of determining whether population structure is stable over time (Larson & Julian 1999; Planes & Lenfant 2002; Lacson & Morizot

1991). Temporal changes in population genetic structure can be caused by several factors include “sweepstake” recruitment, natal homing, and rapid environmental changes (David *et al.* 1997; Smith & Johnston 1985; Thorrold *et al.* 2000; Lacson & Morizot 1991).

Some marine fish populations are thought to have effective population sizes that are orders of magnitude smaller than their census population (Frankham 1994). The idea of a “sweepstake” recruitment process has been used to explain the large differences that are reported between census and effective population size (Larson & Julian 1999). A pattern of “sweepstake” recruitment most likely occurs because of large variance in the reproductive success of adults (Hedgecock 1994), which could in part be due to high mortality of larvae before settlement (Johnson & Black 1982). Large fluctuations in the number of reproductively successful individuals in a population can cause rapid changes in genetic variation within and between populations because of strong genetic drift (David *et al.* 1997).

Lacson & Morizot (1991) proposed that a bottleneck in one population had caused it to become differentiated from a second population, but that the differentiation was lost within three years due to migration. This migration was most likely occurring before, during, and after the bottleneck, but for a short time was not fast enough to keep up with the rate of differentiation caused by genetic drift.

1.2.6: Recent Demographic History

Genetic makers are useful tools for measuring historical population demographics, which in turn can be useful for understanding how populations persist over longer periods of time. Coupled with the knowledge that many environments experience significant changes over a long time period this allows a better appreciation of the response the species will make to future perturbations. Recent demographic changes, such as population bottlenecks, may also influence genetic measurements of population structure (Lacson & Morizot 1991). Consequently, understanding a species recent demographic history should help when interpreting the results of genetic measures of population structure and whether these genetic measures indicate a temporally stable population structure.

1.3: Genetics and Broodstock Selection

1.3.1: Broodstock Selection

Genetic markers can be very informative when constructing a broodstock from wild populations. Broodstock should contain a sufficient amount of the wild source genetic variation because those first individuals will contain the genetic potential that underpins a selective breeding program and limits the problems that can result from inbreeding (Allendorf & Phelpsa 1980; Charlesworth & Willis 2009). A genetic population study is useful when constructing a broodstock as it can be used to determine how much of the wild source genetic variation is present in the broodstock and the number of individuals that are needed to effectively represent that variation. If a large number of markers are used (e.g. 100) it is sometimes possible to identify genetic variation associated with useful production traits (Nielsen *et al.* 2009). Once genetic variation associated with useful production traits has been identified it can be used to selectively breed individuals to increase the frequency of these useful traits. One of the advantages of this type of genotyping-based selection is that it allows the measurement of traits in breeding candidates that would otherwise require destructive sampling (Meuwissen & Goddard 1996); for example selecting for tissue quality in aquaculture broodstock.

1.3.2: Number of Broodstock

When founding a broodstock it is important to consider the number of reproducing individuals that are needed to maintain genetic diversity within a captive group. As a rule-of-thumb Franklin (1980) proposed that 50 individuals are needed to prevent excessive amounts of inbreeding and the loss of genetic diversity from strong genetic drift. For a recent discussion of this rule see Jamieson & Allendorf (2012). The development of minimal kinship breeding methods to reduce inbreeding and the slightly pronounced inbreeding rate used by Franklin (1980) may allow the target to be slightly lower than 50 individuals (Jamieson & Allendorf 2012; Doyle *et al.* 2001); however, the target of 50 individuals should apply accurately to most small populations. While the 50 individuals may be a suitable population size for preventing inbreeding depression it might not be adequate for capturing all the genetic variation that might be useful in a selective breeding program. It is difficult to determine the level of genetic variation required for a selective breeding program because

gene variants cannot be easily matched to specific phenotypic traits. In general, the best approach when establishing a broodstock is to maximize the level of genetic variation from the wild source populations. A population genetic study can be used to determine if the cultured population has sufficient wild genetic variation by comparing the levels of genetic variation and differentiation within and between the wild and cultured populations (e.g. Ha *et al.* 2009, Song *et al.* 2011).

1.4: Genetic Markers for Fishery Management and Broodstock Selection

1.4.1: Selecting Genetic Markers

Choosing the type of genetic marker to use for a project is determined by the project scope and whether the project is measuring population demographics, connectivity, or adaptive variation (Selkoe & Toonen 2006). Statistical power and cost are two of the main factors governing the selection of neutral markers for a project, as well as a few factors unique to each marker type (Schlotterer 2004). Neutral genetic markers are most often used to describe the levels of genetic variation and differentiation within and among populations. DNA markers that detect loci under selection can also be used, but they are often technically difficult and expensive to develop (Conover *et al.* 2006).

1.4.2: Mitochondrial Sequencing

Mitochondrial (mtDNA) sequencing is commonly used to measure population genetic variation and differentiation (e.g. Alves *et al.* 2001; Bernal-Ramı´rez *et al.* 2003; Liu *et al.* 2012). The fast mutation rate, maternal inheritance, and lack of recombination make mtDNA unique for use in determining population structure and demographic history.

The mitochondrial genome has a 5-10 times faster mutation rate than the nuclear genome (Brown *et al.* 1979; Vawter & Brown 1986). In particular, the control region, a neutral sequence located between the tRNA-Pro and tRNA-Phe genes, mutates faster than other areas of the mitochondrial sequence (Brown 1985; Hoelzel *et al.* 1991). Some estimates of mutations rate for the mtDNA control region in fishes range from $1.3x10^{-7}$ to $0.7x10^{-8}$ mutations per base per year (Brown *et al.* 1993; Bowen *et al.* 2006; Mccusker & Bentzen 2010; Padhi 2011).

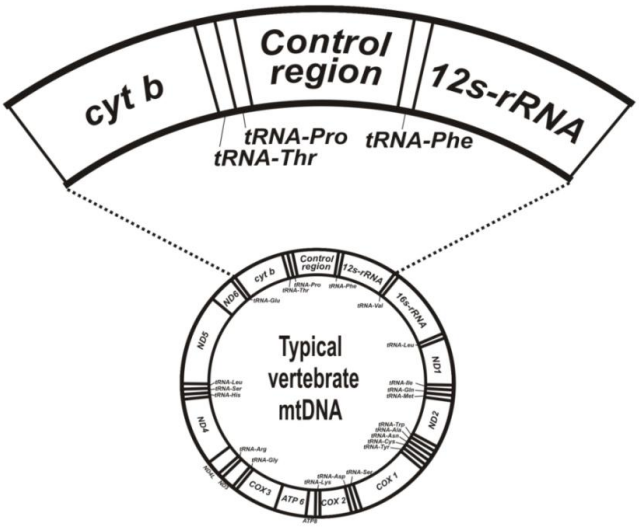


Figure 1.1.1: Typical Vertebrate Mitochondrial Genome and Control Region (Courtesy of Hayden Smith).

A unique feature of the mitochondrial genome is that it is typically maternally inherited; resulting in an effective population size one-fourth the size of nuclear genes in the same population (Carvalho & Hauser 1994). The smaller effective population size increases the sensitivity of mitochondrial sequences as a marker for detecting population subdivisions and demographic changes (Birky *et al.* 1989). The lack of recombination is also unique to the mitochondrial genome and has important implications for how it is used in data analysis (Harrison 1989). Because no reshuffling of the genes occurs between generations a mitochondrial sequence will reflect the mutations that have occurred in a single lineage.

1.4.3: Microsatellite Genotyping

Microsatellites are another commonly used marker for measuring population variation and differentiation (e.g. Berry *et al.* 2012; Bernal-Ramı´rez *et al.* 2003; Larsson *et al.* 2007). Microsatellites are short tandemly repeated sequences of DNA and are highly informative due to being diploid, having a fast mutation rate, and recombination between loci (Ellegren 2000).

The microsatellite mutation rate for a number of terrestrial and aquatic species ranges from $1.0x10^{-3}$ to $1.0x10^{-4}$ mutations per locus per generation; versus $1.0x10^{-9}$ mutations per base per generation for other sequences (Ellegren 2000). The proposed mutation models for microsatellites are the stepwise mutation model (SMM) or two-phase mutation model (TPM) (Liu & Cordes 2004). The TPM model is a combination of the SMM model and the infinity allele model (IAM), which is the mutation model for many other sequences (Sainudiin *et al.* 2004).

Each position on the nuclear genome represents a slightly different genealogy due to recombination of genes between each generation, causing a sample from a single point on the genome to have a high rate of sampling error (Selkoe & Toonen 2006). A data set using multiple microsatellite loci represents multiple points on the genome and decreases the sampling error rate. Comparatively, due to the lack of recombination mitochondrial sequencing represents a single point on the genome. Measuring genetic variation with nuclear markers also allows the measurement of the heterozygosity statistic, which can be used to determine population connectivity and demographics (Wright 1951; Cornuet & Luikart 1997).

1.5: Target Species: New Zealand Snapper (*Pagrus auratus*)

1.5.1: Biology and Distribution

Pagrus auratus (Silver sea bream; Snapper) is an inshore teleost species commonly found around the coasts of New Zealand, Australia and a number of Pacific Islands. They are closely related to Japanese *Pagrus major* (Red sea bream) (Smith *et al.* 1978), and these species are thought to have diverged 2-6 million years ago (Tabata & Taniguchi 2000). The New Zealand *P. auratus* is distributed from the top of the North Island to the northwest coast of the South Island (Smith *et al.* 1978). Some fish are caught further south, but there are no known breeding populations below the top of the South Island; most likely because of the low survival rates of larvae at low temperatures (Cassie 2005). *P. auratus* is commonly found inshore in depths ranging from 20 to 60 meters; they do however range down as far

as 200 meters (Smith *et al.* 1978). Tagging studies indicate that although they can travel long distances adults typically travel less than 100km from their tag site (Crossland 1976; Gauldie and Wood 2002). Other studies have shown extreme site fidelity of adults to within 500 meters (Willis *et al.* 2001). In contrast to the site fidelity shown by tagging studies, chemical studies of *P. auratus* otoliths have shown that 98% of *P. auratus* on the west coast of the North Island originate from the Kaipara Harbor (Morrison 2008). The relatively short pelagic life stage of *P. auratus* should limit the amount of passive dispersal directly after hatching. Cassie (1956) found that *P. auratus* settle after several days. The Tasman Bay population is located at the most southern limit of the *P. auratus* distribution and is known to be negatively affected by a minimum spawning temperature. It has been suggested that in some seasons this could result in a complete lack of recruitment (Annala *et al.* 2000).

1.5.2: The Wild-Capture Fishery

New Zealand *P. auratus* was a source of food for Maori, and has since become an iconic recreational and commercial fish. The 1999 to 2000 Ministry of Fisheries report on *P. auratus* stated that it is one of the largest and most valuable coastal fisheries in New Zealand (Gilbert and Phillips 2002). In 2009 the New Zealand *P. auratus* fishery made up 7% of the total value of New Zealand's commercial fish resource (Statistics NZ 2012), which makes it one of the most valuable inshore New Zealand teleost fish species. The recorded catches have remained relatively stable from 1996 to 2009, with the total allowable catch (TAC) currently set at 6,357 tonnes (Statistics NZ 2012 MFish 2012). Historically, the annual commercial catch was 6,224 tonnes in 1950 and increased to 16,362 tonnes by 1979, but concerns about over-exploitation in the mid 1980's resulted in a decrease to the annual catch levels (Fishery Summary 2012; Gilbert and Phillips 2002). The total allowable catch for both recreational and commercial *P. auratus* fishing is 10,132 of which 3,776 tonnes has been allocated to the recreational catch (Seafood Industry Council 2011).

The New Zealand *P. auratus* fishery is divided into six management areas East Northland/Hauraki Gulf/Bay of Plenty (SNA1), East Coast of the North Island (SNA2), South-East Coast of the North Island (SNA 3), Marlborough/Tasman (SNA7), West Coast of the North Island (SNA8) and Kermadec (SNA 10) (Figure 1.1.2).

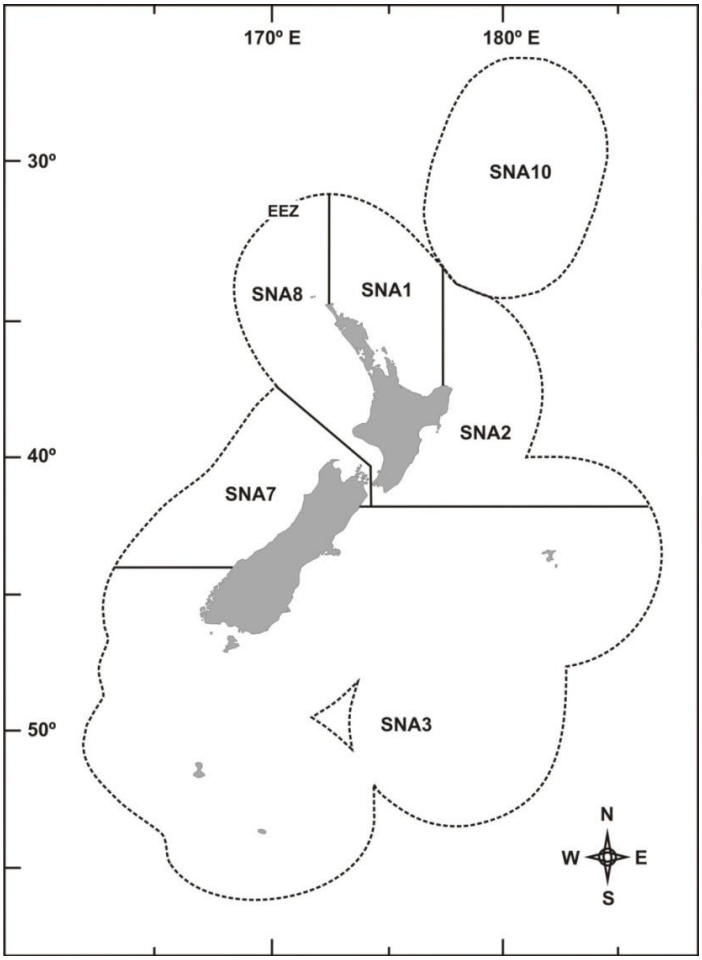


Figure 1.1.3: New Zealand *P. auratus* Management Areas (Courtesy of Sebastian Hernandez).

1.5.3: Aquaculture Developments

There is currently no commercial aquaculture of *P. auratus* in New Zealand, but techniques for growing this species in captivity have been well developed and these developments are being used in other research projects at Plant and Food Research (PFR) in Nelson. The PFR *P. auratus* broodstock were comprised of 26 randomly collected individuals from Tasman Bay with an approximate 1:1 ratio of male and females. Comparison of the broodstock and wild populations has not been carried out to determine how well they represent the genetic variation in the wild population.

1.5.4: Allozymes, Mitochondrial DNA, and Microsatellite DNA analysis

The first genetic studies of New Zealand *P. auratus* used allozyme variation and identified low but significant levels of differentiation between sample sites (Smith *et al.* 1978). Overall, this study found a north to south genetic differentiation of populations, which was attributed to differences in water temperature causing selective difference to arise at the esterase locus. This suggestion was supported by another study that found a correlation of the frequency of alleles at the esterase locus in the Hauraki Gulf larvae with seasonal fluctuations in water temperatures (Smith 1979).

A second study, in 2003 using mitochondrial and microsatellite DNA markers, found a reasonable level of congruency between the results of the allozyme and microsatellite loci; with low but significant levels of differentiation present between the same sample sites (Bernal-Ramı´rez *et al.* 2003). This study identified three genetically distinct populations; north-eastern North Island, southern North Island, and northern South Island. The authors concluded that genetic differentiation between populations is most likely due to water currents acting as barriers to gene flow and genetic drift driving the genetic difference between isolated stocks (Bernal-Ramı´rez *et al.* 2003). The mitochondrial analysis in this study was unable to find this same pattern of significant differentiation (Bernal-Ramı´rez *et al.* 2003).

Hauser *et al.* (2002) investigated the possible effects of overfishing on the genetic diversity of Hauraki Gulf and Tasman Bay populations by sampling the temporal variation of microsatellite DNA loci. The authors reported that the Tasman Bay population had undergone a significant reduction in genetic diversity over the past 50 years, which would most likely have been caused by a reduction in the population size by commercial and recreational fishing pressures.

1.5.5: Tagging and Otolith Chemistry Analysis

Studies of tagged *P. auratus* and otolith microchemistry have provided some seemingly conflicting results about the movement of individual New Zealand *P. auratus* (Willis *et al.* 2001; Morrison 2008). The results of tagging studies suggested a high level of site fidelity, while the results of otolith microchemistry suggested long distance dispersal of individuals

around the western North Island (Crossland 1976; Gauldie and Wood 2002; Willis *et al.* 2001; Morrison 2008). The different findings of these studies could be due to the different types of population connectivity and geographical location measured in each of the studies. Tagging studies detect the movement of adults, while otolith microchemistry detects the movement of larvae, developing adults, and adults. In places affected by strong water currents this may cause increased movement during the larval stage and possibly even with developing adults.

1.6: Thesis Aims

This research had two overall goals; **#1**: Use mitochondrial and microsatellite DNA markers to determine the levels of genetic variation and differentiation within and between wild New Zealand *P. auratus* populations, **#2**: Use mitochondrial and microsatellite DNA markers to determine the level of genetic variation present in the Plant and Food Research (PFR) broodstock, and how well they represent wild population variation.

Chapter two is a presentation of the genetic results obtained from the wild populations. More sampling effort was given to the southern populations of New Zealand because there is particular interest in gene flow around this area that might be involved in the recovery of the Tasman Bay from a bottleneck and it is also the source of the current Broodstock. Samples were collected from a number of North Island sample sites to be used as a comparison for the South Island sites and the cultured population. In particular, this chapter addressed the question of whether there are differentiated populations south of the Cook Strait, if so have they undergone the same bottleneck observed in the Tasman Bay, and how might they influence its genetic recovery. This chapter also takes a look at the demographic history of this species in New Zealand using new Bayesian methods.

Chapter three is a presentation of the genetic variation found in the Broodstock samples and how well this variation represents the wild genetic variation. A basic Python modeling script was developed and used to investigate the best sampling plan for adequately sampling the genetic variation in the wild populations.

Chapter four is a discussion of the overall results of this genetic study, how the findings fit in with the results of previous studies, and what implications all of these studies have for the management of the *P. auratus* fishery and the development of *P. auratus* as an aquaculture species.

Chapter 2: The Population Genetic Structure of New Zealand Snapper (*Pagrus auratus*) based on Mitochondrial and Microsatellite DNA Markers

2.1: Introduction

Pagrus auratus (Silver sea bream; Australasian Snapper) is an inshore teleost species commonly found around the coast of New Zealand, Australia and a number of Pacific Islands. *P. auratus* supports an important recreational and commercial fishery in New Zealand (Gilbert and Phillips 2002) and the 6,357 tonnes of fish that are typically caught each year represent 7% of New Zealand fisheries catch (Statistics NZ 2012). In New Zealand, *P. auratus* are distributed from the top of the North Island to the north-west coast of the South Island (Smith *et al.* 1978). They are commonly found inshore in depths ranging from 20 to 60 meters, but do range as far down as 200 meters (Smith *et al.* 1978). There has been some research into the stock structure of *P. auratus* and movement of individuals using tagging and otolith microchemistry. The studies utilizing these two methods have found different results; with tagging studies indicating site fidelity over a three year period, and otolith studies showing large scale migrations over a 700 km stretch of coastline (Willis *et al.* 2001; Morrison 2008). However, these studies were applied to *P. auratus* in different areas of the New Zealand coast and it is unsure whether this species has different site specific migratory patterns or levels of passive dispersal.

Several genetic studies of New Zealand *P. auratus* have been carried out using mitochondrial DNA (mtDNA) control region analysis and microsatellite DNA genotyping (Adcock *et al.* 2000; Bernal-Ramı´rez *et al.* 2003; Hauser *et al.* 2002). No significant levels of population differentiation have been reported using the mtDNA markers (Adcock *et al.* 2000; Bernal-Ramı´rez *et al.* 2003), but studies using five microsatellite DNA markers have found low but significant levels of differentiation between populations and evidence of a very recent bottleneck in the Tasman Bay (Bernal-Ramı´rez *et al.* 2003; Hauser *et al.* 2002). Hauser *et al.* (2002) suggested a recent genetic bottleneck might have occurred within the last 60 years within the Tasman Bay due to excessive fishing pressure. An analysis of the population genetic structure of *P. auratus* indicated that the Tasman Bay population was

most likely separated from other New Zealand *P. auratus* populations (Bernal-Ramı´rez *et al.* 2003). If the Tasman Bay is a genetically distinct population it may not recover its historic levels of genetic variation since no other populations would be able to replenish the genetic diversity by gene flow. However, there has been limited sampling around the bottlenecked Tasman Bay population so it was possible that there may be other South Island populations that could provide a source of gene flow for the Tasman Bay population. The possibility that there may be other sources of gene flow into the Tasman Bay and that it may have started to regain some of its previous genetic variation needs to be tested with another genetic study of *P. auratus* populations.

The aim of the research reported in this chapter was to sample the New Zealand *P. auratus* populations ten years after the previous genetic studies and determine the level of genetic variation and differentiation within and between a number of sample sites. Temporal sampling in the Tasman Bay was able to investigate whether this population has recovered any of its previous genetic diversity in the last ten years. Further sampling was also carried out at several South Island sample sites other than the Tasman Bay to determine if there were any populations in the South Island, which may provide a source of genetic variation to aid in the bottlenecked population's recovery. Samples from a number of other sites around New Zealand were also collected, including from the North Island. These samples provided estimates of genetic variation against which the South Island populations could be compared and were also used to construct a more detailed demographic history of this species in New Zealand. The use of sequencing rather than the previously used single stranded conformation polymorphism (SCCP) for analyzing the mtDNA control region allowed the use of new Bayesian methods for the determining the demographic history of this species (Bernal-Ramı´rez *et al.* 2003). The samples from all sample sites were also used in chapter 3 of this thesis, as a reference against which to compare the Plant and Food Research (PFR) *P. auratus* aquaculture broodstock. A few of the results for the PFR broodstock have been shown in this chapter, but they are predominantly shown and discussed in chapter 3. Mitochondrial DNA control region sequencing and microsatellite DNA genotyping were used because that were similar markers to those used in previous studies and were currently the best markers available to provide a general overview of genetic variation in *P. auratus* populations.

2.2: Materials and Methods

2.2.1: Primer Development

Mitochondrial DNA (mtDNA) control region primers were designed using the whole mitochondrial genome sequence for *P. major* (Genbank Accession NC_003196.1). Initially, a pair of primers was developed to amplify and sequence the entire control region (tRNA-Pro 5'-CACCATTGGCTCCCAAAGC-3' and tRNA-Phe 5'-GCTTTCTAGGGCCCATC-3'), but a centrally located Poly-A sequence caused problems with amplification and sequencing. Therefore another primer (CR-CCD 5'-GGCACTGTGAGATGTCAACTG-3') was designed to a sequence in the Central Conserved Domain (CCD) of the control region, which was paired with the tRNA-Pro primer to amplify the 641 base pair (bp) sequence from the 5'-end of the control.

The microsatellite DNA primers previously reported in studies on *P. auratus* (Adcock *et al.* 2000) and *P. major* (Hatanaka & Yamada 2006) were used in the present study (see Table 2.1.1). Each forward microsatellite primer had an additional M13 sequence added to the 5'-end and this was used to add a fluorescent label for detecting the fragment on an ABI 3730 Genetic Analyzer (Schuelke 2000).

Table 2.1.1: Microsatellite Primers (M13 primer 5'-TGTAAAACGACGGCCAGT-3')

Locus	Forward 5'-3'	Reverse 5'-3'
Pma1	M13 + CATGCCAGTATTCCATGTGC	AGGACAAATTCCCAAGGTCATCC
Pma2	M13 + TGCCAAGGAGAGGTGAGGAG	TATTCTTGCAAGGGTCCAAACG
Pma3	M13 + CTAAACAAATAATTCTGGTTTAGC	CATGACAAACTGAGGTGAATGC
Pma4-32	M13 + CCTGCCACCTACTGTTTCCT	CGGTGATTACAGTCGGGTTT
Pma5	M13 + CCAATCGGATTGAGTATCTGTGG	GAGAGGTTCTCCGTCACTGTCC
GA2A	M13 + ACGGACAGAGAGGGAGTGG	CATCATCATCAGTCAGAGCTG
GT2	M13 + TAGGGGTCTTGCTGCCTGCTC	ACAGATACAGACTGACAGACG
GT4	M13 + CTTTACACTGTTGAAGCTTTGG	CGCACAGATAAACAAATCAAT

2.2.2: Sample Collection and DNA Preparation

Samples were collected from eight sites around the New Zealand coast and one cultured *P. auratus* population. North Island sample sites included Doubtless Bay (DB), East Coast (EC), Foxton Beach (FB), Hawkes Bay (HB), and North Taranaki Bight (NTB). South Island sample sites included Tasman Bay (TB), Marlborough Sounds (MS), and North Farewell Spit (NFS), (Figure 2.1.0). Sample collecting was conducted by staff at Plant and Food Research (Nelson), as they were able to access fish caught on commercial trawlers and from

recreational fishers. A fin clip was removed from each fish and stored in 85% ethanol at 4°C. Total genomic DNA was extracted using a standard phenol-chloroform protocol (Sambrook *et al.* 1989). The small sample of tissue was individually placed in 400 µL of extraction buffer (10 mM Tris pH 8.0, 50 mM NaCl, 10 mM EDTA, and 0.2% SDS) with 0.5 µg/µL proteinase-K and incubated overnight at 50°C. Following the tissue digestion, DNA was extracted with phenol, followed by chloroform/isoamyl alcohol, and precipitated with 1 mL of ethanol at -20°C for 1 hour. The DNA was pelleted, washed with 70% ethanol, dried and then re-suspended in a Tris-EDTA buffer. The DNA was stored at 4°C. The purified DNA was quantified using a Nanodrop spectrophotometer and the concentration was adjusted to 200 ng/µL using distilled water.

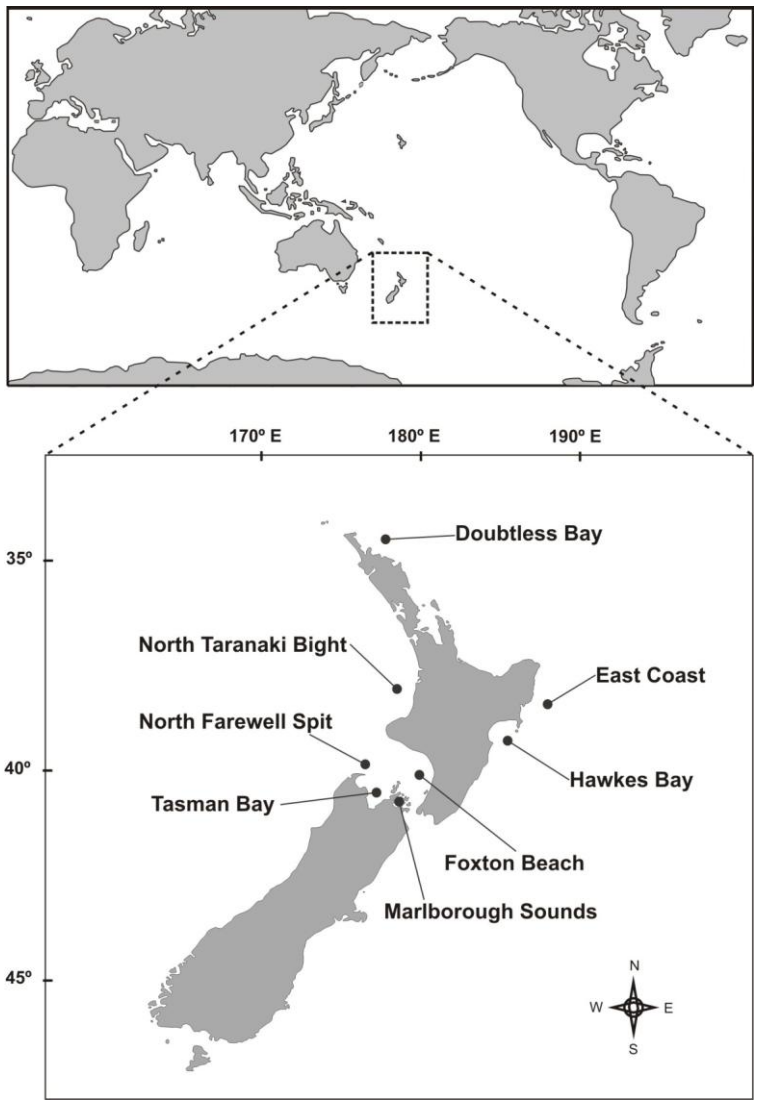


Figure 2.1.0: New Zealand *P. auratus* Samples Sites.

2.2.3: Mitochondrial DNA Amplification and Sequencing

A 641 bp portion of the control region was amplified using the Polymerase Chain Reaction (PCR) and the primers tRNA-Pro and CR-CCD. All PCRs consisted of 10.4 mM Tris-HCl pH 8.8, 52 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.4 µM of each primer, 0.6 µg/µL Bovine Serum Albumin (BSA), 0.5 units of Taq DNA Polymerase (Fisher Scientific), and 200 ng of template DNA. The conditions for thermal cycling were: 33 cycles of 94^{o}C for 30 seconds, 58^{o}C for 30 seconds, and 74^{o}C for 60 seconds, followed by an extension step of 74^{o}C for 10 minutes. The resultant PCR products were purified with ExoSAP-IT (Amersham Parmacia Biotech) and their DNA sequence was determined using the reverse primer and an ABI 3730 Genetic Analyzer (Macrogen, Korea).

2.2.4: Microsatellite DNA Genotyping

Microsatellite loci were amplified by PCR and the allele sizes were determined using an ABI3730 DNA Genetic Analyzer (Massey Genome Services). All PCRs were carried out in 10µL volumes containing 10mM Tris-HCl, 50mM KCl, 200 µM of each dNTP, 0.1µM forward primer, 0.4µM reverse primer, 0.1µM M13 primer, 0.6 µg/µL BSA, 1 unit Fisher *taq* polymerase, and 1 µl of the template DNA. The $MgCl_2$ concentrations for each locus are shown in Table 2.3. The thermal cycling conditions used: were 33 cycles of 94^{o}C for 20 seconds, annealing temperature (see Table 2.1.2 for each locus) for 30 seconds, and 74^{o}C for 60 seconds, followed by 8 cycles of 94^{o}C for 20 seconds, 53^{o}C for 30 seconds, and 74^{o}C for 60 seconds, followed by a final extension step of 74^{o}C for 10 minutes. FAM, VIC, or PET fluorescent labels were incorporated into the PCR products in the amplification step using the M13 labeling method (Schuelke, 2000). Each microsatellite locus was amplified separated and then combined with other amplified loci ('poolplexed') for genotyping; there were either two or three loci in each well on a 96-well plate format (Table 2.1.2).

2.2.5: Mitochondrial DNA Sequence Analysis

DNA sequences were aligned in MEGA 5.0 using a ClustalW alignment with default settings (Tamura *et al.* 2011). Variable sites were checked for possible base calling errors. The final alignment was exported in a FASTA format. The number of segregating sites (*S*), haplotypes (H), private haplotypes (H_P), haplotype (*h*) and nucleotide (π) diversity, and average number of pairwise differences (*k*) were estimated using DNASP 5.0 (Rozas *et al.* 2003). ARLEQUIN

Table 2.1.2: Microsatellite PCR Concentrations and Temperatures.

Poolplex	Locus	T-DNA (ng)	$MgCl_2$ (mM)	T-Ann (^oC)
1	Pma1	100	2.0	60
1	GA2A	150	2.0	60
1	GT2	150	1.5	57
2	Pma2	150	1.5	64
2	Pma4-32	150	2.0	58
2	GT4	150	1.5	59
3	Pma3	150	1.5	57
3	Pma5	150	1.5	59

T-DNA = Template DNA, T-Ann = Annealing Temperature

3.5 was used to calculate the fixation index (Φ_{ST}) for pairwise comparisons between sample sites and 1000 permutations of the data set were used to determine statistical significance (Excoffier *et al.* 2005).

HP-RARE 1.1 was used to construct rarefaction curves for numbers of haplotypes sampled at all sites (Kalinowski 2005). This analysis enabled a comparison of the haplotypes numbers in each population that was unbiased by variation of sample sizes. Two maximum likelihood (ML) haplotype trees were calculated using the PHYML 3.0 (Guindon *et al.* 2010) and viewed with HAPLOVIEWER (Ewing 2012). The best model for constructing the trees was determined using the J-MODEL-TEST 0.1.1 (Posada 2008). The ML tree constructed using the 641bp sequence data set was calculated using the K80 model + G and a fixed proportion of invariable sites of 0.5820; all other settings were left on default. The ML tree constructed using the 460bp sequence data set of both *P. auratus* and *P. major* was calculated using the K80 model + G and default settings for all other settings.

A hierarchical analysis of molecular variance (AMOVA) was conducted between regions using ARLEQUIN 3.5. Five different regional groupings were tested with each region being composed of two or more populations. First, the populations were grouped into North Island and South Island regions. Secondly, the populations were grouped into South Island, mid North Island, and upper North Island; based on the suggested populations in Bernal-Ramı´rez 2003. Thirdly, the populations were grouped into three regions based on the branching on the microsatellite Nei's DA neighbor joining tree (Figure 2.2.3); region 1: North Farewell Spit, Marlborough Sounds, and Foxton Beach, region 2: Hawkes Bay, Tasman Bay, and Broodstock, region 3: North Taranaki Bight, East Coast, and Doubtless Bay. For the fourth and fifth analysis the populations were grouped into regions of 3 or 2 populations,

respectively; these groupings were constructed to reduce the overall Φ_{ST} for each region, based on the Φ_{ST} values in table 2.1.4.

A test for a pattern of isolation-by-distance was conducted using a regression plot of $\Phi_{ST}(1-\Phi_{ST})^{-1}$ and shortest marine geographic distance was constructed in EXCEL and tested in ARLEQUIN 3.5 using a Mantel's test with 10,000 repetitions. The shortest marine geographic distance between sample sites was measured using Google Earth 6.2.2 (Google Inc.).

Fu's Fs and Tajima's D statistics were calculated using ARLEQUIN 3.5; 1000 permutations were used to determine the statistical significance. A mismatch distribution of the DNA sequences from all populations (excluding the sample from the cultured fish) was constructed using DNASP 5.0, a sudden expansion model for the expected distribution was fitted to the data using 1000 repetitions. Theta initial, Theta final, and Tau for the expected distribution were calculated in ARLEQUIN 3.5 and set to 3.60, 21.95, and 344.0 respectively. The significance of deviation from the expected distribution was calculated in Arlequin 3.5 for the total data set and individual populations.

BEAST 1.7.4 was used to calculate two skyline plots for the total pooled data set (Ho & Shapiro 2011; Drummond & Rambaut 2007); the mutation rate of $2.0\text{x}10^{-7}$ and $2.0\text{x}10^{-8}$ were used (Brown *et al.* 1993; Bowen *et al.* 2006; Mccusker & Bentzen 2010; Padhi 2011). A Marko Chain Monte Carlo simulation was run for 10^{-7} iterations using a strict molecular clock, the HKY mutation model, and the stepwise skyline model. TRACER 1.5 was used to construct skyline plots from the Beast output files (Drummond & Rambaut 2007).

2.2.6: Microsatellite DNA Genotype Analysis

Allele peaks were viewed and analyzed using PEAKSCANNER 1.0. The allele size data was exported to an EXCEL worksheet and the raw peak values were manually placed in to size 'bins' that were set for each allele size. The size data was arranged into a GENPOP file format. MICROCHECKER 2.2.3 was used to check for null alleles, allele size shifts and scoring errors (Oosterhout *et al.* 2004).

Allele frequencies for each locus and population were plotted using the Excel addon GENALEX 6.41 (Peakall & Smouse 2006, 2012). The allele frequency plots were also used to construct graphs of allele frequency distribution. Rarefaction curves were plotted for all loci

and then presented as an average for each population using the "PopGenKit" package in R 2.15 (Rioux Paquette 2011).

The number of alleles (A), observed heterozygosity (H_O), and expected heterozygosity (H_E) for all sampled sites were determined using ARLEQUIN 3.5. Deviations from Hardy-Weinberg by population for each locus were detected using the Markov chain Exact test with 10^5 dememorization steps and 10^6 forecast chain length in ARLEQUIN 3.5. A test for pairwise Linkage Disequilibrium (LD) between loci within and among all populations was calculated in ARLEQUIN 3.5 using 10,000 permutations.

Wright's fixation index (F_{ST}) for pairwise comparison between all sites was estimated using ARLEQUIN 3.5. Significance at the 5%-level was determined using 20,000 permutations (Wright 1951). Weir and Cockerham's fixation index (Theta) was measured for all pairwise comparisons between sites in FSTAT 2.9.3.2 (Goudet 2005; Weir & Cockerham 1984). Regression plots were constructed of Theta and F_{ST} versus shortest marine distance between sites (in kilometers) and tested in ARLEQUIN 3.5 using a Mantel's test with 10,000 replicates. Two neighbor-joining (NJ) trees were constructed using 1000 bootstrap replicates in the program POPULATIONS 1.2.31 and viewed with the program TREEVIEW (Langella 2012). Nei's Da and Reynolds weighted distance was used to construct the NJ tree (Nei *et al.* 1983; Takezaki & Nei 1996; Reynolds *et al.* 1983).

A Bayesian inference of population structure and assign individuals to expected populations was made using STRUCTURE 2.3.3 (Pritchard *et al.* 2000). The expected number of populations (K) was allowed to range from 1-8, with 10 repetitions, and the Markov chain Monte Carlo (MCMC) simulations were run for 10^7 iterations with a burn-in time to 4^5 step. STRUCTURE HARVESTER 0.6.92 was used to determine the most likely value for K (Dent & Bridgett 2012) by applying the Delta K estimator approach (Evanno *et al.* 2005). CLUMPP 1.1.2 was used to concatenate the results of 10 runs at the suggested K value (3) and determine each individual's proportion of membership to each K (Jakobsson & Rosenberg 2007).

A Principle Component Analysis (PCA) was conducted using PCAGEN 1.2.1 (Goudet 1999) and used to investigate the possible clustering and separation of populations. GENETIX 4.05 was used to conduct a Factorial Correspondence Analysis (FCA) of individuals and

populations (Belkhir 1999). Two graphs were constructed for the FCA based on combined individual and population data to look at the three main axes.

GENECLASS2 (Piry *et al.* 2004) was used to conduct individual assignment tests to determine which population individuals were best assigned, with a percentage measure used to determine the likelihood of population assignment. A plot of heterozygosity against F_{ST} was constructed in LOSITAN for each microsatellite locus (Beaumont & Nichols 1996; Antao *et al.* 2008) using 100,000 simulations, 99.5% confidence interval, and both the infinite allele model (IAM) and the stepwise mutation model (SMM).

To test for a genetic bottleneck significant deviation of heterozygosity from the expected values was tested by population and locus using Hardy-Weinberg Exact test implemented in GENEPOP 4.0.10 with 1000 dememorizations, 100 batches, and 1000 iterations per batch were used (Raymond & Rousset 1995). Hardy-Weinberg Exact tests for significant heterozygosity deficiency were run with and without the Pma4-32 locus present. Population plots of allele frequency distribution were constructed for all microsatellite alleles using the output from the Excel add-on GENALEX 6.41. M_P_VAl was used to calculate the M-ratio for all loci and populations and whether the M-ratios were significantly below the expected value (Garza and Williamson 2001). Statistical significance was determined by 1000 iterations of the expected value. The program BOTTLENECK 1.2.02 was used to compare of the observed heterozygosity and expected heterozygosity based on the observed allele numbers (Maruyama & Fuerst 1985; Cornuet & Luikart 1997). Significant deviation under the Stepwise Mutation Model (SMM) and the Two Phase Mutation model (TPM) were tested with the Wilcoxon's test based on 1000 replications. 95% single-step mutations and 5% multiple-step mutations were used for the TPM model (Piry *et al.* 1999). BOTTLENECK 1.2.02 was run with and without the Pma4-32 locus in the data set.

2.3: Results

2.3.1: Mitochondrial DNA Control Region Sequencing

The DNA sequence of the control region was determined from 364 fish, sampled from eight sample sites and one population in captive culture. The aligned DNA sequence data set was 641 base pairs and the average nucleotide composition was π_A = 0.315, π_T = 0.308, π_G = 0.205, π_C = 0.171. There were 85 polymorphic sites and two indels in the data set. Overall nucleotide diversity (π) was 0.0310 (Table 2.1.3). There were 88 haplotypes and the average haplotypic diversity (*h*) was 0.858 (Table 2.1.3). A range of haplotype frequencies were found and the three most frequent haplotypes were at 0.33, 0.15, and 0.10. The number of mutational steps between the two most frequent haplotypes was 44, while the largest number of mutational steps between haplotype was 55 differences.

Table 2.1.3: Mitochondrial Sequence Summary Statistics for all Sample Sites

Sample Site	*n*	*S*	H	H_p	*k*	*h*	π
Doubtless Bay	49	58	21	10	20.23	0.867	0.0317
East Coast	45	54	19	7	21.24	0.882	0.0333
Foxton Beach	27	54	13	5	18.96	0.815	0.0297
Hawkes Bay	51	60	23	13	19.66	0.910	0.0308
Marlborough Sounds	25	54	14	5	21.28	0.883	0.0334
North Farewell Spit	49	57	20	10	20.90	0.889	0.0328
North Taranaki Bight	44	50	18	10	19.21	0.829	0.0301
Tasman Bay	50	52	15	6	19.19	0.828	0.0301
Total	**364**	**85**	**88**	**68**	**19.73**	**0.858**	**0.0310**

n = number of samples, S = variable sites, H = number of haplotypes, HP = number of private haplotypes, h = haplotype diversity, k = average number of pairwise nucleotide differences, π = nucleotide diversity

The rarefaction plot of haplotype numbers showed that when corrected for the differences in sample size, the haplotype numbers fall within a relatively narrow range (Figure 2.1.1). One exception to this trend was the Tasman Bay, which had lower haplotype numbers relative to other sample sites (Figure 2.1.1).

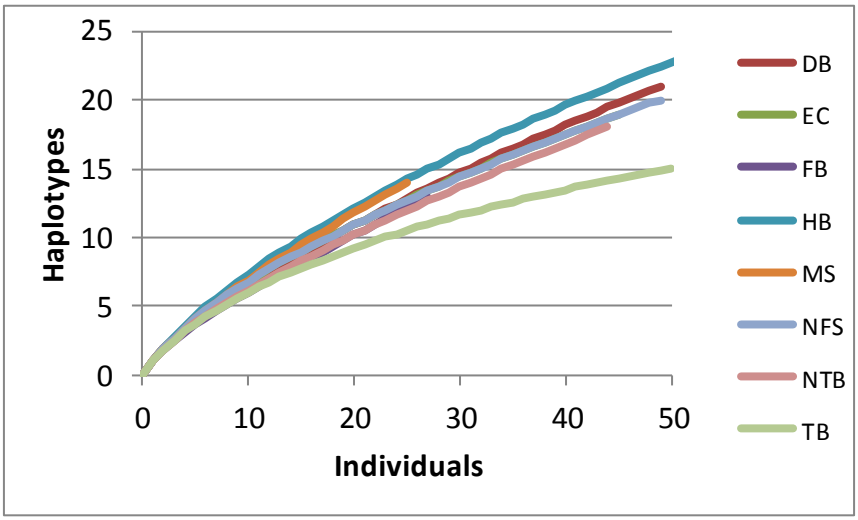


Figure 2.1.1: Haplotype Rarefaction Plot for *P. auratus* Sample Sites.

At the 5%-level there was only one significant Φ_{ST} comparison (PFR broodstock to East Coast), but none were significant after a sequential Bonferroni correction (Table 2.1.4). The uncorrected significance could be due to the small sample size of the PFR broodstock data set (n = 24). The AMOVA analysis of the grouped sites showed no significant difference between all five regional comparisons (Table 2.1.5) and over 99% of the variation was found within sample sites. Only a very small amount of variation was found between regions with the fourth and fifth regional groupings, however this was not significant.

Table 2.1.4: Pairwise Φ_{ST} for all Sample Sites.

	DB	EC	FB	HB	MS	NFS	NTB	TB
DB	-	-0.00387	-0.01632	-0.00989	-0.01863	-0.01559	-0.01447	-0.0136
EC	0.42383	-	0.02056	0.01261	-0.02565	-0.00756	-0.00029	0.00639
FB	0.64648	0.19141	-	-0.01577	-0.00151	-0.0086	-0.00432	-0.01634
HB	0.61328	0.18164	0.64551	-	0.00108	-0.00598	0.00818	0.0025
MS	0.69727	0.91211	0.33887	0.30762	-	-0.01627	-0.01917	-0.0181
NFS	0.86133	0.50781	0.45801	0.45703	0.61816	-	-0.00884	-0.00515
NTB	0.75586	0.33008	0.34961	0.23242	0.65234	0.5332	-	-0.01426
TB	0.75488	0.24609	0.63281	0.32031	0.65527	0.47754	0.72559	-

DB = Doubtless Bay, EC = East Coast, FB = Foxton Beach, HB = Hawkes Bay, MS = Marlborough Sounds, NFS = North Farewell Spit, NTB = North Taranaki Bight, TB = Tasman Bay

Table 2.1.5: AMOVA Within and Between Regional Groupings.

	Group	Regions	df	Var	%Var	$\Phi_{CT/SC/ST}$	p-value
Among Regions	**1**	NI/SI	1	-0.00089	-0.20	-0.00205	0.38416
	2	SI/mNI/uNI	2	0.00015	0.04	0.00036	0.48583
	3	M-Tree	2	0.00001	0.00	0.00003	0.53079
	4	Reduced ΦST 3	2	0.00079	0.18	0.00183	0.25904
	5	Reduced ΦST 2	3	0.00217	0.50	0.00501	0.06256
Among Sites Within Regions	**1**	NI/SI	6	0.00080	0.19	-0.00020	0.31672
	2	SI/mNI/uNI	5	0.00021	0.05	0.00048	0.40958
	3	M-Tree	5	0.00032	0.07	0.00074	0.36950
	4	Reduced ΦST 3	5	-0.00026	-0.06	-0.00060	0.49267
	5	Reduced ΦST 2	4	-0.00153	-0.35	-0.00356	0.71652
Within Sites	**1**	NI/SI	332	0.43253	100.02	0.00185	0.86901
	2	SI/mNI/uNI	332	0.43253	99.92	0.00084	0.36755
	3	M-Tree	332	0.43253	99.92	0.00076	0.38221
	4	Reduced ΦST 3	332	0.43253	99.88	0.00124	0.38416
	5	Reduced ΦST 2	332	0.43253	99.85	0.00147	0.37830

df = degrees of freedom. Var = Variation between regions or sites, %Var = % variation between regions or sites, $\Phi_{CT/SC/ST}$ = genetic distance between regions or sites.

The regression plot of $\Phi_{ST}(1-\Phi_{ST})$-1 versus shortest marine distance (km) between populations was not significant when tested with a Mantel's test (R2= 3.0 x 10-5, p-value = 0.5316)(Figure 2.1.2).

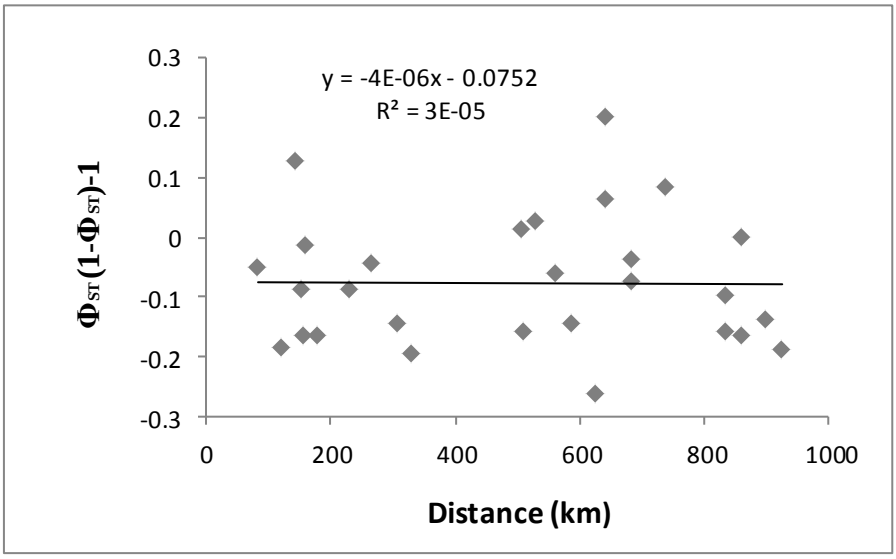


Figure 2.1.2: $\Phi_{ST}(1-\Phi_{ST})^{-1}$ versus Shortest Marine Distance (km).

The unrooted maximum likelihood (ML) tree of the 88 haplotypes shows that there are a large number of mutational steps between the major haplotypes (Figure 2.1.6); there is 44 mutations between the two largest haplotype groups. There was no distinct pattern to the sample sites on the haplotype tree; rather the sample sites are spread evenly across all the different haplotypes. The haplotype frequency plots show similar distributions in all populations (Figure 2.1.8). Haplotypes that are found in all sample sites are listed in the legend, from highest to lowest frequency. A second unrooted maximum likelihood (ML) tree constructed with 460 bp *P. auratus* and *P. major* sequences shows how the *P. auratus* species has branched off from *P. major* into two major haplogroups (Figure 2.1.7). The number of mutational steps between the two main *P. auratus* haplotypes, the farthest *P. auratus* haplotypes, the closest main *P. auratus* and *P. major* haplotypes is 27, 41, and 29, respectively (Figure 2.1.7).

The mismatch distribution for all wild snapper sample sites combined had a predominantly bimodal distribution and was significantly different from the expected distribution under an expansion model (p-value = 0.012) (Figure 2.1.3; Table 2.1.6).

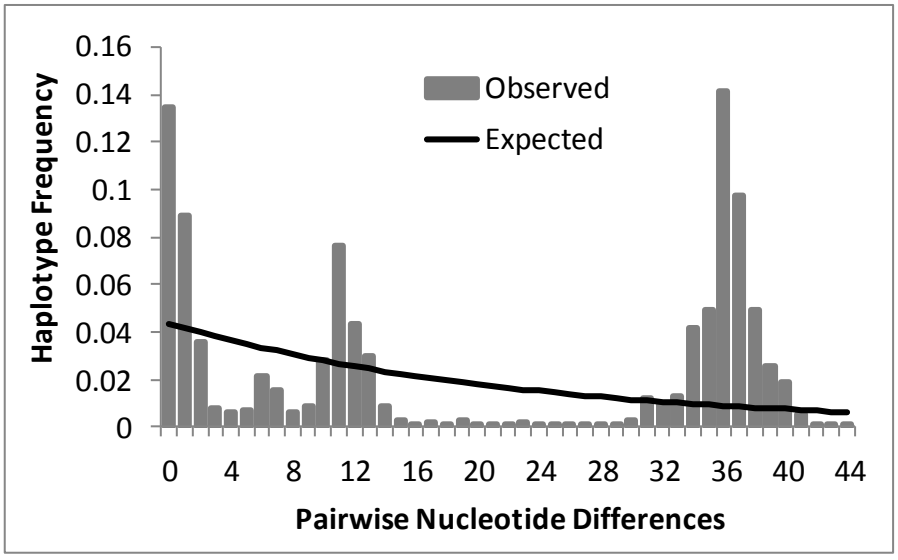


Figure 2.1.3: Mismatch Distribution for *P. auratus* Population.

Fu's Fs statistic was positive and significant for Tasman Bay and the PFR broodstock, indicating that these two samples have gone through recent bottlenecks (Fu 1997) (Table 2.1.6). The significant Fu's Fs statistic in the PFR broodstock is probably the result of a

founder effect when removed from their original population. Tajima's D statistic was positive and significant for all populations except Foxton Beach and the PFR broodstock (Table 2.1.6). This could indicate that these samples have gone through genetic bottlenecks (Tajima 1989) or that it this statistic is being affected by secondary contact between two historical populations (see Discussion).

Table 2.1.6: Demographic Statistics, Sum of Squared Deviations for the Mismatch Distribution, and Harpending's Raggedness Index.

Sample Site	*Fs*	*D*	*R*	$R_{P\text{-}value}$	*SSD*	$SSD_{P\text{-}value}$
Doubtless Bay	2.697	**1.935***	0.0340	0.049	0.0508	0.024
East Coast	3.813	**2.532***	0.0244	0.017	0.0391	0.004
Foxton Beach	3.698	1.347	0.0861	0.002	0.0817	0.017
Hawkes Bay	1.679	**1.649***	0.0243	0.026	0.0265	0.053
Marlborough Sounds	2.727	**1.885***	0.0480	0.011	0.0561	0.004
North Farewell Spit	3.579	**2.213***	0.0327	0.002	0.0534	0.003
North Taranaki Bight	3.542	**2.358***	0.0556	0.001	0.0773	0.006
Tasman Bay	**7.838***	**2.259***	0.0344	0.007	0.0614	0.009
Total	-11.169	1.538	0.0259	0.019	0.0507	0.012

* = Significant at 95%, Fs = Fu's Fs, D = Tajima's D, R = Raggedness index, $R_{P\text{-}value}$ = Raggedness index p-value, SSD = sum of squared standard deviations, and SSDp-value = sum of squared standard deviations p-value

The bayesian skyline plots indicated that a reduction in population size has occured in the New Zealand *P. auratus* followed by a period of population expansion (Figure 2.1.4 and Figure 2.1.5). With the mutation rate of $2.0x10^{-7}$ and $2.0x10^{-8}$ the reduction in population size occurrs at around 3,500 and 35,000 years ago, respectively. In both plots, prior to the reduction the population sizes appears to have remained constant for a long period of time.

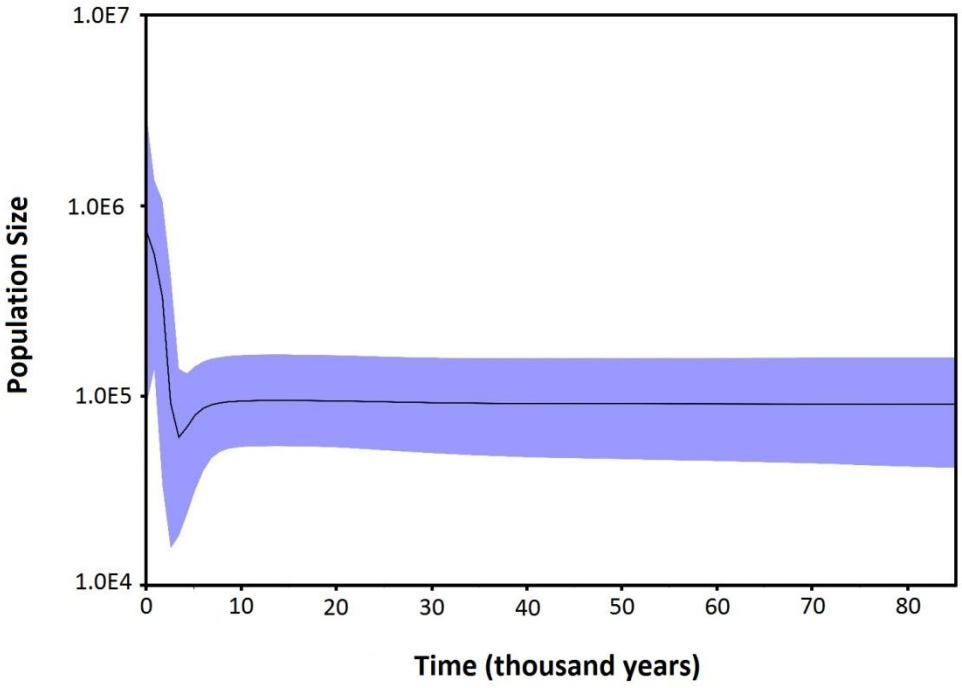


Figure 2.1.4: Bayesian Skyline Plot for $2.0x10^{-7}$ Mutation Rate.

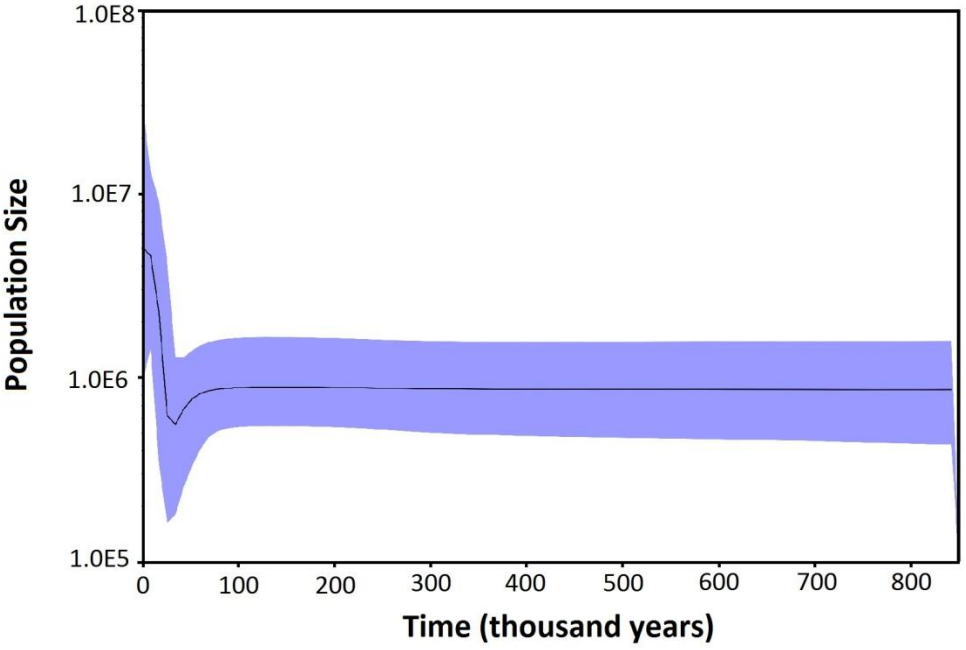


Figure 2.1.5: Bayesian Skyline Plot for $2.0x10^{-8}$ Mutation Rate.

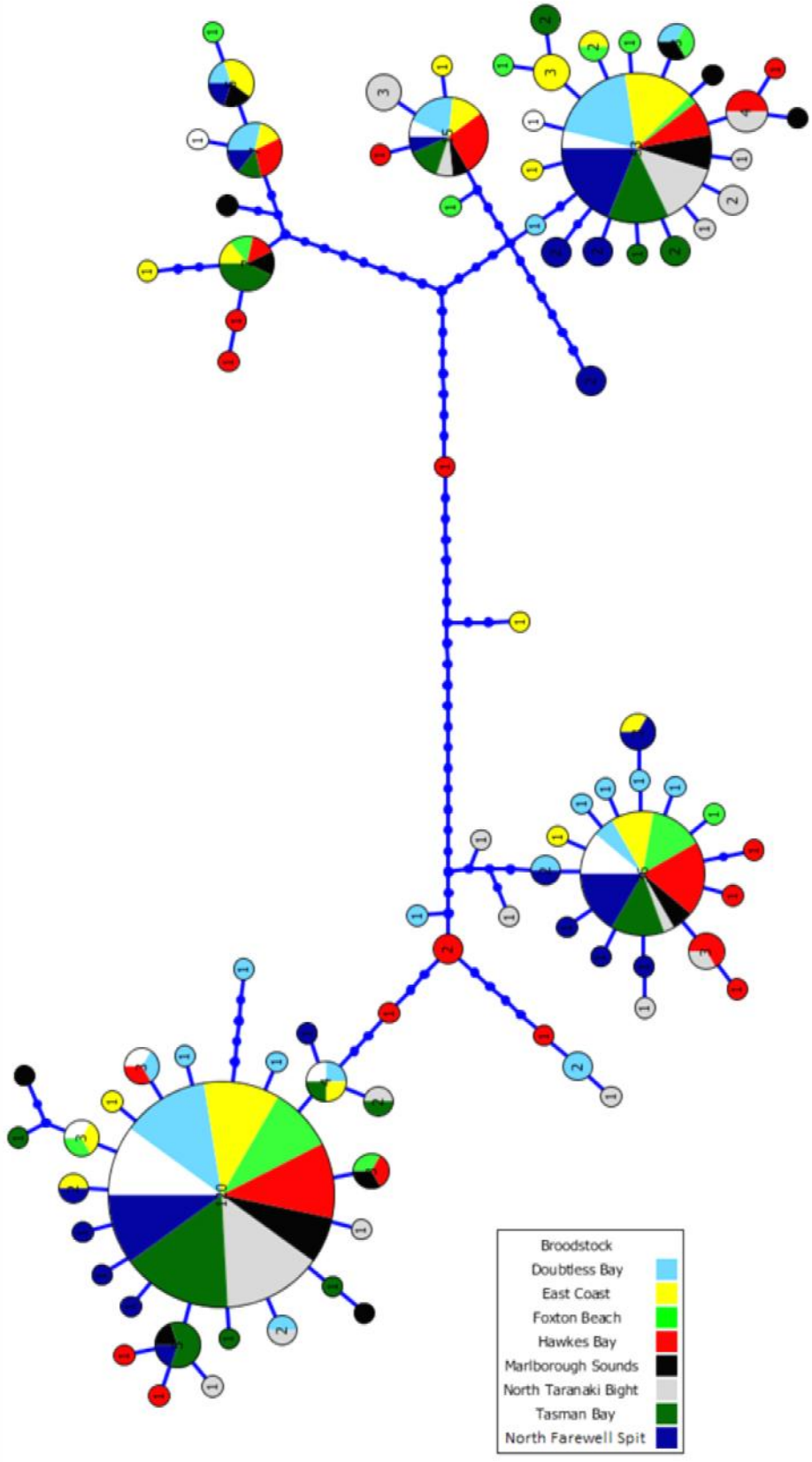


Figure 2.1.6: Maximum Likelihood Haplotype Tree by Sample Site for 641 bp mtDNA Sequence.

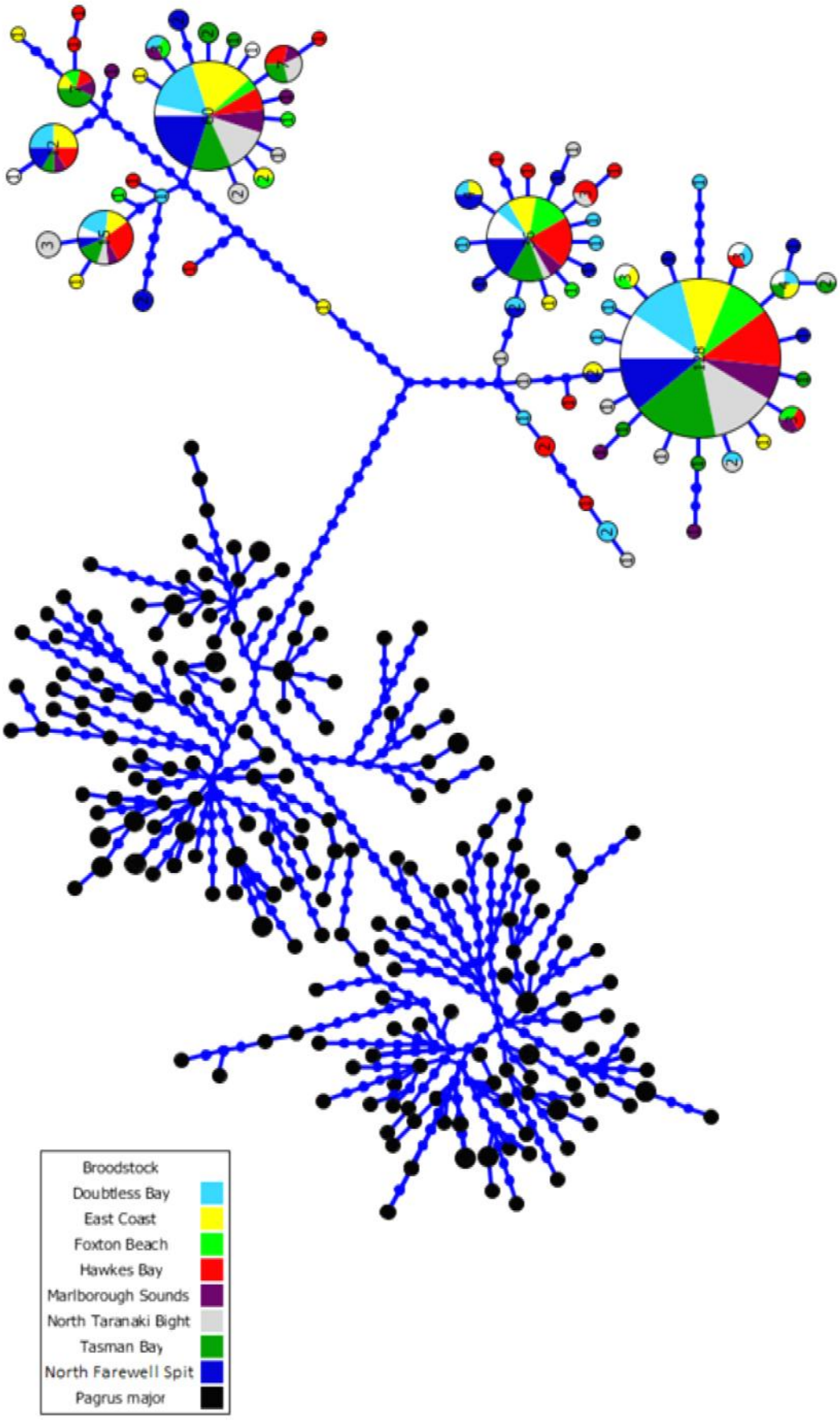


Figure 2.1.7: Maximum Likelihood Haplotype Tree by Sample Site for 460 bp mtDNA Sequence.

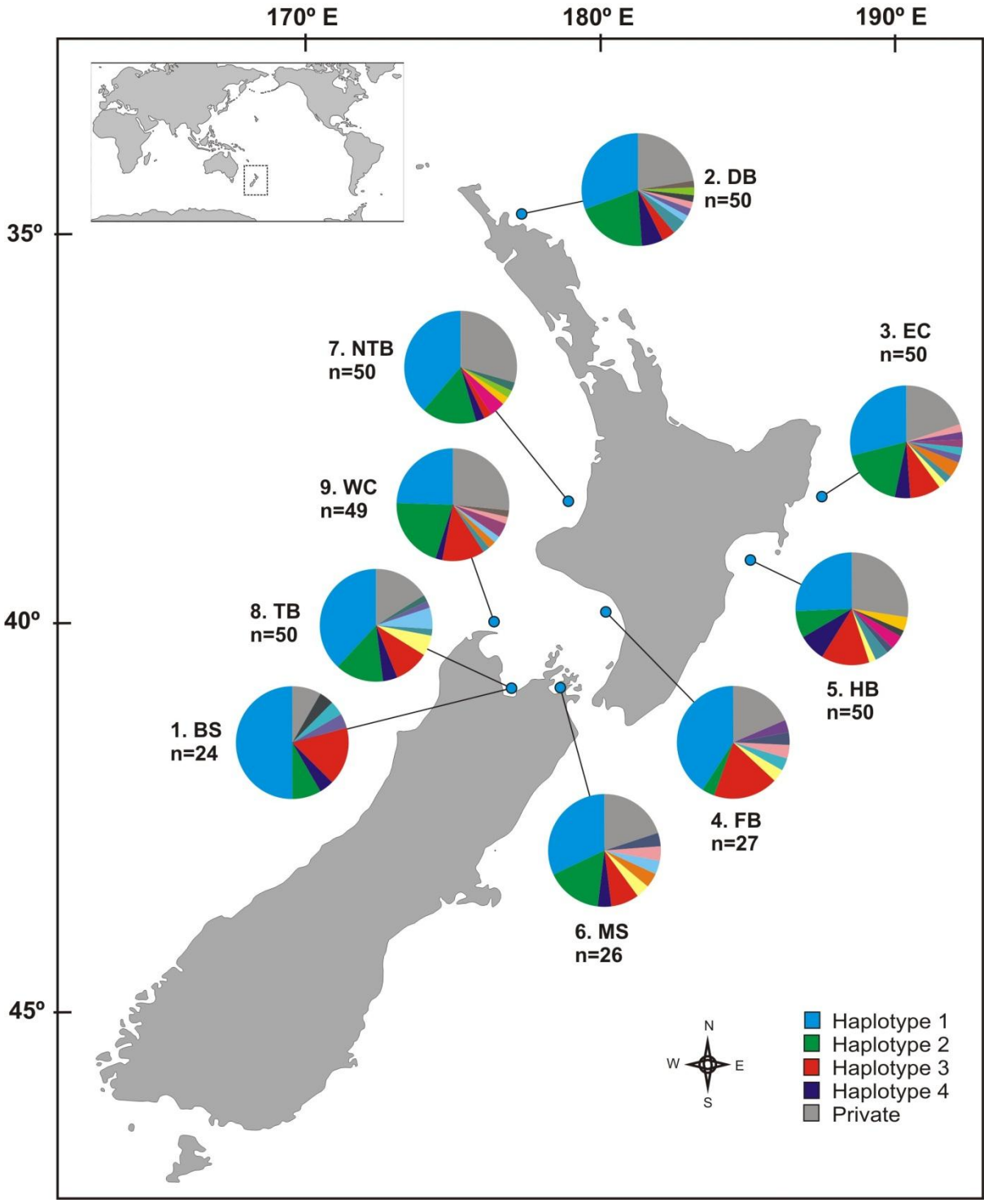


Figure 2.1.8: Haplotype Frequency Distributions by Sample Site.

2.3.2: Microsatellite DNA genotyping

The eight microsatellite DNA loci were successfully genotyped for 374 fish samples. The Pma4-32 locus probably had null alleles and was significantly deviated from Hardy-Weinberg equilibrium in all sample sites. However, when Pma4-32 was excluded from the analyses it did not affect the results in most cases, any analyses that did show a difference when this locus was included have been reported with and without this locus. The results from the LOSITAN outlier analysis showed that all eight microsatellite loci fell within the 99% confidence interval range under both the Infinite Allele Model (IAM) and Stepwise Mutation Model (SMM), and there was no evidence to reject null hypothesis of neutrality for all loci. No linkage dis-equilibrium was observed for any pairs of loci that consistently appeared across all sample sites.

The average number of alleles per locus was 10.11 per sample site, and ranged from 9.00 to 11.13 (Table 2.1.7). The rarefaction curves for all loci and sample sites had started to flatten (Figure 2.1.9), which indicated that the sample sizes used in the study were able to sample most of the variation.

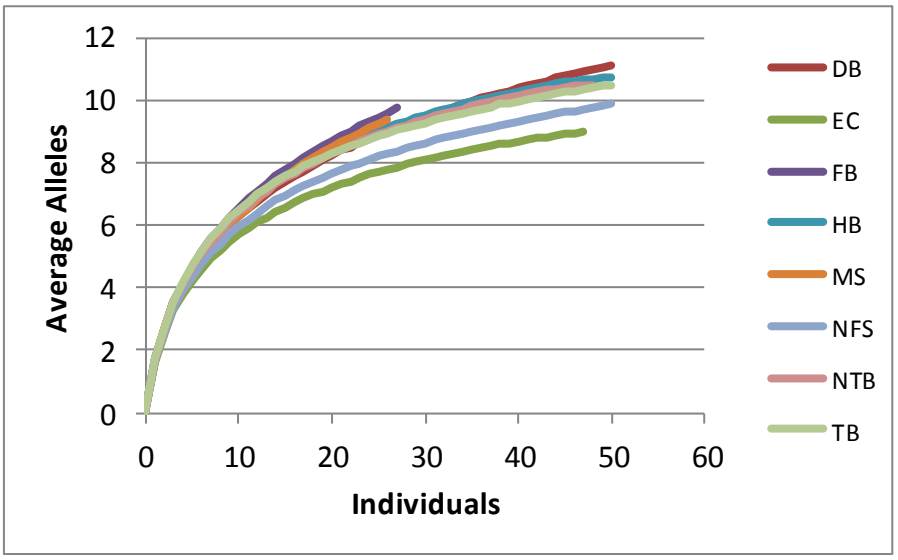


Figure 2.1.9: Rarefaction Curve for Average Microsatellite Alleles by Sample Site.

Table 2.1.7: Microsatellite Allelic Diversity and Heterozygosity.

Sample Site	N	He	Ho	A_{Avg}	A_P
Doubtless Bay	50	0.747	0.723	11.13	2
East Coast	49	0.722	0.650	9.00	5
Foxton Beach	27	0.751	0.690	9.75	2
Hawkes Bay	49	0.730	0.724	10.75	7
Marlborough Sounds	26	0.708	0.654	9.38	5
North Farewell Spit	49	0.718	0.672	9.88	1
North Taranaki Bight	49	0.736	0.743	10.50	3
Tasman Bay	49	0.748	0.702	10.50	6
Total / Average	**374**	**0.733**	**0.695**	**10.11**	**3.6**

Weir and Cockerham's F-statistic (Theta) and Wright's F-statistic (F_{ST}) among sample sites showed similarly patterns of genetic differentiation (Table 2.2.2 and Table 2.2.3). Marlborough Sounds to Doubtless Bay pairwise comparisons had the highest level of differentiation for both F_{ST} and Theta (both 0.018); followed by Marlborough Sounds to Foxton Beach (0.015 & 0.014). The third highest level of differentiation for both F_{ST} and Theta was Marlborough Sounds to Hawkes Bay (both 0.013), followed closely by Foxton Beach to Hawkes Bay (0.012 & 0.013). The fifth highest level of differentiation for both F_{ST} and Theta was the North Farewell Spit to Doubtless Bay (0.012 & 0.011). The only F_{ST} comparisons significant after sequential Bonferroni correction were the Marlborough Sounds to Doubtless Bay and North Farewell Spit to Doubtless Bay. The Marlborough Sounds to East Coast, North Taranaki Bight to Foxton Beach, Marlborough Sounds to North Taranaki Bight, and North Farewell Spit to Tasman Bay comparisons were significant for F_{ST}, but not for Theta.

The regression plot of $F_{ST}(1- F_{ST})$ against the shortest marine distance (km) between sample sites showed no significant correlation ($R^2 = 0.0075$) (Figure 2.2.0) and Mantel's test showed not significant relationship either (p-value = 0.5323).

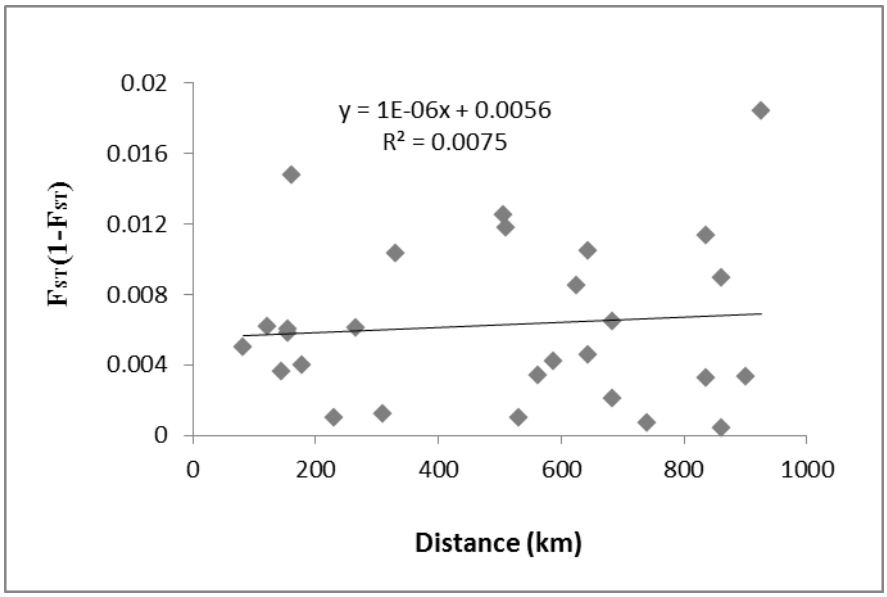


Figure 2.2.0: $F_{ST}(1-F_{ST})$ Versus Shortest Marine Distance (km).

The STRUCTURE 2.3.3 analysis showed that there were either two or three groups (K) in given the data set based on the delta k values, but three groups were most likely (Figure 2.2.1). The three expected genotype groups were present at all sample sites but in different frequencies. Overall, the analysis indicated that the New Zealand *P. auratus* population was panmictic (Figure 2.2.2).

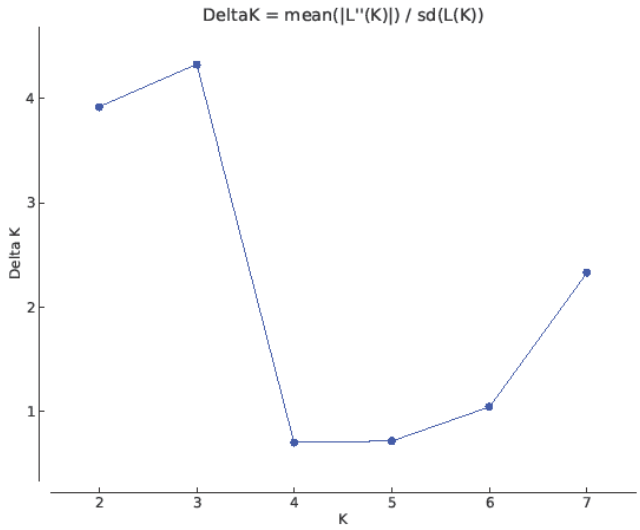


Figure 2.2.1: Delta K for Each Expected Population (K).

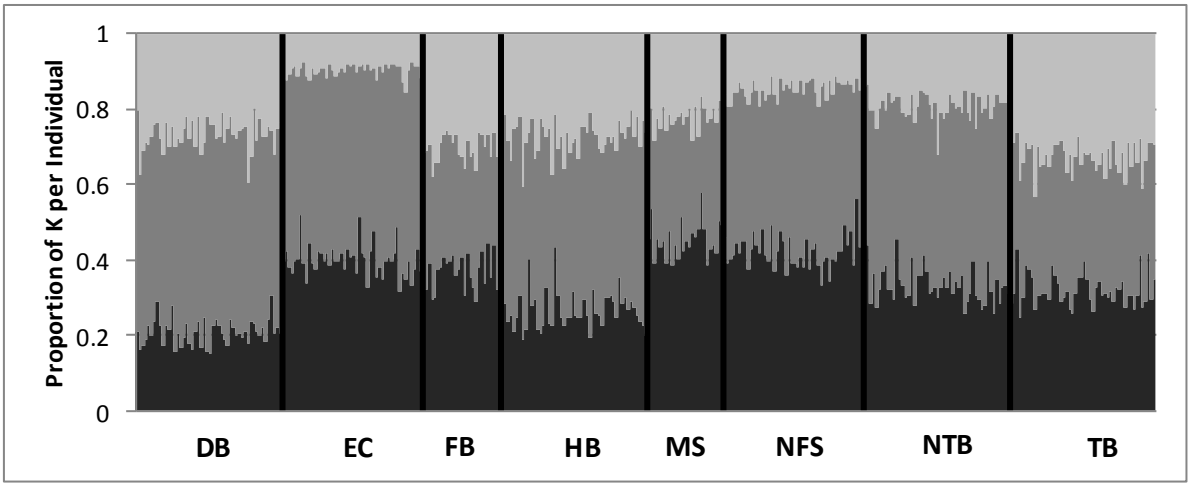


Figure 2.2.2: Results for STRUCTURE Analysis of *P. auratus* Sample Sites.

The neighbor-joining (NJ) analysis showed a tree with a starburst-like structure and there did not appear to be any distinctive structure within the data set (Figure 2.2.3 and Figure 2.2.4). The trees that were constructed with Reynolds weighted distance and Nei's DA distance showed very different topologies.

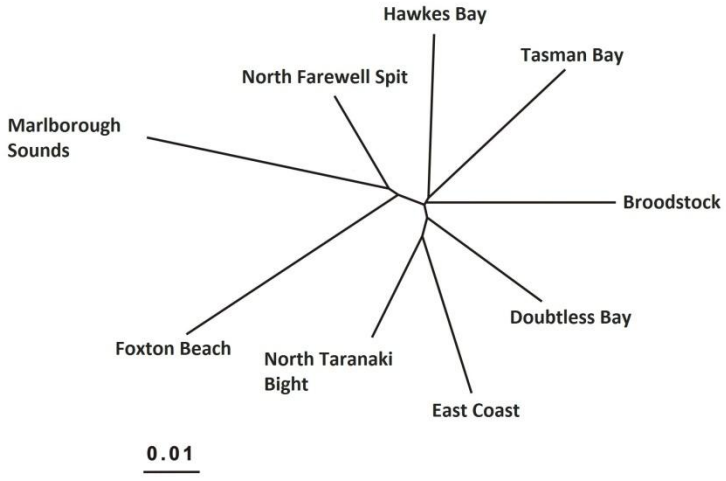


Figure 2.2.3: Nei's Da Distance Neighbor-Joining Tree.

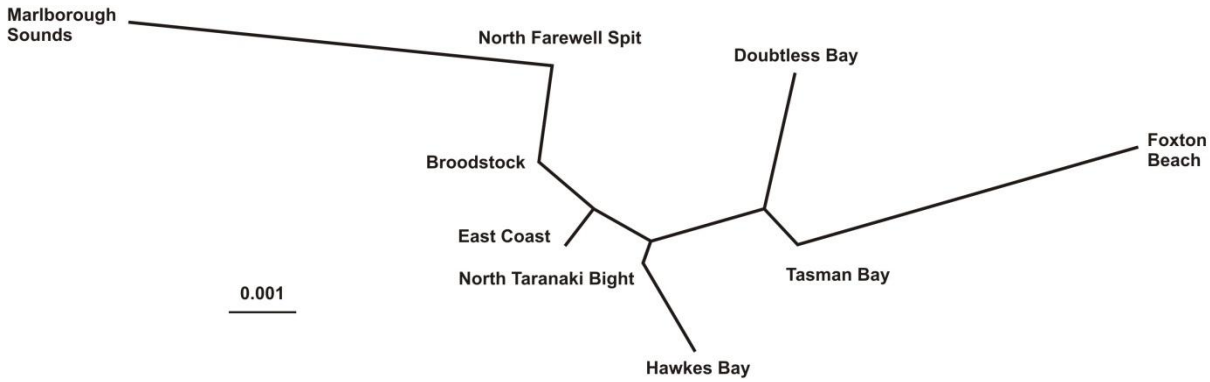


Figure 2.2.4: Reynolds Weighted Distance Neighbor-Joining Tree.

The individual assignment test found that on average 65% of individuals were able to be assigned to their correct sample site, with the highest rate being 81% (Marlborough Sounds) and the lowest 48% (Doubtless Bay) (Table 2.1.8).

Table 2.1.8: Percentage of Individuals from each Sample Site (columns) Assigned to each Sample Site (rows) with correctly assigned individuals in **bold**.

		Individuals							
		DB	**EC**	**FB**	**HB**	**MS**	**NTB**	**TB**	**WC**
Sample Site	**DB**	**48**	0	0	6	4	6	8	2
	EC	12	**77**	4	4	0	8	2	4
	FB	6	4	**78**	4	0	2	4	2
	HB	6	4	0	**60**	4	4	4	4
	MS	4	4	4	6	**81**	4	4	4
	NTB	4	2	7	4	0	**52**	6	10
	TB	8	4	0	6	0	8	**66**	4
	WC	4	0	4	2	8	8	0	**66**

No significant differences between populations was found for any of the 8 axis for the principle component analysis when tested in GENETIX 4.05. The combined population and individual factorial correspondence analysis (FCA) showed that the sample sites were largely overlapping with a few individuals in most sample sites spreading out near the edges of the distribution (Figure 2.2.5 and Figure 2.2.6). This result fits well with the other analyses, which indicate a largely panmictic population. Overall, the three main axes explain 49.68% of the variation in the data set.

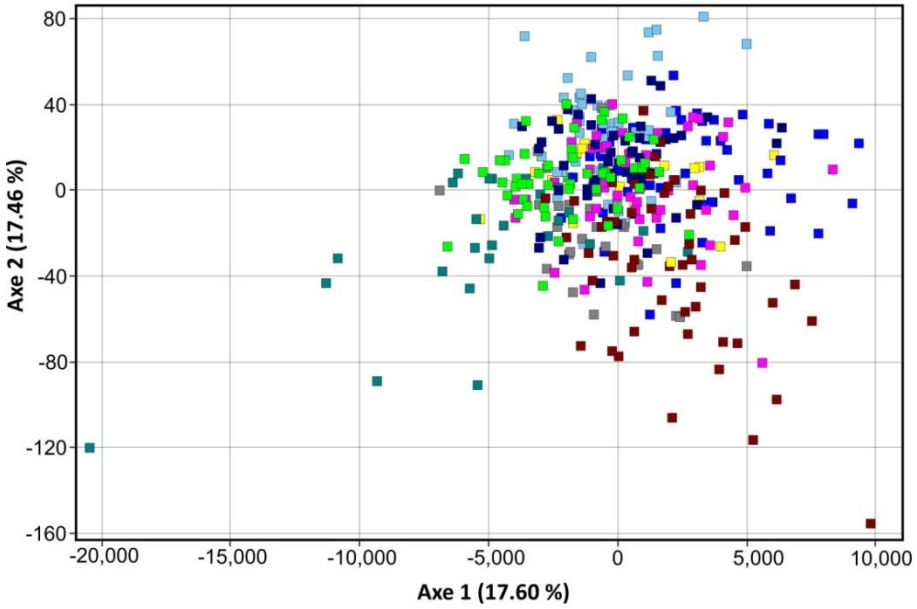


Figure 2.2.5: Plot for Axis 1 and 2 of the FCA.

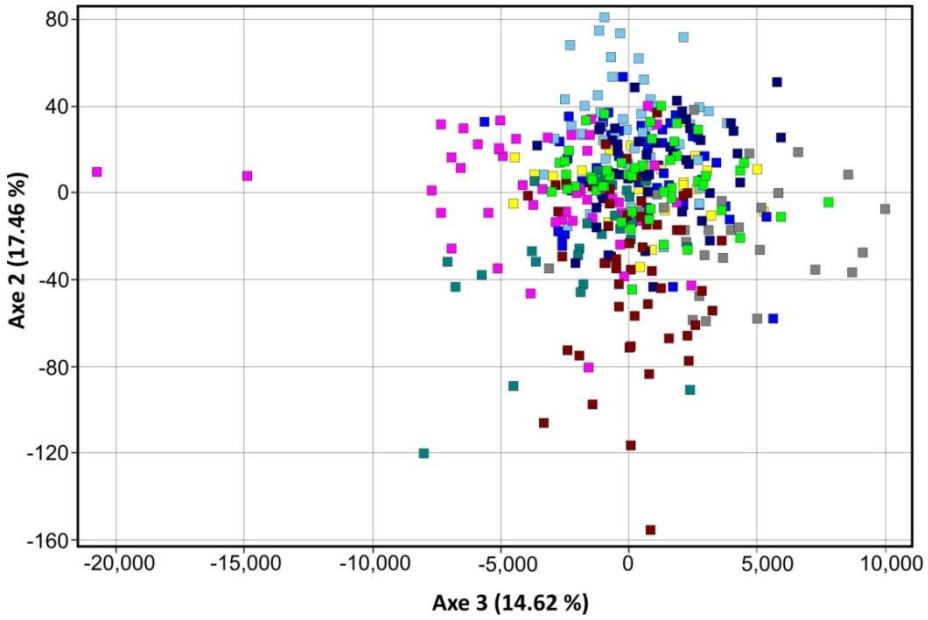


Figure 2.2.6: Plot for Axis 2 and 3 of the FCA.

The Hardy-Weinberg Exact test for heterozygosity deficiency and excess showed that all sample sites except North Taranaki Bight had a significant level of heterozygosity deficiency (Table 2.1.9). When the Pma4-32 locus was removed PFR broodstock, East Coast, and Marlborough Sounds were no longer significant. A locus-by-locus analysis showed that five of the eight loci had a significant level of heterozygosity deficiency. The loci with the highest numbers of alleles typically had the most significant levels of heterozygosity deficiency.

Table 2.1.9: Significance of Heterozygosity Deficiency by Sample Site (Table A) and Locus (Table B).

Table A

Sample Site	p-value (with Pma4-32)	p-value (without Pma4-32)
Doubtless Bay	0.0000	0.0015
East Coast	0.0000	0.1477
Foxton Beach	0.0000	0.0000
Hawkes Bay	0.0000	0.0139
Marlborough Sounds	0.0010	0.1788
North Farewell Spit	0.0000	0.0201
North Taranaki Bight	0.3264	0.7683
Tasman Bay	0.0003	0.0600

Table B

Locus	p-value
Pma1	0.5488
Pma2	0.0000
Pma3	0.0152
Pma4-32	0.0000
Pma5	0.9197
GA2A	0.0070
GT2	0.5050
GT4	0.0046

All the allele frequency distributions had an L-shape distribution (Appendix B). There was no evidence from this test that the sample sites have experienced a bottleneck in the last 24 generations that was strong enough to reduce the number of rare alleles (Luikart *et al.* 1998). The Hawkes Bay sample site appeared to have a reduced number of rare alleles, but not enough of a reduction to show a significant mode-shift.

All sample sites except Doubtless Bay, East Coast, and Hawkes Bay had an M-ratio value below the established ratio of 0.68 used to indicate a bottleneck (Garza and Williamson 2001) (Table 2.2.0). However, even the North Island sites with M-ratios just above 0.68 had less than 5% of the M_P_Val simulated replicates below the observed value. This suggests that these sample sites have experienced a bottleneck within the last 500 generations (Garza and Williamson 2001).

Table 2.2.0: M-ratio and Replicates Below Observed Value.

Sample Site	Value	S.D.	%
Doubtless Bay	0.6999	0.126	1.74%
East Coast	0.6836	0.188	1.06%
Foxton Beach	0.6606	0.141	0.84%
Hawkes Bay	0.7092	0.235	2.59%
Marlborough Sounds	0.5993	0.124	0.12%
North Farewell Spit	0.6592	0.174	0.51%
North Taranaki Bight	0.6525	0.193	0.35%
Tasman Bay	0.6319	0.130	0.22%

The results from BOTTLENECK 1.2.02 indicated that there was a significant heterozygosity deficiency in all sample sites when using the stepwise mutation model (SMM) and two phase mutation model (TPM) (Table 2.2.1). Removal of the Pma4-32 locus did not affect the number of sample sites with significant level of heterozygosity deficiency.

Table 2.2.1: Bottleneck Heterozygosity Deficiency p-value.

	With Pma4-32		Without Pma4-32	
Sample Site	SMM	TPM	SMM	TPM
Doubtless Bay	0.0039	0.0039	0.0039	0.0039
East Coast	0.0020	0.0020	0.0039	0.0039
Foxton Beach	0.0137	0.0371	0.0078	0.0117
Hawkes Bay	0.0137	0.0195	0.0117	0.0117
Marlborough Sounds	0.0039	0.0039	0.0039	0.0039
North Farewell Spit	0.0195	0.0273	0.0078	0.0273
North Taranaki Bight	0.0039	0.0039	0.0195	0.0273
Tasman Bay	0.0098	0.0273	0.0039	0.0039

SMM = Stepwise mutation model, TPM = Two-phase model

Table 2.2.2: Wrights Fixation Index (F_{ST}) Above the Diagonal with p-value Below the Diagonal.

	Doubtless Bay	East Coast	Foxton Beach	Hawkes Bay	Marlborough Sounds	North Farewell Spit	North Taranaki Bight	Tasman Bay
Doubtless Bay		0.0043	**0.0075***	0.0025	**0.0181**	**0.0118**	0.0025	0.0025
East Coast	0.0926		**0.0106**	0.0032	0.0083	0.0013	0.0004	0.0046
Foxton Beach	**0.0345**	**0.0230**		**0.0119**	**0.0149**	0.0056	**0.0061***	0.0042
Hawkes Bay	0.1463	0.1425	**0.0039**		**0.0128**	0.0036	0.0004	0.0017
Marlborough Sounds	**0.0005**	**0.0719***	**0.0083**	**0.0030**		0.0064	**0.0103**	0.0062
North Farewell Spit	**0.0007**	0.3828	0.1049	0.0922	0.0933		0.0006	**0.0477**
North Taranaki Bight	0.1465	0.4315	**0.0602***	0.3462	**0.0159**	0.3861		0.0015
Tasman Bay	0.1911	0.0944	0.1705	0.2494	0.0924	**0.0055**	0.2663	-

Underlined = Significant with and without Pma4-32, Asterix* = Significant only without Pma4-32

Table 2.2.3: Weir and Cockerham's Fixation Index (Theta) Above the Diagonal with Lower 95% CI Below the Diagonal.

	Doubtless Bay	East Coast	Foxton Beach	Hawkes Bay	Marlborough Sounds	North Farewell Spit	North Taranaki Bight	Tasman Bay
Doubtless Bay		**0.004**	0.007	0.003	**0.018****	**0.011**	0.003	0.003
East Coast	**0.002**		**0.010****	0.003	0.007	0.001	0.000	0.004
Foxton Beach	-0.002	**0.001****		**0.013**	**0.014**	0.005	0.007	0.003
Hawkes Bay	-0.001	-0.002	**0.006**		**0.013**	0.004	0.001	0.002
Marlborough Sounds	**0.006****	-0.002	**0.003**	**0.001**		0.005	0.011	0.005
North Farewell Spit	**0.005**	-0.004	-0.001	-0.001	-0.005		0.001	0.004
North Taranaki Bight	-0.001	-0.005	0.000	-0.002	0.000	-0.004		0.002
Tasman Bay	-0.004	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003	

Underlined = Significant with and without Pma4-32, Double = Significant only with Pma4-32**

2.4: Discussion

2.4.1: Demographic History

The demographic analysis and mtDNA haplotype trees indicated that the New Zealand P. auratus population has been stable for a long period of time, but may have experienced some recent fluctuations in population size. Two major haplogroups were identified, which were separated by a large number of mutational steps (steps=44) and present at all sample sites (Figure 2.1.6). The distinct haplogroups produced a mismatch distribution that had a predominantly bimodal distribution (Figure 2.1.3). Several possible scenarios which could have lead to the formation of these two haplogroups include, separation of the New Zealand population into two geographically isolated groups (e.g. East-West split), a recent migration event from outside of New Zealand (e.g. Australia), an unusual lineage sorting event that occurred by chance, or the New Zealand population could represent two geographically panmictic but reproductively distinct groups (e.g. two cryptic species).

During the Last Glacial Maximum (*ca*. 19-20k yr. BP) the sea level was about 120m lower than the present day, which would have exposed land around the coast and formed a land bridge connecting the North and South Island from the Tasman Mountains to Taranaki (Proctor & Carter 1989). There was no marine corridor between the west and east coast through the Cook Strait, which may have isolated marine populations on each side of New Zealand and led to genetically differentiated populations. If the long distance dispersal of *P. auratus* were mainly due to dispersal driven by water currents, as suggested by Bernal-Ramı´rez *et al.* (2003), then a land barrier across the Cook Strait would have caused a significant amount of disruption to gene flow because the southward moving water currents would have isolated populations on either side of the New Zealand coastline. In opposition to this idea, the distance among breeding groups at the time of last glacial maximum would most likely have been reduced because the southern populations would have been pushed further north to more favorable breeding temperatures (Cassie 2005). The submergence of the Cook Strait land-barrier at the end of the Last Glacial Maximum would have then allowed secondary contact between the two previously isolated populations result in intermixing across all sample sites. Whether connectivity between *P. auratus* populations is solely

the result of water currents is still somewhat uncertain, but would be a possible mechanism for driving the divergence of population on either side of New Zealand.

If the New Zealand population has experienced a bottleneck, which increased the loss of mtDNA lineages, it may have facilitated the development of the two haplogroups. The large number of mutational steps between the two haplogroups suggest they are relatively old, but the starburst-like patterns of haplotypes surrounding each of the major haplogroups suggests that many of the haplotypes have arisen only recently (Chen *et al.* 2004). This type of pattern could occur if a population remained stable for a long period of time allowing the development of distant haplotypes within a single population, followed by a bottleneck removing all but a few of these distant haplotypes. Recovery of the population would then result in the starburst formation around each of the distant haplotypes. This scenario is supported by the results of the Bayesian skyline plots which show that the New Zealand P. auratus population has been stable for a long period of time, with a recent reduction occurring between 3 to 30 thousand years ago, followed by a population expansion. The environmental upheavals caused by the Last Glacial Maximum 19 to 20 thousand years ago could have had a significant effect on coastal fish populations (Clark *et al.* 2009). During the Last Glacial Maximum *P. auratus* most likely had a significant reduction in suitable shoreline due to colder sea temperatures and minimum temperature breeding (Cassie 2005) requirement, which would have displaced southern populations and reduced the overall carry capacity of the New Zealand population. Together with separation around New Zealand this reduced populations size during the Last Glacial Maximum may have driven a rapid lineage sorting event to form the two haplogroups.

Another possibility is that the haplotype groups could represent an historic dispersal event from a genetically different population outside of New Zealand. The ancestral sequences of one haplogroup may have been present around the coast of New Zealand and another in populations, such as around the coast of Australia. Contact between the two ancestral populations might have occurred and the New Zealand population became a mixture of two distinct haplogroups. Comparison of haplotype samples from Australia could be used to confirm whether this scenario was or was not responsible for the presence of the two distant New Zealand haplogroups.

Comparison of the haplotypes from *P. auratus* and *P. major* indicated that the two New Zealand haplogroups were more distant from each other than a number of *P. auratus* haplotypes were from even *P. major* haplotypes. Consequently, the question arose as to whether the two haplotype groups may be reproductively distinct groups (e.g. two cryptic species) within what is considered a single species. To test this idea the microsatellite data was divided into two groups based on the two haplogroups and differentiation between the two groups determined using Theta. The results of this analysis between the two groups showed no evidence of genetic differentiation at nuclear DNA markers (Theta p-value = 0.526), which indicated that the two haplogroups did not represent two distinct reproductive groups.

Evidence for several possible bottlenecks was detected in the current study with both mtDNA and microsatellite DNA markers. The significant Tajima's D statistics suggested that there has been a recent genetic bottleneck in seven of the eight wild sample sites; however, significant Fu's Fs statistic provided evidence for a bottleneck only in the Tasman Bay. The differences between these two statistics may be caused by the presence of the two major haplogroups and the methods with which each estimate is calculated. The large number of mutational steps between the two haplogroups would have increased the number of polymorphic sites without equally increasing the mutations. Tajima's D statistic is calculated as the number of mutations relative to the number of polymorphic sites and consequently may have been strongly affected by the presence of the two distant haplogroups (Tajima 1989). In contrast, Fu's Fs statistic is calculated based solely on the number of mutations and therefore should not be affected by the presence of the two haplogroups (Fu 1997). The results of the Fu's Fs statistic test suggested that a recent bottleneck has occurred in the Tasman Bay, which was congruent with the finding of Hauser *et al.* (2002) using microsatellite DNA genotyping data from modern and historic samples. The fact that no significant Fu's Fs statistics were observed in any other sample sites, including other South Island sites, indicated that the proposed bottleneck may have been localized to a relatively small area within the Tasman Bay. In the microsatellite DNA data, the M-ratio test indicated that there has been a bottleneck at all sample sites. This bottleneck was unlike to have been a recent event because it was not detected using the Fu's Fs statistic. This putative bottleneck could have occurred anytime within the last 500 generations (Garza and

Williamson 2001). It is unlikely that this bottleneck coincides with the Last Glacial Maximum (ca. 19-20k yr. BP) as it would require a generation time of 40 years to fit within the 500 generations, which is longer generation time than expected for *P. auratus* (Clark *et al.* 2009; Paul and Tarring 1980).

In contrast to the bottlenecks discussed above, the significant amount of heterozygosity deficiency at mitochondrial loci suggests that New Zealand *P. auratus* has undergone recent population expansion following an old bottleneck event (Maruyama & Fuerst 1985; Cornuet and Luikart 1996). Heterozygosity deficiency is typical following a reduction in population size as mutation replenishes the number of alleles in a population and the distribution of allelic sizes returns to normal (Cornuet and Luikart 1996). This population expansion is most likely linked to the bottleneck indicated by the microsatellite M-ratio test, as significant M-ratio tests are the result of disruption of microsatellite allele distributions and significant heterozygosity deficiency can be caused as those allele distributions returns to normal.

2.4.2: Population Structure

The results of this study indicated that the New Zealand *P. auratus* population was a largely panmictic with was some low level differentiation between sites. This overall pattern of panmixia can be observed in the factorial correspondence analysis (FCA) plots (Figure 2.2.5 and Figure 2.2.6), which showed that most individuals form a single cluster with only a few individuals separating at the edges of the clusters.

No significant genetic differentiation was found between any sample sites using the mtDNA control region sequencing, which is consistent with the findings of previous studies of *P. auratus* mtDNA (Adcock *et al.* 2000; Bernal-Ramı´rez *et al.* 2003). This was most likely due to the relatively low statistical power of mitochondrial DNA markers. While there was no significant genetic differentiation between Tasman Bay and other South Island sample sites, a significant Fu's Fs statistic was found solely in the Tasman Bay. Significant Fu's Fs statistics can indicate that a population has been through a recent population bottleneck and would support the results of Hauser *et al.* (2002). If this were the case then the presence of the significant Fu's Fs statistic solely in the Tasman Bay would suggest that limited gene flow has occurred between the Tasman Bay and other South Island sample sites within the last 60 years since the putative bottleneck was suggested to have occurred (Hauser et al. 2002).

This is particularly surprising in the case of the North Farewell Spit and Tasman Bay samples as the sample sites are located approximately 50 km apart with no obvious barriers to fish movement between them.

The North Farewell Spit (NFS) and Marlborough Sounds (MS) were significantly differentiated from Doubtless Bay with both F_{ST} and Theta after sequential Bonferroni correction (NFS = 0.011 and MS = 0.018). Apart from the significant differentiation observed between East Coast and Doubtless Bay this was the only other significant differentiation in the data set not involving at least one small sample size. Lacson and Morizot (1991) presented results suggesting that measures of population differentiation can be affected by demographic changes such as bottlenecks. The presence of the significant Fu's Fs statistic, the putative bottleneck proposed by Hauser et al. (2002), and the differences in results between the current study and the results of Bernal-Ramı´rez *et al.* (2003) indicate that genetic differentiation measurements in the Tasman Bay may not be temporally stable and could be being influenced by its recent demographic history.

The break observed by Bernal-Ramı´rez *et al.* (2003) between the East Coast and Hawkes Bay was not present in this data set. Instead, low but significant differentiation was observed between the East Coast and Doubtless Bay (Theta = 0.004). Several North Island samples including the East Coast were collected from within a set geographical range, but the exact sample sites were not known. If the East Coast sample was from the southern side of the genetic break proposed by Bernal-Ramı´rez *et al.* (2003) then this would explain how the East Coast has changed from being differentiated from southern sample sites to being differentiated from northern sample sites.

While the methodology used in the calculation of Theta attempts to correct for sample size, there were some significant Theta values in the present data set between sites with small sample sizes, which were inconsistent with the results of previous studies and the results for sample sites with larger sample sizes (Weir & Cockerham 1984). For example, Hawkes Bay (sample size = 49) was significantly differentiated from Foxton Beach (sample size = 27, Theta = 0.13), but not from North Taranaki Bight (sample size =49, Theta = 0.001). However, Foxton Beach and North Taranaki Bight should be comprised of individuals from an ecologically panmictic population (Morrison 2008). These three sample sites are also located

along the coast, with Foxton Beach located between Hawkes bay and North Taranaki Bight. Individuals from North Taranaki bight would have to travel through Foxton Beach to mix with Hawkes Bay, however the significant differentiation between Foxton Beach and Hawkes Bay indicate that this is not happening. This suggests that the smaller sample sizes in the current data set may be susceptible to indicating falsely significant differentiation.

After sequential Bonferroni correction and removing sample sites with small sample sizes (26 and 27 individuals) there was no significant differentiation between South Island and lower North Island sample sites with either microsatellite DNA F_{ST} or Theta. This result was inconsistent with the findings of Bernal-Ramı´rez *et al.* (2003) who reported a significant level of differentiation between Hawkes Bay and Tasman Bay and West Coast of the North Island and Tasman Bay. The difference between these two studies could be explained by variation in the number of samples and microsatellite loci used and possible effects of using Tasman Bay as a representative South Island population. In the current study sample sites with smaller sample sizes appear to be susceptible to falsely significant levels of genetic differentiation (e.g. Foxton Beach, sample size = 27). However, these smaller sample sizes may still have had higher statistical power than the Tasman Bay sample size of 43 in Bernal-Ramı´rez *et al.* (2003) due to an increased number of loci and total alleles (loci = 8 versus 6, total alleles = 78 versus 56) (Selkoe and Toonen 2006). In addition, if the allele frequencies in the Tasman Bay were affected by a recent bottleneck, as proposed by Hauser *et al.* (2002), then the significant amount of genotype differentiation could have been an artifact of strong genetic drift.

In conclusion the genetic differentiation results in this study indicate that the New Zealand *P. auratus* population is largely panmictic with a small amount of genetic differentiation between the most distant sample sites. This study also suggests caution should be taken with further studies using small sample sizes (< 50 individuals) and low numbers of microsatellite markers (e.g. < 10). While limited genetic differentiation was observed with most estimates of genetic differentiation, some data, such as the presence of Fu's Fs statistic solely in the Tasman Bay, suggest that in some situations movement and gene flow among *P. auratus* may still take decades to travel across relatively small distances (60 kms).

Chapter 3: Constructing a Broodstock That Best Represents the Genetic Resources of the Wild Population

3.1: Introduction

The first step in the developing of new aquaculture species is establishing a broodstock of individuals sourced from the wild population. The broodstock should contain a sufficient amount of genetic variation because those first individuals will provide the genetic potential that underpins a selective breeding program and it will limit the problems that can result from inbreeding (Allendorf & Phelpsa 1980; Charlesworth & Willis 2009). A population genetic study is needed to determine the levels of genetic variation contained in the wild source populations. This study of wild populations can then be used to guide the collection of individuals from each genetically distinct population and the number of individuals that are needed to sufficiently represent the wild genetic variation. For example, an allele that is found at a frequency of 0.05 in a genetically panmictic population will require a larger sample of individuals for it to be included in the broodstock compared to an allele that is at a frequency of 0.25; or if the source population is found to be genetically subdivided then individuals may need to be collected from each area of the population in order to better represent the wild genetic variation.

Inbreeding and loss of traits over successive generations due to strong genetic drift in the relatively small breeding groups of a typical aquaculture population needs to be considered when establishing the broodstock (Franklin 1980; Jamieson & Allendorf 2012). Franklin (1980) used models to show that a minimum population size of 50 breeding individuals was needed to mitigate the risk of inbreeding depression in a small population. The development of methods to reduce inbreeding, such as minimal kinship crossbreeding and the slightly pronounced inbreeding rate used by Franklin (1980) may allow the target of 50 individuals to be slightly lower; however this minimum population size should be accurate for most cultured populations (Jamieson & Allendorf 2012; Doyle *et al.* 2001; Ortega-Villaizan *et al.* 2011).

In some aquaculture programs it may be possible to regularly supplement the broodstock with individuals from a wild source population. However, this could have a negative effect

on a selective breeding program because carefully selected fish lines may experience a reduction in productivity as wild genetic variation is added to the gene pool. Integration of individuals from the wild into selected lines may have the positive effect of reducing inbreeding depression, but it could also have the negative effect of causing outbreeding depression (Pekkala *et al.* 2012). Starting a selective breeding program with the high levels of genetic variation that will be required at latter stages of the program and maintaining as much of that variation as possible throughout the program should prevent inbreeding depression and the need to supplement the selected line with wild genetic variation.

It is difficult to determine the level of genetic variation required for a selective breeding program because gene variants cannot be easily matched to specific phenotypic traits. In general, the best approach when establishing a broodstock is to maximize the level of genetic variation from the wild source populations. The sampling plan needs to consider how much variation is in the wild populations and how that variation is structured. Simply obtaining an estimate of the level of genetic variation may be of limited value when determining whether broodstock sufficiently represent wild genetic variation because it does not take into account how the broodstock variation is represented in relation to its geographic structuring. In some cases, the broodstock may appear to have similar levels of genetic variation to the wild source population, but the allele frequencies and genotypes might be significantly differentiated (e.g. Ma *et al.* 2011). Previous studies typically compared cultured and wild populations using both estimates of genetic variation and differentiation, which should help take into account the geographic structuring of the genetic variation (Ha *et al.* 2009, Song *et al.* 2011, Ma *et al.* 2011). In most of these studies the presence or lack of significant genetic differentiation between a cultured and wild population is used to imply that the broodstock lack or have sufficient genetic variation, respectively. However, some care needs to be taken when interpreting the results of these estimates as the presence or lack of significance can be affected in part by the power of the marker being used.

Many studies that have compared the genetic variation in cultured and wild populations have used either mitochondrial DNA (mtDNA) markers or microsatellite DNA markers (Ha *et al.* 2009; Song *et al.* 2011; Lundrigan *et al.* 2005; Sekino *et al.* 2002). Microsatellite DNA marker sets only sample small amounts of the genome (e.g. 10-20 loci) and therefore only

provide a very general overview about the levels of genetic variation and differentiation within and among populations. Mitochondrial DNA markers can be even less representative of the levels of genetic variation because mtDNA is an extra-nuclear genome and does not undergo recombination (Avise 1994). However, the low cost of these markers does make them a cost effective option for use in projects such as the current study.

The aim of the research reported in this chapter was to investigate how well the Plant and Food Research (PFR) *Pagrus auratus* broodstock (henceforth referred to as "PFR broodstock" and BS) represent the genetic variation found at the wild sample sites. The mtDNA and microsatellite DNA data from the PFR broodstock has been compared to the wild population data reported in chapter 2 using the same estimates of genetic variation and differentiation as several previous studies (Ha *et al.* 2009; Song *et al.* 2011; Lundrigan *et al.* 2005; Sekino *et al.* 2002). A python scripted simulation was written to investigate the proportion of wild alleles at each frequency that were represented within the PFR broodstock. This method was developed based on an idea proposed by Tave (1999), in which it was suggested that maintaining alleles that were above a frequency of 0.05 in cultured populations should prevent the loss of traits that were under positive selection in the cultured environment. The script was also used to look at the possible effects of increasing the number of the PFR broodstock with individuals from wild sample sites. Simulated broodstock groups were also constructed by randomly combining individuals from the wild sample sites into a new population without replacement. This process of simulating broodstock groups could be useful at the start of broodstock collection to determine the likely genetic variation that would be collected with different sampling plans.

3.2: Materials and Methods

3.2.1: DNA extraction and sequencing, and microsatellite DNA genotyping

The mitochondrial DNA (mtDNA) control-region and microsatellite primers used in Chapter 2 of this study were also used for this Chapter. The forward and reverse primers for the mtDNA control-region were tRNA-Pro (H-tRNA-Pro 5'-CACCATTGGCTCCCAAAGC-3') and Central Conserved Domain (CCD) (L-CR-CCD 5'-GGCACTGTGAGATGTCAACTG-3'), respectively. The microsatellite primers are shown below in Table 3.1.0.

Table 3.1.0: Microsatellite Primers (M13 primer 5'-TGTAAAACGACGGCCAGT-3').

Locus	Forward 5'-3'	Reverse 5'-3'
Pma1	M13 + CATGCCAGTATTCCATGTGC	AGGACAAATTCCCAAGGTCATCC
Pma2	M13 + TGCCAAGGAGAGGTGAGGAG	TATTCTTGCAAGGGTCCAAACG
Pma3	M13 + CTAAACAAATAATTCTGGTTTAGC	CATGACAAACTGAGGTGAATGC
Pma4-32	M13 + CCTGCCACCTACTGTTTCCT	CGGTGATTACAGTCGGGTTT
Pma5	M13 + CCAATCGGATTGAGTATCTGTGG	GAGAGGTTCTCCGTCACTGTCC
GA2A	M13 + ACGGACAGAGAGGGAGTGG	CATCATCATCAGTCAGAGCTG
GT2	M13 + TAGGGGTCTTGCTGCCTGCTC	ACAGATACAGACTGACAGACG
GT4	M13 + CTTTACACTGTTGAAGCTTTGG	CGCACAGATAAACAAATCAAT

Samples from 26 *P. auratus* broodstock were collected by Plant and Food Research (Nelson). Fin clips were removed from each fish and stored in 85% ethanol at 4^{o}C. Total genomic DNA was isolated using a standard phenol-chloroform protocol (Sambrook *et al.* 1989). The small sample of tissue was individually placed in 400 µL of extraction buffer (10 mM Tris pH 8.0, 50 mM NaCl, 10 mM EDTA, and 0.2% SDS) with 0.5 µg/µL proteinase-K and incubated overnight at 50°C. Following the tissue digestion, DNA was extracted with phenol, followed by chloroform/isoamyl alcohol, and precipitated with 1 mL of ethanol at -20°C for 1 hour. The DNA was pelleted, washed with 70% ethanol, dried and then re-suspended in a Tris-EDTA buffer. The DNA was stored at 4^{o}C. The purified DNA was quantified using a Nanodrop spectrophotometer and the concentration was adjusted to 200 ng/µL using distilled water.

A 641 bp portion of the control region was amplified using the Polymerase Chain Reaction (PCR) and the primers H-tRNA-Pro and L-CR-CCD. All PCRs consisted of 10.4 mM Tris-HCl pH 8.8, 52 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.4 µM of each primer, 0.6 µg/µL Bovine Serum Albumin (BSA), 0.5 units of Taq DNA Polymerase (Fisher Scientific), and 200 ng of template DNA. The conditions for thermal cycling were: 33 cycles of 94^{o}C for 30 seconds,

58°C for 30 seconds, and 74°C for 60 seconds, followed by an extension step of 74°C for 10 minutes. The resultant PCR products were purified with ExoSAP-IT (Amersham Parmacia Biotech) and their DNA sequence was determined using the reverse primer and an ABI 3730 Genetic Analyzer (Macrogen, Korea).

The microsatellite loci were amplified by PCR using the primers in Table 1.1 and the amplified fragments genotyped at Massey Genome Services on an ABI3730 Genetic Analyzer. FAM, VIC, or PET fluorescent labels were incorporated into the PCR products in the amplification step using the M13 labeling method (Schuelke 2000). The PCR amplification was carried out on individual microsatellite loci which were then poolplexed for genotyping; with either two or three loci per genotyping well (Table 3.1.1).

Each 10µL PCR contained 10mM Tris-HCl, 50mM KCl, 0.8mM dNTP, 0.1µM forward primer, 0.4µM reverse primer, 0.1µM M13 primer, 4.0µg BSA, 1 unit Fisher *taq* polymerase, and the template DNA and $MgCl_2$ shown in Table 2.3. The conditions for thermal cycling were: 33 cycles of 94°C for 20 seconds, annealing temperature (Table 3.1.1) for 30 seconds, and 74°C for 60 seconds, followed by 8 cycles of 94°C for 20 seconds, 53°C for 30 seconds, and 74°C for 60 seconds, followed by a final extension step of 74°C for 10 minutes.

Table 3.1.1: Microsatellite PCR Concentrations and Temperatures.

Poolplex	Locus	T-DNA (ng)	$MgCl_2$ (mM)	T-Ann (°C)
1	Pma1	100	2.0	60
1	GA2A	150	2.0	60
1	GT2	150	1.5	57
2	Pma2	150	1.5	64
2	Pma4-32	150	2.0	58
2	GT4	150	1.5	59
3	Pma3	150	1.5	57
3	Pma5	150	1.5	59

T-DNA = Template DNA, T-Ann = Annealing Temperature

3.2.2: Data Analysis

DNA sequences were aligned in MEGA 5.0 using a ClustalW alignment with default settings (Tamura *et al.* 2011). Variable sites were checked for base calling errors and the final alignment exported in fasta format to be used in further analysis. The haplotype (*h*) and nucleotide (π) diversity, and average number of pairwise differences (*k*) were estimated

using DNASP 5.0 (Rozas *et al.* 2003). Mitochondrial haplotype richness for 24 individuals and average microsatellite allele richness for 26 individuals was calculated in HP-RARE and R (Kalinowski 2005; Rioux Paquette 2011). Rarefaction plots for haplotype and average allele richness were calculated in HP-RARE and R. Fu's Fs statistics was calculated using ARLEQUIN 3.5 with 1000 permutations to determine the statistical significance. The number of alleles (A), observed heterozygosity (H_O), and expected heterozygosity (H_E) for all sampled sites were determined using ARLEQUIN 3.5. ARLEQUIN 3.5 was used to calculate the fixation index (Φ_{ST}) for pairwise comparisons between sample sites and 1000 permutations of the data set were used to determine statistical significance (Excoffier *et al.* 2005). Wright's fixation index (F_{ST}) for pairwise comparison between all sites was estimated using ARLEQUIN 3.5. Significance at the 5%-level was determined using 20,000 permutations (Wright 1951). Weir and Cockerham's fixation index (Theta) was measured for all pairwise comparisons between sites in FSTAT 2.9.3.2 (Goudet 2005; Weir & Cockerham 1984).

3.2.3: Population Genetic Simulations: Python Script Design

A script was written using Python 2.6, which determined the percentage of alleles at each frequency in the wild population that were not represented in the broodstock (representation method). The script was also able to test the effect that adding new individuals to the broodstock would have on the results and also construct simulated broodstock groups to test the effects of different sampling plans on the results. The script supported the use of one broodstock group and up to ten wild populations. If multiple permutations or populations were used then the script calculated the mean and standard deviation of all simulations or comparisons. The Python script ran through eight steps, which are described below.

1. **Import populations**

The populations were set out in a Python file (.py) with a nested list for each population. When the script was run it automatically imported the populations from the data file (Data.py).

Example: the input setup for two populations with two loci and two individuals:

Broodstock = [[[146, 146], [151, 153]],

[[145, 147], [151, 153]]

Wild Pop = [[[147, 146], [149, 151]],

[[145, 149], [153, 155]]

2. ***(Optional) Increasing and constructing broodstock populations***

This step was only used if individuals were being added to the broodstock or the script was being used to simulate broodstock groups. In this step individuals were randomly transferred with replacement into the broodstock from any of the wild populations. Any combination and number of wild populations could be used in this step to source broodstock.

3. ***Alleles at each locus were combined into lists by locus and population.***

Example: from step 1:

Broodstock = [[146,146,145,147], [151,151,153,153]]

Wild Total = [[147,146,145,149], [149, 151,153,155]]

4. ***Comparison of the Broodstock and Wild Total list was used to construct a list that contains alleles found only in the wild population.***

Example: from step 3:

Wild Only = [[149], [149,155]]

5. ***The allele frequency in Wild Total of alleles found in Wild Only and Wild Total was calculated and stored in a Python dictionary.***

Example: locus 1 from step 3 and 4:

Wild Total locus 1 = {‘147’: 0.25, ‘146’: 0.25, ‘145’:0.25, ‘149’:025}

Wild Only locus 1 = {‘149’: 0.25}

6. ***<u>The number of alleles at each frequency was then calculated for Wild Only and Wild Total</u>***

Example: locus 1 from step 5:

Wild Total = 4 alleles at 0.25 Wild Only = 1 allele at 0.25

7. ***<u>The percentage of alleles not found in the Broodstock at each frequency was calculated using the number of alleles in the Wild Only and Wild Total</u>***

Example: locus 1 from step 6:

Wild Total = 4 alleles at 0.25 Wild Only = 1 allele at 0.25 Not Represented = (4/6)*100

Output: 25% of alleles at 0.25 at locus 1 in the wild population were not found in broodstock

8. ***<u>Multiple loci, permutations, and population comparisons.</u>***

The results for multiple loci were calculated simultaneously using further nested lists. Multiple permutations of the simulation were run by cycling through steps 2 to 7; but were only needed if individuals were being added to the broodstock in step 2. This script also allowed the broodstock to be compared with between one and ten wild populations from the input file. Mean and standard deviation were calculated if multiple permutations were used or if the broodstock were compared to more than one population. Finally, the results were written to an Excel sheet; the output format was three columns; 1 = Percent allele frequency in wild population, 2 = Average alleles not represented at each allele frequency, 3 = standard deviation for multiple permutations or population comparisons. If only a single permutation was used and the broodstock were only compared to one population then only column 1 and 2 was written to the results sheet.

3.2.4: Current Broodstock Allele and Haplotype Representation

The proportion of haplotypes at each frequency were not represented in the Broodstock was calculated for each population individually using the output from GENALEX 6.41 (Peakall & Smouse, 2006, 2012). The Python script was used to determine the percentage of alleles

at each frequency that were not represented in the PFR broodstock for each wild population individually and for the average of all wild populations. The allele representation plot for all wild populations was constructed with two standard deviations.

3.2.5: Increasing the Current Broodstock

The effect of increasing the broodstock population with individuals sourced from the Tasman Bay was tested by comparison with the Tasman Bay and all sample sites. Increasing with increments of five individuals was used for both plots up to a total of 50 new broodstock. The effect of increasing the broodstock with individuals from all wild sample sites was tested by increasing the broodstock with an equal number of individuals from each wild population. Increments of eight individuals (one per wild population) were used to increase the number of broodstock up to a total of 48. 1000 permutations were used when determining the effect of increasing the broodstock.

3.2.6: Selecting New Broodstock Groups

The Python script was used to simulate the collection of broodstock groups from the Tasman Bay in increments of five individuals up to a total of 50, with 1000 permutations at each sample size. Two plots were constructed to compare the current Broodstock and simulated broodstock groups from the Tasman Bay against the Tasman Bay and North Farewell Spit. The same number of individuals as the current Broodstock and 1000 permutations was used in the simulated broodstock groups. Two plots were constructed to determine the representation of alleles from all sample sites in simulated broodstock groups collected from the Tasman Bay and all sample sites; 1000 permutations was used at each sample size and an even number of individuals from each sample site was used when constructing the broodstock from multiple sample sites.

3.3: Results

3.3.1: Comparison of PFR broodstock and Wild Sample Sites using Standard Measures of Genetic Variation and Differentiation

The comparison of PFR broodstock genetic data with wild sample sites from Chapter Two indicated slightly different results with mtDNA and microsatellite DNA markers. The broodstock had lower mtDNA haplotype diversity, nucleotide diversity, and average number of pairwise nucleotide differences than the wild sample sites (PFR broodstock = 0.736, 0.0240, and 15.30, respectively; ranges for wild sample sites = 0.815-0.910, 0.0297-0.0334, and 18.96-21.28, respectively). However, at microsatellite markers the PFR broodstock had similar estimates of variation to those observed in the wild populations. The expected and observed heterozygosity were 0.699 and 0.654, respectively, for the PFR broodstock and ranged from 0.708-0.748 and 0.650-0.743, respectively, for the wild sample sites. The allele richness adjusted for sample size was 8.63 in the PFR broodstock and ranged from 7.80-9.61 for the wild sample sites. The PFR broodstock also had a significant positive Fu's Fs statistic which is indicative of a reduction in population size (Fu's Fs = 6.091)

Table 3.1.2 shows the haplotype and allelic richness for the PFR broodstock and all the wild sample sites sampled in Chapter two of this study. The haplotype and allele richness was based on resampling of 24 and 26 individuals from each of the sample sites, respectively. The results show that the PFR broodstock had the lowest haplotype richness, but were similar to wild sample sites with allele richness.

Table 3.1.2: Haplotype and Allele Richness for 24 and 26 Individuals, Respectively.

	BS	DB	EC	FB	HB	MS	NFS	NTB	TB
Haplotype Richness (24)	9.00	12.46	12.37	11.78	13.81	13.56	12.33	11.60	10.26
Allele Richness (26)	8.63	9.05	7.80	9.61	9.15	9.38	8.28	9.02	8.94

There was no significant differentiation between the PFR broodstock and wild sample sites with Φ_{ST}, F_{ST}, or Theta after sequential Bonferroni correction (Table 3.1.3). The only significant differentiation was between the PFR broodstock and the East Coast site, before sequential Bonferroni correction.

Table 3.1.3: P-value for Φ_{ST} and F_{ST} and Lower 95% Confidence Interval for Theta Between PFR broodstock and Wild Sample Sites.

	Broodstock		
Wild Sample Site	**Φ_{ST}**	**F_{ST}**	**Theta**
Doubtless Bay	0.157	0.326	-0.005
East Coast	**0.028**	0.550	-0.006
Foxton Beach	0.598	0.228	-0.003
Hawkes Bay	0.115	0.507	-0.006
Marlborough Sounds	0.116	0.291	-0.005
North Farewell Spit	0.078	0.724	-0.006
North Taranaki Bight	0.133	0.769	-0.007
Tasman Bay	0.289	0.808	-0.006

3.3.2: Broodstock Representation of mtDNA Haplotypes and Microsatellite DNA Alleles

The results of the representation method for comparing broodstock and wild population genetic variation and differentiation were similar with both mtDNA markers and microsatellite DNA markers, although the microsatellite DNA markers were much more precise. The representation method results indicate that over all sample sites the PFR broodstock represent most haplotypes and alleles above a frequency of 0.7 (Figure 3.1.0 and Figure 3.1.1).

The mtDNA markers had significantly lower power for measuring the representation of medium frequency variation due to the low number of none-rare haplotypes. While all wild haplotypes above a frequency of 0.07 were represented in the PFR broodstock, most wild sample sites had low numbers of haplotypes above a frequency of 0.04, which meant it was not possible to determine the exact representation of medium frequency variation in the PFR broodstock. For example, haplotypes from Foxton Beach were only found at three frequencies (frequencies = 0.41, 0.19, and 0.04). The large gap between 0.19 and 0.04 prevented exact measurement of the level of the wild genetic variation between these frequencies that was represent in the PFR broodstock.

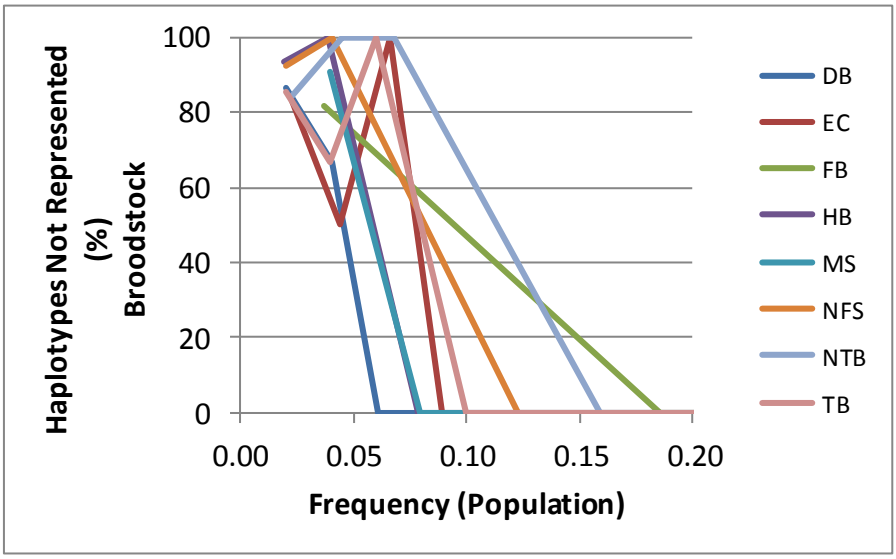


Figure 3.1.0: Haplotypes at Each Frequency in Each Sample Sites Not Represented (%) in the PFR broodstock.

The PFR broodstock included all alleles that were over a frequency of 0.12 in all wild sample sites, 0.08 in the Tasman Bay, and 0.05 in the North Farewell Spit. The microsatellites had better coverage of allele frequencies than the mtDNA markers with one or more alleles at each frequency up to 0.15 in most sample sites and good coverage above this in some sample sites.

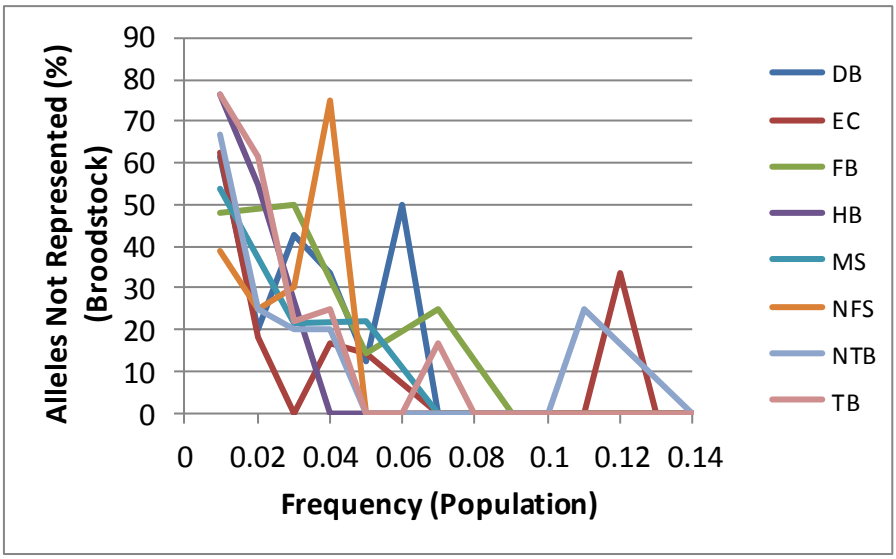


Figure 3.1.1: Alleles at Each Frequency in Each Sample Site Not Represented (%) in the PFR broodstock.

Figure 3.1.2 shows the average for all plots shown in Figure 3.1.1 with two standard deviations. The average percentage of alleles not represented in the PFR broodstock was under 10% for all alleles at a frequency of 0.07 or higher in the wild sample sites. All alleles at a frequency of 0.12 or higher in the wild sample sites were represented in the PFR broodstock.

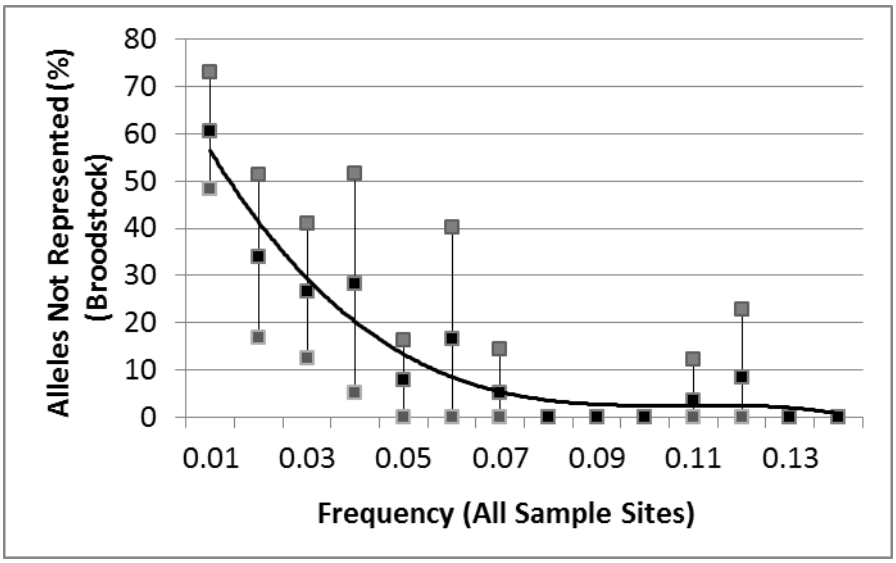


Figure 3.1.2: Alleles at Each Frequency from All Sample Sites Not Represented (%) in the PFR broodstock, With Mean, Two Standard Deviations, and Polynomial Fit (4).

3.3.3: Increasing the Number of PFR broodstock Individuals

Single Sample Site

Figure 3.1.3 and 3.1.4 show the percentage of alleles from the Tasman Bay and wild population, respectively, that were not represented in the PFR broodstock, with simulated addition of new individuals to the group from the Tasman Bay. All alleles with a frequency of 0.05 or more were represented by increasing the PFR broodstock with 20 individuals from the Tasman Bay (Figure 3.1.3). Increasing the PFR broodstock from a single sample site did not improve the representation of alleles that were not present at that site. Consequently, increasing the PFR broodstock with even 50 individuals from the Tasman Bay did not allow the representation of all alleles above the frequency of 0.05.

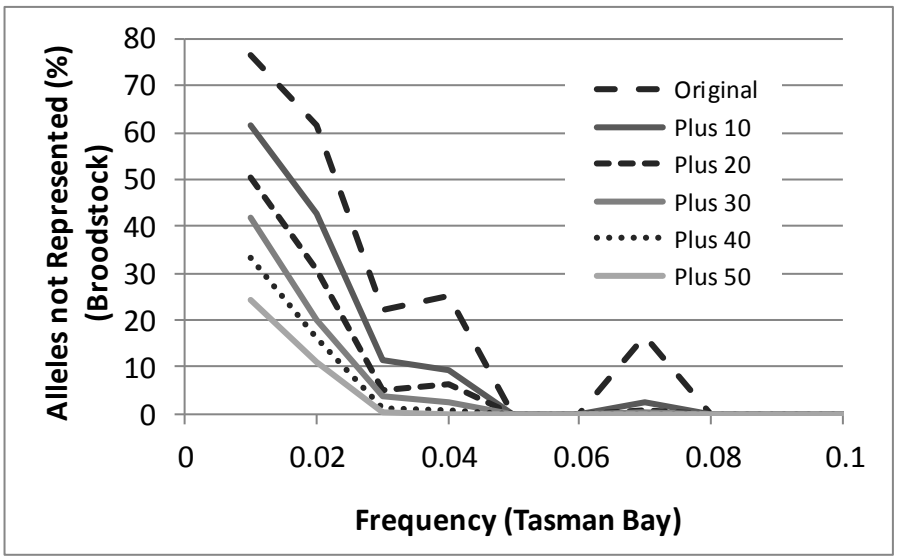


Figure 3.1.3: Alleles at Each Frequency from Tasman Bay Not Represented (%) in the PFR broodstock When Increased with Individuals from the Tasman Bay.

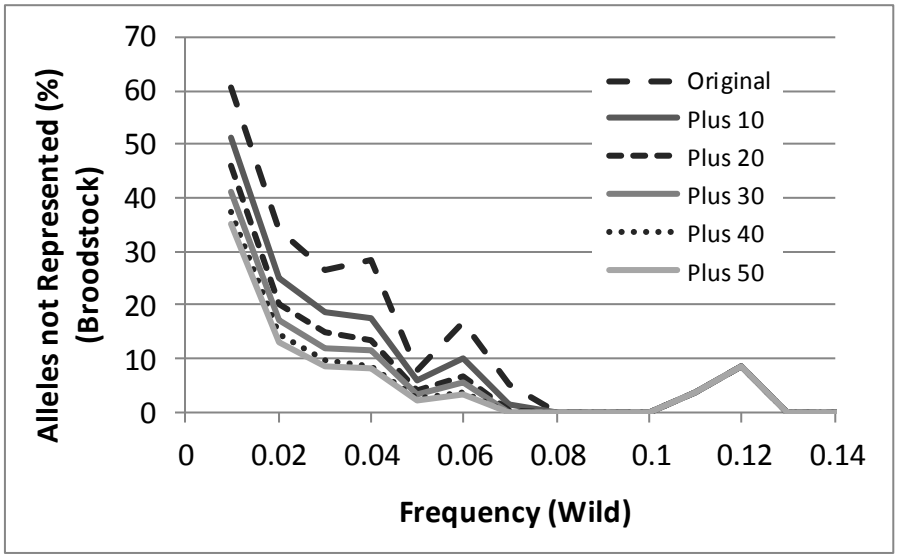


Figure 3.1.4: Alleles at Each Frequency from All Sample Sites Not Represented (%) in the PFR broodstock When Increased with Individuals from the Tasman Bay.

Multiple Sample Sites

Figure 3.1.5 shows the percentage of alleles from the wild population that were not represented in the PFR broodstock, with simulated addition of new individuals to the group from multiple sample sites. Around 48 new individuals were needed to reach the target representation frequency of 0.05.

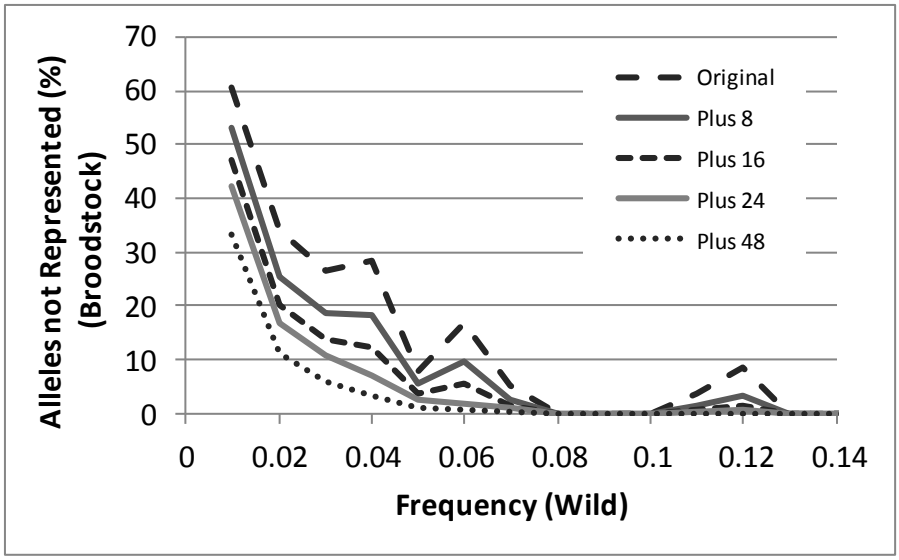


Figure 3.1.5: Alleles at Each Frequency from All Sample Sites Not Represented (%) in the Current PFR Broodstock When Increased with Individuals from All Sample Sites.

3.3.4: Selecting New Broodstock Groups

Single Sample Site

Figure 3.1.6 shows the percentage of alleles found in the Tasman Bay that were not represented in simulated broodstock groups collected from the Tasman Bay. Representation of all alleles over 0.05 in the Tasman Bay required a simulated broodstock sample size of approximately 50 individuals.

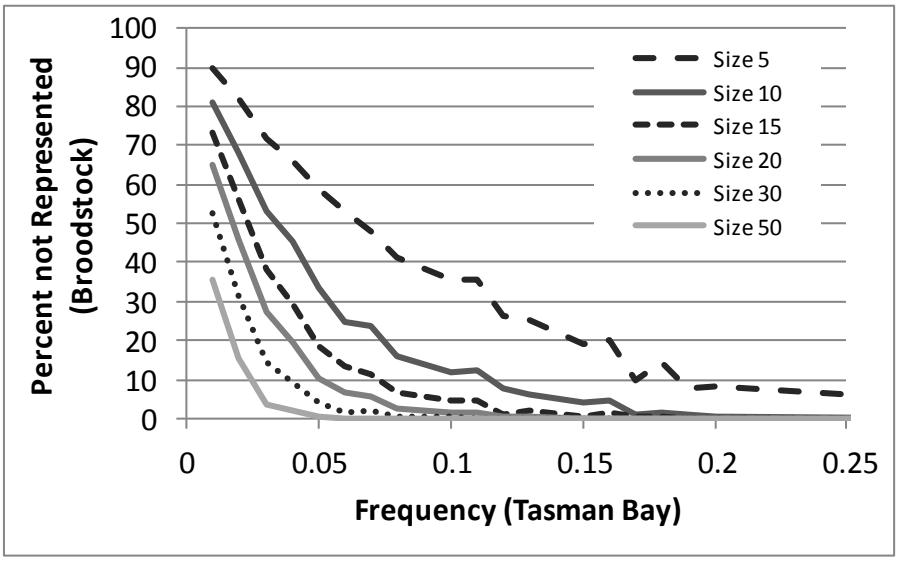


Figure 3.1.6: Alleles at Each Frequency from Tasman Bay Not Represented (%) in the Simulated Tasman Bay Broodstock.

Figure 3.1.7 and Figure 3.1.8 shows the percentage of alleles from the Tasman Bay that were not represented in simulated broodstock groups from the Tasman Bay and North Farewell Spit. This was compared with the results for the PFR broodstock. The PFR broodstock had a similar pattern of allele representation to the simulated broodstock groups. The best results were observed when comparing broodstock groups with their source population as opposed to sample sites other than their source population.

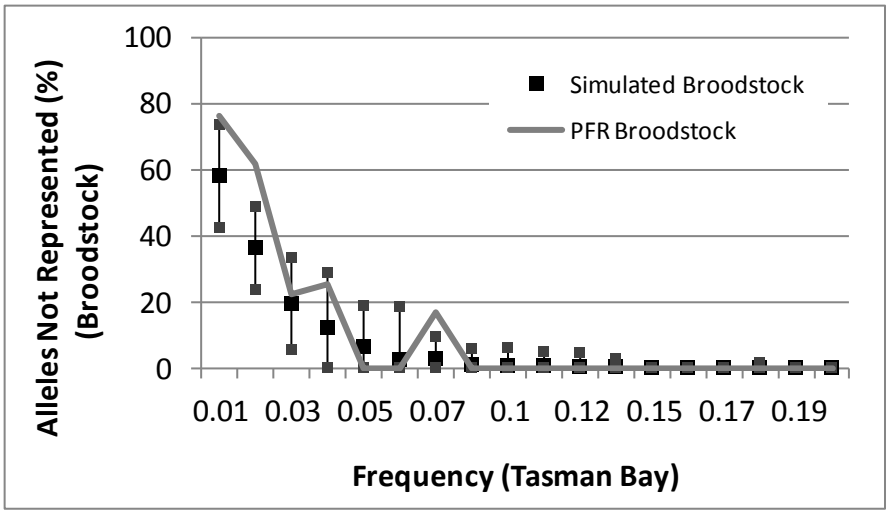


Figure 3.1.7: Alleles at Each Frequency from Tasman Bay Not Represented (%) in the ■ Simulated Tasman Bay and ▬ PFR Broodstock.

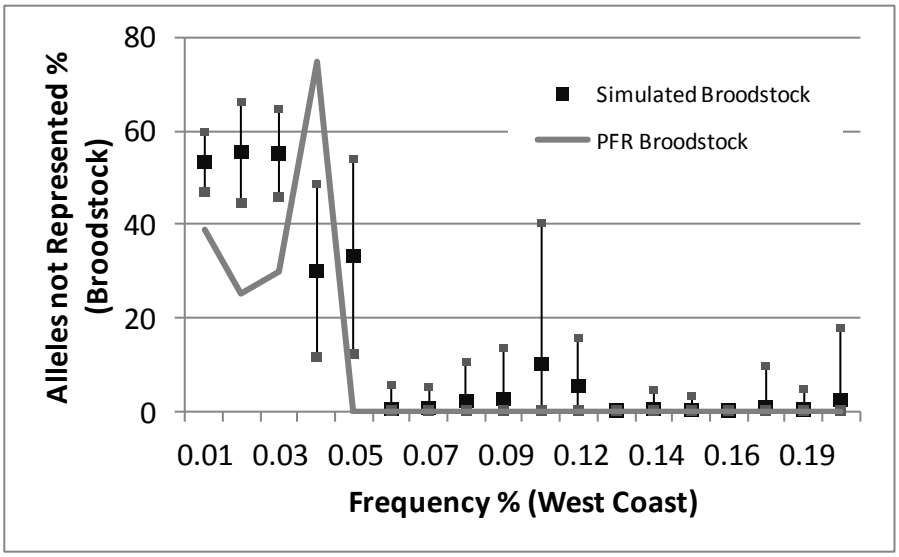


Figure 3.1.8: Alleles at Each Frequency from North Farewell Spit not Represented (%) in the ■ Simulated Tasman Bay and ▬ PFR Broodstock.

Figure 3.1.9 shows the percentage of alleles found in the wild population that were not represented in simulated broodstock groups of different samples sizes. The simulated broodstock groups were constructed using individuals solely from the Tasman Bay sample. Most of the major improvements to allele representation are made by a sample size of 25 individuals. Further increases in the number of individuals past 25 resulted in relatively low improvements to the representation. Even with a total of 50 individuals collected from the Tasman Bay some alleles at a frequency of 0.12 were still not represented in the simulated broodstock groups.

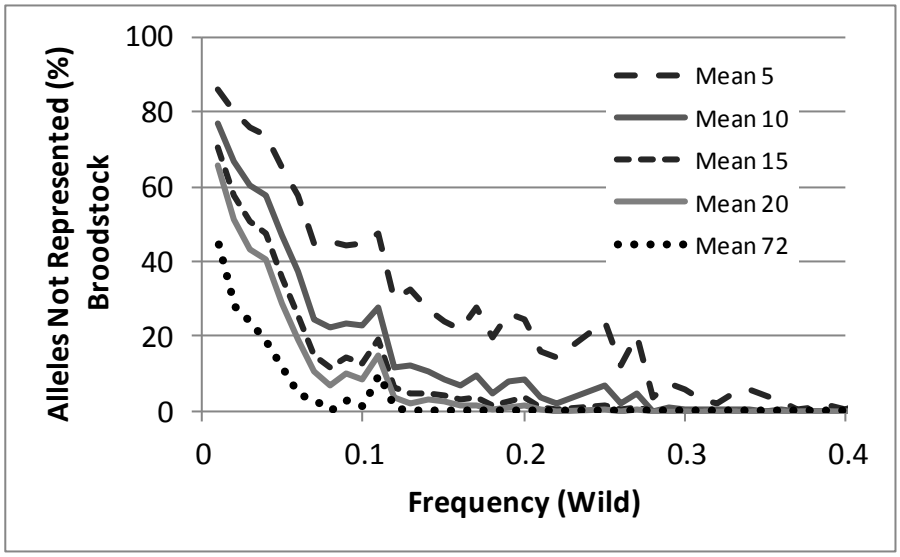


Figure 3.1.9: Alleles at Each Frequency from All Sample Sites not Represented (%) in the Simulated Tasman Bay Broodstock.

Multiple Sample Sites

Figure 3.2.0 shows the percentage of alleles found in the wild population that are not represented in simulated broodstock groups of different samples sizes. An even number of individuals were taken from each of the wild sample sites to construct the simulated broodstock groups. Most of the major improvements to representation of alleles above a frequency of 0.05 occurred by a sample size of 40. Increasing the sample size from 40 to 72 slightly improved the representation of alleles that were below and slightly above 0.05. These simulations indicate that more than 72 individuals collected from multiple sample sites would be needed to represent all wild alleles above a frequency of 0.05.

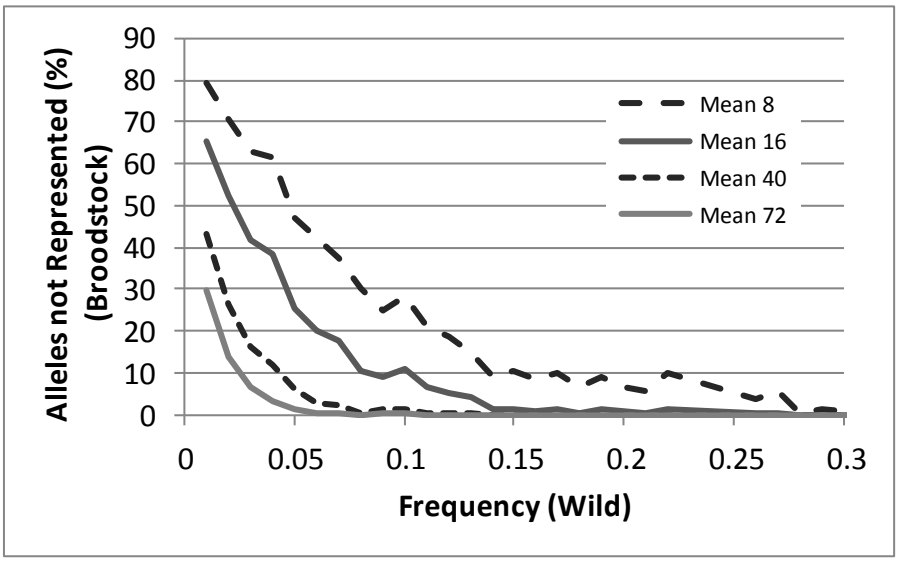


Figure 3.2.0: Alleles at Each Frequency from All Sample Sites not Represented (%) in the Simulated Broodstock Collected from All Sample sites.

3.4: Discussion

3.4.1: Broodstock Genetic Diversity

The haplotype diversity, nucleotide diversity, and average number of pairwise nucleotide differences were lower in the PFR broodstock than in any of the wild sample sites, including those with small sample sizes (e.g. 26 and 27). The level of haplotype diversity, nucleotide diversity, and average number of pairwise differences was 9.7, 19.2, and 19.3 percent lower, respectively, in the PFR broodstock than the lowest wild sample site. By comparison, the same measurements in a study of cultured Japanese flounder, which were suggested to need genetic supplementation, were found to be 14.5, 6.2, and 4.9 percent lower in the highest cultured population than the lowest wild population (Song *et al.* 2011). Haplotype richness, which adjusts for differences in sample size, was lower in the broodstock than at wild sample sites (9.0 versus 10.3). A genetic bottleneck was also detected in the PFR broodstock with Fu's *Fs* statistic (Fu's *Fs* = 6.091). The lack of significance for a bottleneck test in a broodstock population was used by Ha *et al.* (2009) to support the suggestion that the broodstock had a sufficient level of wild genetic variation. Based on that criterion, the bottleneck that the present study detected in the PFR broodstock would suggest that they were a poor representation of the genetic variation present in the wild population.

Overall, the mtDNA results suggested that the PFR broodstock had a reduced level of genetic variation compared to all wild sample sites, including the source population in the Tasman Bay. The PFR broodstock are wild-caught individuals, which indicates that the difference between PFR broodstock and wild sample sites was most likely due to random sampling variance when they were collected or an artifact of reduced variation in the source population at the time of collection. The presence of the Tasman Bay bottleneck proposed by Hauser *et al.* (2002) could support the suggestion that some of the difference between broodstock and wild sample sites was due to reduced variation in the source population at the time of collection.

The same level of reduced genetic variation in the PFR broodstock compared to the wild sample sites was not observed with the microsatellite DNA markers. Expected heterozygosity in the PFR broodstock was slightly lower than at the lowest wild sample site (0.690 versus 0.708) and observed heterozygosity was slightly higher than the lowest wild

sample site (0.654 versus 0.650). The average number of alleles per locus was slightly lower in the PFR broodstock than at the lowest wild sample site (8.63 versus 9.00), but allelic richness, which adjusts for differences in sample size, was not (8.63 versus 7.80), with lower allelic richness observed at the East Coast and North Farewell Spit sites (Leberg 2002). Expected heterozygosity is influenced by the number of alleles and consequently adjustments for sample size, such as used when calculating allelic richness, would most likely result in broodstock expected heterozygosity being similar to the wild sample sites. A lack of difference in allelic richness and heterozygosity, similar to the current study, was the criterion that Ha *et al.* (2009) used to suggest that cultured populations were founded with a sufficient number of broodstock.

As discussed in the introduction of this chapter it is important to use estimates of genetic differentiation to determine how the variation in cultured populations is represented in relation to its geographic structuring. The PFR broodstock were not significantly differentiated at mtDNA and microsatellite DNA loci from any wild sample sites after sequential Bonferroni correction, which in previous studies was the criterion used to support the suggestion that the broodstock sufficiently represented the wild genetic variation (Ha *et al.* 2009) or in cases where there was significant differentiation to support that there was insufficient wild genetic variation (Song *et al.* 2011).

In the current study the mtDNA results suggested that the PFR broodstock did not sufficiently represent the wild genetic variation but the microsatellite DNA markers suggested that they did. There are several attributes of these markers that could be responsible for these differences, including differences in effective population size and proportion of rare and private alleles. Mitochondrial DNA has a quarter of the genetically effective population size of nuclear loci, which means it will be more sensitive than microsatellite DNA to the strong level of genetic drift that can be experienced during a founding event (Carvalho & Hauser 1994). However, the wild-caught PFR broodstock still contain the genetic variation that will be lost in this founding event as a result of the smaller mtDNA effective population size and consequently this will only affect the offspring of the PFR broodstock and subsequent generations. The differences between the two marker types may also be influenced by differences in the distribution of private, rare, and total haplotypes and alleles; in the current study 77% of haplotypes were private while only 25%

of alleles were private and 79% of haplotypes were below a frequency of 0.05 while only 50% of alleles were below a frequency of 0.05. Rare haplotypes and alleles are more likely to be lost in a bottleneck or founding event compared to the higher frequency variants. Consequently, the mtDNA marker should be more sensitive to the loss of variation than the microsatellite DNA markers, which was the most likely reason for differences between the results of the two marker types observed in this chapter.

3.4.2: Development of the Representation Method

There were several issues with the approaches used in the previous studies, which lead to the development of the representation method. First, the combined use of multiple estimates was needed to determine the level of wild genetic variation represented in the broodstock and how this variation is structured in the wild (Ha *et al.* 2009; Song *et al.* 2011). These multiple estimates can difficult to apply back to some questions that might be asked in a breeding program (e.g. what is the likelihood that traits of X frequency will be present in the broodstock?). The applicability of these estimates to answering such questions was also brought into question in the current study by the fact that wild sample sites that were not genetically differentiated from the PFR broodstock contained an allele at a frequency of 0.11 that was not present in the PFR broodstock. A single allele at this frequency did not have a significant effect on the genetic differentiation estimates, but may be important when determining suitability of broodstock for detecting low frequency genetic variation. Finally, in the current study using the approach from previous studies (e.g. Ha *et al.* 2009; Song *et al.* 2011) resulted in different conclusions for the two marker types. This was most likely due to the differences between the allele and haplotype frequency distributions (discussed above).

The representation method developed in this Chapter attempted to mitigate the issues described in the preceding paragraph. First, it was able to apply directly to the question presented in the previous paragraph, which would likely be of interest in a breeding program. It also calculated and presented the representation of each wild allele and haplotype in the broodstock individually rather than a single estimation across all alleles and haplotypes. This allowed the observation that an allele at a frequency of 0.11 in a wild sample site was not represented in the PFR broodstock, even though the estimates of

differentiation did not detected a significant difference between the two sample sites. By separating the allele and haplotype representation into different frequencies it should also remove the influence of different allele and haplotype frequency distributions on the results.

3.4.3: Haplotype and Allele Representation in the Broodstock

The results of the representation method with the mtDNA markers indicated that the PFR broodstock included all wild haplotypes that were present at a frequency of 0.07 or more in the wild sample sites. The large number of frequencies with no haplotypes in each sample site means that the frequency at which wild haplotypes stopped being represented was not able to be determined precisely. For example haplotypes in Foxton Beach were only present at three frequencies (frequencies = 0.41, 0.19, and 0.04). The large gap between 0.19 and 0.04 meant that it could be said that all Foxton Beach haplotypes above 0.19 were represented in the broodstock, when actually it may have been that all Foxton Beach haplotypes above 0.08 were represented.

The representation method was more precise when used with microsatellite DNA markers because most populations had one or more alleles at each frequency up to 0.15 and relatively good coverage of frequencies above this level. The microsatellite DNA markers would also have had less sampling error than mtDNA because they sampled multiple independent points on the genome (Harrison 1989; Selkoe & Toonen 2006). The PFR broodstock contained all alleles that were at a frequency of 0.08 or higher in the Tasman Bay for the eight microsatellite loci. This level of allele representation was similar when compared to all other populations including those in the North Island. Two exceptions were the East Coast and North Taranaki Bight; in which the PFR broodstock had representatives of all alleles above a frequency of 0.11 and 0.12, respectively. These results were similar to the result of the mtDNA marker, which showed that all wild haplotypes above a frequency of 0.07 were represented in the PFR broodstock. The results of both mtDNA and microsatellite DNA markers suggest that the PFR broodstock are close to representing the target frequency of 0.05, but will require an increased number of individuals to do so.

3.4.4: Increasing PFR broodstock with Wild Sourced Individuals and Simulating New Broodstock Groups

Simulations were used to predict the increased representation of wild alleles in the PFR broodstock, by collecting new individuals from one or multiple populations and adding them to the broodstock. All Tasman Bay alleles above the target frequency of 0.05 were represented by adding 20 new individuals from the Tasman Bay to the PFR broodstock. However, even if all alleles above the target frequency of 0.05 in the Tasman Bay were represented, there were still a large number of alleles above the target frequency in other sample sites that were not. Increasing the PFR broodstock with approximately 48 new individuals from multiple populations was suitable to represent all allele above the target frequency of 0.05.

Simulating broodstock groups may provide a useful tool when constructing an initial broodstock from wild populations. In the current study, the PFR broodstock had similar levels of representation to simulated broodstock groups from the Tasman Bay. This provides support for the accuracy of the Python script for predicting the likely representation of alleles in broodstock groups based on the allele distributions in the wild sample sites and using specific sampling plans.

3.4.5: Further Improvements to Methods

Significant differences between cultured and wild populations was used as the criteria in several previous studies to determine if cultured populations sufficiently represented wild genetic variation (Ha *et al.* 2009; Song *et al.* 2011). However, different criteria were needed for the representation method. Tave (1999) suggested that alleles above a target frequency of 0.05 should be maintained within cultured populations, because alleles under selection would be above this frequency. This target frequency has been used in the current study. However, the adaption of most domesticated fish to captive conditions indicates some traits under selection in a captive populations may not be under selection in wild populations and consequently may be at a lower frequency than 0.05 (Gjedrem 2005; Christie *et al.* 2012). It is not entirely clear how differential selection influences variation in cultured and wild populations, but the target frequency suggested by Tave (1999) should be applicable to most gene variants under positive selection in at least one wild sample site.

The second improvement that could be made to the method used in this chapter would be to increase the number of loci used. The neutral genetic markers used in this study sample relatively few loci on the genome and consequently only provide a rough proxy to guide the construction of a broodstock so that they have the maximum level of genetic variation. Using a larger number of loci would increase the statistical power and accuracy when designing a sampling plan (Selkoe & Toonen 2006). Sampling a large number of points on the genome could also allow adaptive variation and quantitative trait loci (QTL) to be detected (e.g. Nielsen *et al.* 2009; Zhang *et al.* 2012). Adaptive variation and QTL could be used for other work within aquaculture programs, such as enhancement of selective breeding programs.

Chapter 4: Implications of the Current Study for Fisheries Management and Aquaculture

4.1: Fishery Management

4.1.1: Fisheries Overview

Pagrus auratus is an important inshore recreational and commercial fishery around the coast of New Zealand. In 2009 the fishery made up 7% of the total value of New Zealand's commercial fish resource (Statistics NZ 2012). The current annual total allowable commercial catch (TACC) for New Zealand *P. auratus* is set at 6,357 tonnes and the fishery is divided into six management areas (Figure 4.1.0) (MFish 2012). The management area with the highest total allowable commercial catch (TACC) is SNA1 (4,500 tonnes), followed by SNA8 with (1,300 tonnes). The remaining four management areas have TACC ranging from 10 to 315 tonnes (MFish 2012).

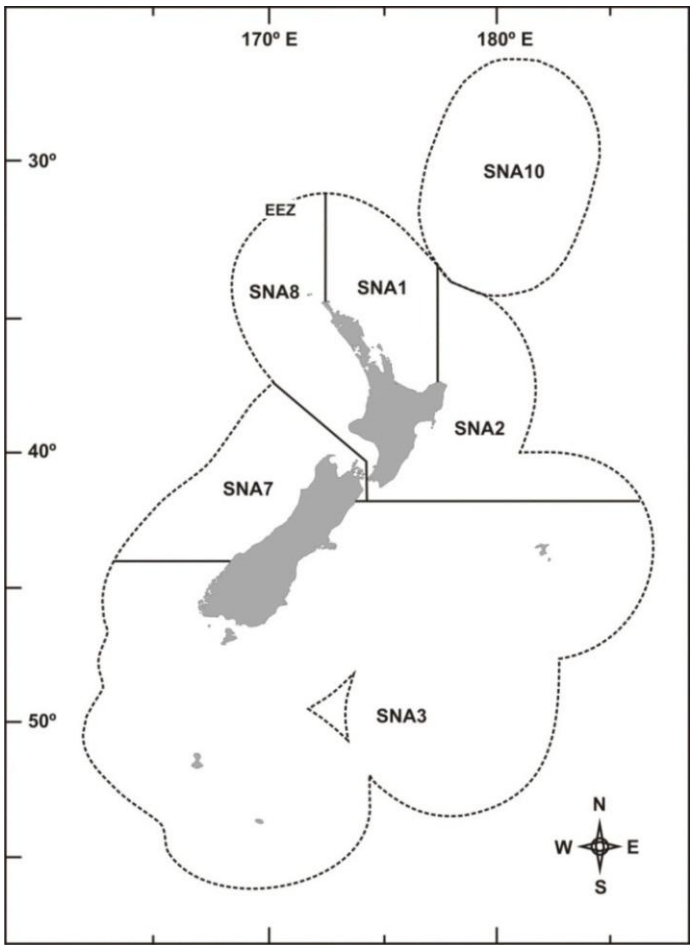


Figure 4.1.0: New Zealand *P. auratus* Management Areas.

4.1.2: Population Differentiation and Management Areas

Management of more than one distinct stock within a single management area can result in one stock being depleted due to uneven fishing pressure (Carvalho & Hauser 1994). To minimise the risk of this, management areas should be aligned with the biological population structure of fisheries stocks and the total allowable catch (TAC) set to a sustainable limit for each area. A number of studies using direct and indirect methods have been used to look at biological stock structuring in New Zealand *P. auratus* (Bernal-Ramı´rez 2003; Willis *et al.* 2001; Morrison 2008). Each of these studies observed different aspects of the biological structure, but all were useful for better understanding the biological structure of the *P. auratus* population.

The results of the current study suggested that the New Zealand *P. auratus* population is predominantly panmictic with weak differentiation observed between the North Farewell Spit and Doubtless Bay. Bernal-Ramı´rez *et al.* (2003) proposed that barriers to gene flow around the east coast of the North Island and across the Cook Strait led to population differentiation in New Zealand *P. auratus*. The current study provided further support for the significant differentiation observed between the most northern and southern populations by Bernal-Ramı´rez *et al.* (2003), but did not find the same level of significant differentiation across the Cook Strait (except with the smaller samples sizes).

While managing multiple stocks within a single management area can have the negative effect of depleting one or more stocks due to overfishing, there should be no such issues when managing a single panmictic population within multiple management areas (Carvalho & Hauser 1994). Based on the panmixia observed in the current study, the management areas for New Zealand *P. auratus* (Figure 4.1.0) fall under this second category, with multiple management areas for a single panmictic population. However, the small amount of genetic differentiation observed in this study (between the North Taranaki Bight and Doubtless Bay) together with the differentiation observed in previous studies indicate that it may be prudent to maintain the current management areas until more powerful genetic markers can be used to further examine the gene-flow within the New Zealand population (Bernal-Ramı´rez *et al.* 2003).

4.1.3: Historical Demographics

The results of the Bayesian skyline plots in the current study indicated that the New Zealand *P. auratus* population has been stable for a long period of time, but has undergone several recent demographic fluctuations during or recently after the Last Glacial Maximum. These changes have most likely occurred before the introduction of commercial fishing pressure and consequently may have limited use for informing current fisheries management. Determining the effect of climate change on fisheries is one area of concern where long term demographics may be applied to fisheries management (Brander 2009). New Zealand *P. auratus* are at their southern geographical limit due to a minimum breeding temperature (Cassie 2005). Predicted water temperature increases due to climate change would most likely be advantageous to this species in New Zealand and increase the capacity of the fisheries due to increased coastline with suitable breeding temperatures and decreased seasonal fluctuations in temperature related larval survival (Wratt & Mullan 2012; Cassie 2005). The long term historical demographic stability of this species indicates that even relatively drastic climate change events have and may continue to have limited effect on their persistence.

4.1.4: Recovery of the Tasman Bay

A recent bottleneck (< 60 years ago) was proposed to have occurred in the Tasman Bay due to the effects of over-exploitation and was also detected in the current study (Hauser *et al.* 2002). Bernal-Ramı´rez *et al.* (2003) suggested that the Tasman Bay represented an isolated population, which meant that the only source of migration to aid its recovery from the proposed bottleneck would be across the Cook Strait. The genetic differentiation between the Tasman Bay and lower North Island in prior studies indicated that less than one individual per generation was traversing across the Cook Strait and breeding with individuals on the other side (Bernal-Ramı´rez *et al.* 2003; Cowen *et al.* 2000). Together these results implied that the recovery of the Tasman Bay was uncertain and would most likely take a long period of time.

The current study identified two South Island sample sites that had not been through the recent bottleneck detected in the Tasman Bay, including the North Farewell Spit and Marlborough Sounds sample sites. The close geographical distance and lack of significant

genetic differentiation between these two South Island sites and the Tasman Bay indicates that they should provide a source of migration much closer than across the Cook Strait. In addition, the current study suggests that the Cook Strait may be a weaker barrier to gene flow than observed by Bernal-Ramı´rez *et al.* (2003), which should increase the chances that gene flow from across the Cook Strait will play a greater role in the recovery of the Tasman Bay population.

While the identification of two South Island sites that have not been through bottlenecks and a higher rate of gene flow across the Cook Strait in this study suggests that the recovery rate of the Tasman Bay may be higher than implied from the results of previous studies, it is still unsure how long this recovery is likely to take. The large scale and long distance dispersal of *P. auratus* in some areas of New Zealand and lack of significant differentiation between South Island populations suggests that this recovery should occur over a relatively short period of time (Morrison 2008; chapter 2). However, the fact that the genetic signal from the proposed bottleneck has not dissipated throughout the South Island populations after more than 40 years suggests that the recovery time may be much longer (Hauser *et al.* 2002). Accurate projection for the recovery of the Tasman Bay population would require further analysis of the localized population structuring within and between the South Island sites and could be aided by further temporal sampling.

4.1.5: Individual Mobility and Marine Protected Areas

Marine Protected Areas (MPAs) are important for conservation and fishery management as they may provide a local population that can act as a source for the surrounding unprotected areas and a buffer to over-fishing problems (Allison *et al.* 1998). Increased species mobility is suggested to reduce the benefits of MPAs, while improving the surrounding fisheries with migration out of the MPAs (Grüss *et al.* 2011). Because of this the effective use of MPAs for fishery management requires matching the size of the MPAs with the mobility of any target species. The location of MPAs may also be important as differences in site to site species mobility could alter their effectiveness at different locations.

Morrison (2008) used otolith microchemistry to show that most individuals on the west coast of the North Island originated from the Kaipara Harbour and some had travelled

distances of up to 700 km. However, tagging studies in the Hauraki Gulf showed high levels of site fidelity, with some individuals remaining within a 500 meter range over a 3 year period (Willis *et al.* 2001). Bernal-Ramı´rez *et al.* (2003) suggested that New Zealand *P. auratus* subdivision may be due to water currents. If true, the differences between otolith microchemistry and tagging studies may be due to differences in water currents at each site and the effect of water current on species mobility at different life stages. Indirect measurement of mobility by genetic differentiation appear to support long distance dispersal, possible due to differential water currents, but is limited by the fact that only one migrant per generation is needed to genetically homogenise populations (Wright 1931; Wang 2002; Bernal-Ramı´rez *et al.* 2003). On the other hand, the identification of a possible bottleneck from more than 40 years ago in Tasman Bay but not at the North Farewell Spit site (50 km apart) indicates that there may have been limited mixing of these populations during the last 40 years (Hauser *et al.* 2002). These conflicting results may provide further support for site specific differences in mobility of New Zealand *P. auratus*.

A study of several MPAs around the Hauraki Gulf indicated that *P. auratus* within them were at higher density and of a larger size than those outside the MPAs (Willis *et al.* 2003). However, if similar MPAs to these were constructed on the west coast of the North Island they may be less effective if *P. auratus* mobility is greater at this location. If the differences in mobility are due to life stage specific movement, such as passive larval dispersal by water currents, then it may not completely reduce the effectiveness of the MPAs as adults and breeding individuals should still be protected. Further studies would be needed to corroborate the proposed difference in mobility and the influence this may have on the formation of MPAs.

4.2: Aquaculture

4.2.1: Current Broodstock and Suggested Changes

Ensuring that broodstock populations have sufficient wild source genetic variation before undertaking a selective breeding program is important as those first individuals will contain the genetic potential that underpins a selective breeding program and limits the problems that can result from inbreeding (Allendorf & Phelpsa 1980; Charlesworth & Willis 2009). Variation of neutral genetic markers can be used as a proxy to guide construction of broodstock so that they best represent the wild genetic variation. A genetic population study, such as the one carried out in Chapter 2, is needed so that broodstock can be compared with wild populations and the number of individuals that are needed from each source population to represent variation from all populations can be determined.

In the current study, several measures of genetic variation and differentiation, which have been used in previous studies to compare cultured and wild populations, were used to determine if the Broodstock sufficiently represented the wild genetic variation, including haplotype diversity, nucleotide diversity, average number of pairwise nucleotide differences, haplotype richness, allele richness, F_{ST}, and Φ_{ST} (Ha *et al.* 2009; Song *et al.* 2011). In these studies differences between the broodstock and wild populations were typically used to imply that broodstock did not sufficiently represent wild genetic variation and lack of difference was typically used to support the suggestion that broodstock did. The results of the current study suggested that the Broodstock did not sufficiently represent wild genetic variation based on the mtDNA results and using the criteria from previous studies. Using the same criteria the microsatellite DNA results suggested that the Broodstock did represent sufficient wild genetic variation. The difference between the two markers was most likely due to the fact that mtDNA marker had a higher proportion of low frequency variants than microsatellite DNA markers, and was consequently more sensitive to the loss of low frequency variation.

The representation method developed in this study determined how many alleles or haplotypes at each frequency in the wild populations were represented in the broodstock population and whether the broodstock represent all alleles above the target frequency of 0.05 suggested by Tave (1999). The results of this study indicated that while the broodstock

represented most medium to high frequency haplotypes and alleles above a frequency of 0.08-0.12 in the wild populations a large number of low frequency variants below this were not represented in the broodstock. According to the target frequency of 0.05 the Broodstock need further individuals to reach this target frequency.

Simulations using the representation method indicated that increasing the Broodstock with approximately 15 and >48 individuals from multiple populations was needed to reach a target frequency of 0.05 in the Tasman Bay and all sample sites, respectively. Because some alleles were only present at certain sample sites it was necessary to collect individuals from multiple populations in order to detect them. In addition to the suggestions for further broodstock collection based on the representation method a minimum increase in the number of Broodstock, by 24 individuals, is needed to reach the target size of 50, which is necessary to prevent excessive inbreeding (Franklin 1980).

Individuals that are larger due to increased age require more resources to maintain than smaller individuals but only contribute the same amount of genetic variation to the offspring. Some studies have shown a correlation between increased size of older individuals and increased egg size, yolk sac size, growth rates and larval survival (Kennedy *et al.* 2007; Springate and Bromage 1985; George 1994; Sehgal and Toor 1991; Monteleone and Houde 1990). However, other studies suggest that the effects of these factors on the overall performance of the offspring appears to be limited to the first few weeks after hatching and may only provide an advantage during harsh conditions, which may not be encountered in an established culture population (Kennedy *et al.* 2007; Louzao *et al.* 2008). The Broodstock are comprised of individuals that that are relatively large and most likely >25 years old (Paul and Tarring 1980). If the Broodstock were increased with further wild-caught individuals the broodstock maintenance cost could be reduced, while still maintaining the quality of the progeny, by collecting 4-6 year old individuals, which is the age of broodstock used in a *P. major* selective breeding program (Murata *et al.* 1996).

4.3: Future Work

Using a larger number of genetic markers would improve the accuracy when determining whether broodstock sufficiently represent wild genetic variation by increasing the number of points on the genome that are sampled and for the representation method by increasing the coverage of different allele frequencies (Selkoe & Toonen 2006). Large numbers of markers, which construct fine scale genetic marker maps, are able to detect selection acting on different parts of the genome and how that selection is distributed in the wild population (Nielsen *et al.* 2009; Zhang *et al.* 2012). Simulations have also indicated that typically used neutral loci may have a different distribution to quantitative trait loci (QTL) under positive selection within wild populations (Corre and Kremer 2003). Consequently, using only neutral genetic markers may result in inaccurate estimates about the number of individuals that are needed to sufficiently represent wild genetic variation. Utilizing markers that detect both neutral and selective variation would ensure that the effects of natural selection within the wild population are also considered when selecting individuals for aquaculture broodstocks.

Fine scale genetic markers, such as single nucleotide polymorphisms (SNPS), can be used to detect QTL, which may be linked to traits of interest for selective breeding programs (e.g. Gutierrez *et al.* 2012). Quantitative trait loci are powerful tools for use in marker assisted selection (MAS), which can be used to enhance the gains from selective breeding (Hayes *et al.* 2007). One particularly useful feature of MAS is that it allows the measurement of traits in breeding candidates that would otherwise require destructive sampling. Meuwissen & Goddard (1996) suggested that increases in genetic gain from MAS will be greatest for these types of traits that cannot typically be measured on breeding candidates.

Identification of quantitative trait loci linked to traits of interest using fine scale genetic markers is an important step towards identifying the genes and genetic components underlying those traits (Gutierrez *et al.* 2012). Comparison of gene maps formed by fine scale genetic markers against fully sequenced species can then be used to increase the efficiency with which candidate genes are identified (Sarropoulou and Fernandes 2011). In some cases where genes or genetic variation is conserved between species it may also be possible to transfer knowledge about genetics underlying particular traits directly between species (e.g. Edmunds *et al.* 2009).

In conclusion while the mtDNA and microsatellite DNA markers were suitable for the scope of the current study, future work would be most effective if carried out with fine scale genetic maps of the *P. auratus* genome and potentially even large scale sequencing. A fine scale genetic map could be used when selecting broodstock and for enhancing selective breeding programs. Quantitative trait loci could then potentially be compared to genome sequencing (preferably *P. auratus*) to help determine the genetic and biological processes underlying specific traits and how they might best be utilized for meeting the needs of the fisheries and aquaculture industry.

References:

Allendorf, F.W., England, P.R., Luikart, G., Ritchie, P.A. & Ryman, N. (2008) Genetic effects of harvest on wild animal populations. Trends in Ecology and Evolution, 23, 327-337.

Allendorf, F. W. & Hard, J. J. (2009) Human-induced evolution caused by unnatural selection through harvest of wild animals. Proceedings of the National Academy of Sciences, 106 (suppl. 1), 9987-9994.

Adcock, G. J., Bernal-Ramı´rez, J. H., Hauser, L., Smith, P. & Carvalho, G. R. (2000) Screening of DNA polymorphisms in samples of archived scales from New Zealand snapper. Journal of Fish Biology, 56, 1283-1287.

Allison, G.W., Lubchenco, J. & Carr, M.H. (1998) Marine reserves are necessary but not sufficient for marine conservation. Ecological Applications, 8(1), 79-92.

Alves, M. J., Coelho, H., Collares-Pereira, M. J. & Coelho, M. M. (2001) Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanica*: importance for conservation. Heredity, 87, 463-473.

Annala, J. H., Sullivan, K. J., O'Brien, C. J., & Iball, S. D. (2000) Report from the Fishery Assessment Plenary. Stock Assessments and Yield Estimates. NIWA Research, Wellington, New Zealand.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. (2008) LOSITAN: A workbench to detect molecular adaptation based on an Fst-outlier method. BMC Bioinformatics, 9, 323.

Ball, A. O., Beal, M. G, Chapman, R.W. & Sedberry, G. R. (2007) Population structure of red porgy, *Pagrus pagrus*, in the Atlantic Ocean. Marine Biology, 150, 1321-1332.

Avise, J.C. (1994) Molecular Markers, Natural History, and Evolution. Chapman and Hall, New York, NY, 511.

Beaumont, M. A. & Nichols, R. A. (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London, 263, 1619-1626.

Belkhir, K. (1999) GENETIX. Cited 2 times. v. 4.0. Laboratoire Génome, Populations, Interactions. Centre National de la Recherche Scientifique Unite Mixte de Recherche, Université, Montpelier, France.

Bernal-Ramı´rez , J. H., Adcock, G. J., Hauser, L., Carvalho, G. R. & Smith, P. J. (2003). Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents. Marine Biology, 142, 567-574.

Berry, O., England, P., Fairclough, D., Jackson, G, & Greenwood, J. (2012) Microsatellite DNA analysis and hydrodynamic modelling reveal the extent of larval transport and gene flow between management zones in an exploited marine fish (*Glaucosoma hebraicum*). Fisheries Oceanography, 21(4), 243-254.

Birky, C. W, Fuerst, P. & Maruyamat, T. (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics, 121, 613-627.

Bowen, B. W., Muss, A., Rocha, L. A. & Grant, W. S. (2006). Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus *Centropyge*) indicates a recent invasion from the Indian Ocean. The Journal of Heredity, 97(1), 1-12.

Brander, K. (2009) Impacts of climate change on fisheries. Journal of Marine Systems, 79, 389-402.

Brown, J. R., Beckenbach, A. T. & Smith, M. J. (1993) Intraspecific DNA Sequence Variation of the Mitochondrial Control Region of White Sturgeon (*Acipenser transmon tanus*). Molecular Biology and Evolution, 10(2), 326-341.

Brown, W. M., George, M. Jr. & Wilson, A. C. (1979) Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences, 76, 1967-1971.

Brown, W. M. (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, 95-130.

Brown, W. M., George, M. & Wilson, A. C. (1979) Rapid evolution of animal mitochondrial DNA. Genetics, 76(4), 1967-1971.

Carvalho, G. R. & Hauser, L. (1994). Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology and Fisheries, 4(3), 326-350.

Cassie, R. M. (1956) Early development of the snapper, *Chrysophrys auratus* Forster. Transactions of the Royal Society of New Zealand, 83, 705-713.

Charlesworth, D. & Willis, J. H. (2009) The genetics of inbreeding depression. Nature, 10, 783-796.

Chen, C. A., Ablan, M. C. A., McManus, J. W., Bell, J. D., Tuan, V. S., Cabanban, A. S. & Shao, K. (2004) Population Structure and Genetic Variability of Six Bar Wrasse (*Thallasoma hardwicki*) in Northern South China Sea Revealed by Mitochondrial Control Region Sequences. Marine Biotechnology, 6, 312-326.

Christie, M. R., Marine, M. L., French, R. A. & Blouin, M. S. (2012) Genetic adaptation to captivity can occur in a single generation. Proceedings of the National Academy of Sciences, 109(1), 238-242.

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W. & McCabe, A. M. (2009) The Last Glacial Maximum. Science, 325(5941), 710-714. doi: 10.1126/science.1172873

Connor, R. (2001) Initial Allocation of Individual Transferable Quota in New Zealand Fisheries. In Shotton (ed), Case Studies on the Allocation of Transferable Quota Rights in Fisheries. Rome: FAO.

Conover, D. O., Clarke, L. M., Munch, S. B. & Wagner, G. N. (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. Journal of Fish Biology, 69 (Supplement C), 21-47.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics, 144,2001-2014.

Cowen, R. K., Lwiza, M. K. K., Sponaugle, S., Paris, C. B. & Olson, D. B. (2000). Connectivity of Marine Populations: Open or Closed? Science, 287(5454), 857-859.

David, P., Perdieu, M. A., Pernot, A. F. & Jarne, P. (1997) Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*, Evolution, 51, 1318-1322.

Dent, E. A. & Bridgett, V. M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources vol. 4 (2) pp. 359-361 doi: 10.1007/s12686-011-9548-7

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214-218.

Edmunds, R. C., Herwerden, L. V., Smith-Keune, C. & Jerry, D. R. (2009) Comparative characterization of a temperature responsive gene (lactate dehydrogenase-B, ldh-b) in two congeneric tropical fish, *Lates calcarifer* and *Lates niloticus*. International Journal of Biological Science, 5, 558-569.

Ellegren, H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics, 16(12), 551-558.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. Molecular Ecology, 14, 2611-2620.

Ewing, G. B. (2012) Retrieved from http://www.cibiv.at/~greg/haploviewer on 20 Feb 2012.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1, 47-50.

FAO (2012) The State of World Fisheries and Aquaculture. Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture Department.

Fishery Summary (2012) Retrieved from http://fs.fish.govt.nz/Doc/21790/92_SNA_09%20.pdf.ashx on 20 Feb 2012.

Frankham, R. (1994) Effective population size/adult population size ratios in wildlife: A review. Genetical Research, 66(2), 95-107.

Franklin, I.R. (1980) Evolutionary change in small populations. In Conservation Biology: an Evolutionary–Ecological Perspective (Soule´, M.E. and Wilcox, B.A., eds), Sinauer Associates, 135-150.

Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 147, 915-925.

George, S. B. (1994) Population differences in maternal size and offspring quality for *Leptasterias epichlora* (Brandt) (Echinodermata: Asteroidea). Journal of Experimental Marine Biology and Ecology, 175, 121-131.

Gilbert, D.J., McKenzie, J.R., Davies, N.M. & Field, K.D. (2000) Assessment of the SNA1 Stocks for the 1999–2000 Fishing Year. New Zealand Fisheries, Assessment Report 2000/38, Wellington, New Zealand.

Gilbert, D. J. & Philips, N. L. (2002) Assessment of the SNA 2 and Tasman Bay/Golden Bay (SNA 7) snapper fisheries for the 2001-02 fishing year. New Zealand Fisheries, Assessment Report 2003/45, Wellington, New Zealand.

Gjedrem, T. (2005) Selection and Breeding Programs in Aquaculture. Springer, Dordrecht.

Goudet, J. (1999) PCA-GEN, (version 1.2) Lausanne, Switzerland. Retrieved from www.unil.ch/izea/softwares/pcagen.html on the 20 Feb 2012.

Goudet, J. (2005) FSTAT [Online]. (accessed on November 14, 2012) Available at http://www2.unil.ch/popgen/softwares/fstat.htm

Grüss, A., Kaplan, D. M., Guénette, S., Roberts, C. M. & Botsford, L. W. (2011) Consequences of adult and juvenile movement for marine protected areas. Biological Conservation, 144, 692-702.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology, 59(3), 307-21.

Gutierrez, A. P., Lubieniecki, K. P., Davidson, E. A., Lien, S., Kent, M. P., Fukui, S., Withler, R. E., Swift, B. & Davidson, W. S. (2012) Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. Aquaculture, 358-359, 61-70.

Ha, H. P., Nguyen, T. T. T., Poompuang, S. & Na-Nakorn, U. (2009) Microsatellites revealed no genetic differentiation between hatchery and contemporary wild populations of striped catfish, *Pangasianodon hypophthalmus*. Aquaculture, 291, 154-160.

Harrison, R. G. (1989) Animal mitochondrial DNA as genetic markers in population and evolutionary biology, TREE, 4(1), 6-11.

Hatanaka, A. & Yamada, S. (2006) Isolation and Application of Microsatellite DNA Markers for Pedigree Tracing of Seedlings of Red Sea Bream (Pagrus major). Journal of the World Aquaculture Society, 37 (1), 139-143.

Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramı´rez, J. H. & Carvalho, G. R. (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, 99(18), 11742-11747

Hayes, B., Baranski, M., Goddard, M. E. & Robinson, N. (2007) Optimisation of marker assisted selection for abalone breeding programs. Aquaculture, 264(1-4), 61-69.

Hedgecock, D. (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In Genetics and evolution of aquatic organisms, A. R. Beaumont, ed. London: Chapnian & Hall, 122-134.

Ho, S. & Shapiro, B. (2011). Skyline---plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources, 1111(10), 998-1755.

Hoelzel, A. R., Hancock, J. A. & Dover, G. A (1991) Evolution of the Cetacean Mitochondrial D-Loop Region. Molecular Biology Evolution, 8(3), 475-493.

Idyll, C. P. (1973) The anchovy crisis. Scientific American, 228(6), 22-29.

Jakobsson, M. & Rosenberg, N. A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23(14), 1801-1806.

Jamieson, I. G. & Allendorf, F. W. (2012) How does the 50/500 rule apply to MVPs? Trends in Ecology and Evolution, 27(10), 578-584.

Johnson, M. S. & Black, R. (1982) Chaotic patchiness in an intertidal limpet, *Siphonaria sp.* Marine Biology, 70,157-164.

Kalinowski, S. T. (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. Molecular Ecology Notes, 5, 187-189.

Kennedy, J., Geffen, A. J. & Nash, R. D. M. (2007) Maternal influences on egg and larval characteristics in plaice (*Pleuronectes platessa L.*). Journal of Sea Research, 58, 65-77.

King Salmon (2012) Retrieved from http://kingsalmon.co.nz on the 20 Feb 2012.

Lacson, J. M. & Morizot, D. C. (1991) Temporal genetic variation in subpopulations of bicolor damselfish (*Stegastes partitus*) inhabiting coral reefs in the Florida Keys. Marine Biology, 110, 353-357.

Laikre, L., Palm, S. & Ryman, N. (2005) Genetic Population Implications for Coastal of Fishes: Structure Zone Management. Ambio, 34(2), 111-119.

Langella, O. (2012) POPULATIONS 1.2.31. Population genetic software (individuals or populations distances, phylogenetic trees). Available from http://bioinformatics.org/~tryphon/populations/

Larson, R. J. & Julian, R. M. (1999) Spatial and Temporal Genetic Patchiness in Marine Populations and Their Implications for Fisheries Management. California Cooperative Oceanic Fisheries Investigations, 40, 94-99.

Larsson, L. C., Laikre, L., Palm, S., André, C., Carvalho, G. R. & Ryman, N. (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. Molecular Ecology, 16(6), 1135-1147.

Le Corre, V. & Kremer, A. (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics, 164, 1205-1219.

Leberg, P. L. (2002) Estimating allelic richness: effects of sample size and bottlenecks. Molecular Ecology, 11, 2445-2449.

Liu, Z. J. & Cordes, J. F (2004) DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238, 1-37. doi:10.1016/j.aquaculture.2004.05.027

Liu, G., Zhou, J. & Zhou, D. (2012) Mitochondrial DNA reveals low population differentiation in elongate loach, *Leptobotia elongata* (Bleeker): implications for conservation. Environmental Biology of Fishes, 93, 393-402. DOI 10.1007/s10641-011-9927-7

Lock, K. & Leslie, S. (2007). New Zealand's Quota Management System: a history of the first 20 Years. Public Policy Research (p. 75).

Louzao, M. Igual, J. M., Genovart, M., Forero, M. G., Hobson, K. A. & Oro, D. (2008) Spatial variation in egg size of a top predator: Interplay of body size and environmental factors? Acta Oecologica, 34(2), 186-193.

Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. The Journal of Heredity, 89(3), 238-247.

Luikart, G. & Cornuet, J.-M. (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology, 12(1), 228-237.

Lundrigan, T. A., Reist, J. D. & Ferguson, M. M. (2005) Microsatellite genetic variation within and among Arctic charr (*Salvelinus alpinus*) from aquaculture and natural populations in North America. Aquaculture, 244, 63-75.

Ma, Y., Wang, C., Wang, J., Yang, X., Bi, X. & Xu, L. (2011) Genetic differentiation of wild and hatchery Oujiang color common carp: potential application to the identification of escapees. Fisheries Science, 77, 591-597.

Maruyama, T. & Fuerst, P.A. (1985) Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics, 111, 675-689.

Mccusker, M. R. & Bentzen, P. (2010)Phylogeography of 3 north Atlantic Wolffish species (*Anarhichas spp.*) with phylogenetic relationships within the family *Anarhichadidae*. Journal of Heredity, 101(5), 591-601.

Meuwissen, T. H. E. & Goddard, M. E. (1996) The use of marker halotypes in animal breeding schemes. Genetics Selection Evolution, 28, 161-176.

MFish (2012) Retrieved from http://fs.fish.govt.nz/Page.aspx?pk=7&sc=SNA.

Monteleone, D. M. & Houde, E. D. (1990) Influence of maternal size on survival and growth of striped bass *Morone saxatilis* Walbaum eggs and larvae. Journal of Experimental Marine Biology and Ecology, 140, 1-11.

Morrison, M. (2008). Tracking snapper origins. Water and Atmosphere 16, 4.

Murata, O., Harada, T., Miyashita, S., Izumi, K., Maeda, S., Kato, K. & Kumai, H. (1996) Selective Breeding for Growth in Red Sea Bream. Fisheries Science, 62(6), 845-849.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C., Taranger, G., Ogden, R. & Carvalho, G. R. (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evolutionary Biology, (9)276.

Olsson, J., Mo, K., Florin, A. B., Aho, T. & Ryman, N. (2011) Genetic population structure of perch *Perca fluviatilis* along the Swedish coast of the Baltic Sea. Journal of Fish Biology, 1-16.

Oosterhout, C. V., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4, 535–538.

Padhi, A. (2011) Climate-driven population dynamics of a North American freshwater fish*, Ictalurus furcatus*: inference from the mitochondrial control region sequence data. Ecology of Freshwater Fish, 20, 263-269.

Peakall, R. & Smouse, P.E. (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6, 288-295.

Peakall, R. & Smouse, P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics In press.

Pekkala, N., Knott, K. E., Kotiaho, J. S., Nissinen, K. & Puurtinen, M. (2012) The benefits of inter-population hybridization diminish with increasing divergence of small populations. Journal of Evolutionary Biology, 25, 2181-2193.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of Heredity, 95, 536-539.

Planes, S. & Lenfant, P. (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. Molecular Ecology, 11, 1515-1524.

Plant and Food Research (2012) Retrieved from http://www.plantandfood.co.nz on the 20 Feb 2012.

Posada, D. (2008) jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution, 25(7), 1253-1256.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Proctor, R. & Carter, L. (1989) Tidal and sedimentary response to the Late Quaternary closure and opening of Cook Strait, New Zealand: Results from numerical modeling. Paleoceanography, 4(2), 167-180. doi:10.1029/PA004i002p00167

Python (2012) retrieved from http://pypi.Python.org/pypi/xlwt on the 15 Feb 2012.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.

Reiss, H., Hoarau, G., Dickey-Collas, M. & Wolff, W. J. (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries, 10, 361-395.

Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics, 105, 767-779.

Rioux-Paquette, S. (2011) PopGenKit: useful functions for (batch) file conversion and data resampling in microsatellite datasets. R package, version 1.0. Available at http://cran.r-project.org/web/packages/PopGenKit/index.html.

Rousset, F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources, 8, 103-106.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19 (18), 2496–2497.

Ruzzante, D.E., Wroblewski, J.S., Taggart, C.T., Smedbol, R.K., Cook, D. & Goddard, S.V. (2005) Bay-scale population structure in coastal Atlantic cod in Labrador and Newfoundland, Canada. Journal of Fish Biology, 56(2), 431-447.

Sainudiin, R., Durrett, R. T., Aquadro, C. F. & Nielsen, R. (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. Genetics, 168, 383-395.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, New York.

Sarropoulou, E. & Fernandes, J. M. O. (2011) Comparative genomics in teleost species: Knowledge transfer by linking the genomes of model and non-model fish species. Comparative Biochemistry and Physiology, Part D 6, 92-102

Seafood Industry Council (2011). 'Snapper.' Retrieved from http://www.seafoodindustry.co.nz/n1128,208.html on 15 July 2011.

Sehgal, H. S. & Toor, H. S. (1991) Offspring fitness and fecundity of an Indian major carp *Labeo rohita* (Ham.), in relation to egg size. Aquaculture, 97, 269-279.

Sekino, M., Hara, M. & Taniguchi, N. (2002) Loss of microsatellite and mitochondrial DNA variation in hatchery strains of Japanese flounder *Paralichthys olivaceus*. Aquaculture, 213, 101-122.

Selkoe, K. A. & Toonen, R. J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters, 9, 615-629. doi: 10.1111/j.1461-0248.2006.00889.x

Schlotterer, C. (2004) The evolution of molecular markers – just a matter of fashion? Nature Reviews Genetics, 5, 63-69.

Schuelke, M. (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology, 18, 233-234.

Shanks, A. L. (2009) Pelagic larval duration and dispersal distance revisited. The Biological Bulletin, 216(3), 373-385.

Skaala, O., Hoyheim, B., Glover, K. & Dahle, G. (2004) Microsatellite analysis in domesticated and wild Atlandic salmon (*Salmo salar L.*): allelic diversity and identification of individuals, Aquaculture, 240, 131-143.

Song, N., Zhang, X. & Gao, T. (2011) Genetic variability of eight cultured and two wild populations of Japanese flounder, *Paralichthys olivaceus*, based on the mitochondrial DNA control region. Journal of the World Aquaculture Society, 42(4), 512-521.

Springate, J. R. C. & Bromage, N. R. (1985) Effects of egg size on early growth and survival in rainbow trout (*Salmo Gairdneri* Richardson). Aquaculture, 47, 163-172.

Smith, P. J. & Johnston, A. D. (1985) Glucosephosphate isomerase and α-glycerophosphate dehydrogenase electromorph frequencies in groper *Polyprion oxygeneiosis* from central New Zealand. New Zealand Journal of Marine and Freshwater Research, 19(2), 173-177.

Statistics NZ (2012) Fish Monetary Stock Account: 1996–2009. Retrieved from http://www.stats.govt.nz/browse_for_stats/environment/natural_resources/fish-monetary-stock-account-1996-2009/results.aspx#snapper on the 15 Nov 2012.

Stephenson, R. L. (1999) Stock complexity in fisheries management: a perspective of emerging issues related to population sub-units. Fisheries Research, 43, 247-249.

Tabata, K. & Taniguchi, N. (2000) Differences between *Pagrus major* and *Pagrus auratus* through mainly mtDNA control region analysis. Fisheries Science, 66, 9-18.

Tajima, F. (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics, 123, 585-595.

Takezaki, N. & Nei, N. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics, 144, 389-399.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 2731-2739.

Taniguchi, N. & Perez-Enriquez, R. (2000) Genetic evaluation of broodstock for aquaculture of red sea bream by DNA markers. In: Fingerman, M. and Nagabhushanan, R. (eds.), Recent Advances in Marine Biotechnology, Volume 4. Aquaculture, Science Publishers, New Hampshire, 1-16.

Tave, D. (1999) Inbreeding and Management. FAO Fisheries Technical Paper, 392.

Thorrold, S. R., Latkoczy, C., Swart, P. K. & Jones, C. M (2000) Natal Homing in a Marine Fish Metapopulation. Science, 291, 297-299.

Vawter, L. & Brown, W. M. (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science, 234(4773), 194-196.

Walters C. & Maquire J. (1996) Lessons for stock assessment from the northern cod collapse. Reviews in Fish Biology and Fisheries, 6, 125-137.

Wang, J. (2002) Application of the one-migrant-per-generation to conservation and management. Conservation Biology, 12(2), 332-343.

Weir, B. S. & Cockerham, C. C. (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution, 38(6), 1358-1370.

Weersing K. & Toonen R. J. (2009) Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series, 393, 1-12.

Willis, T. J., Millar, R. B. & Babcock, R. C. (2003) Protection of exploited fish in temperate regions: high density and biomass of snapper *Pagrus auratus* (Sparidae) in northern New Zealand marine reserves. Journal of Applied Ecology, 40(2), 214-227.

Willis T. J., Parsons D. M., & Babcock R. C. (2001) Evidence for long-term site fidelity of snapper (*Pagrus auratus*) within a marine reserve. New Zealand Journal of Marine and Freshwater Research, 35, 581-590

Winkler, S. (2004) Lichenometric dating of the 'Little Ice Age' maximum in Mt Cook National Park, Southern Alps, New Zealand. The Holocene, 14(6), 911-920.

Wratt, D. & Mullan, B. (2012) Climate change scenarios for New Zealand. Retrieved from http://www.niwa.co.nz/our-science/climate/information-and-resources/clivar/scenarios on 9 Jan 2013.

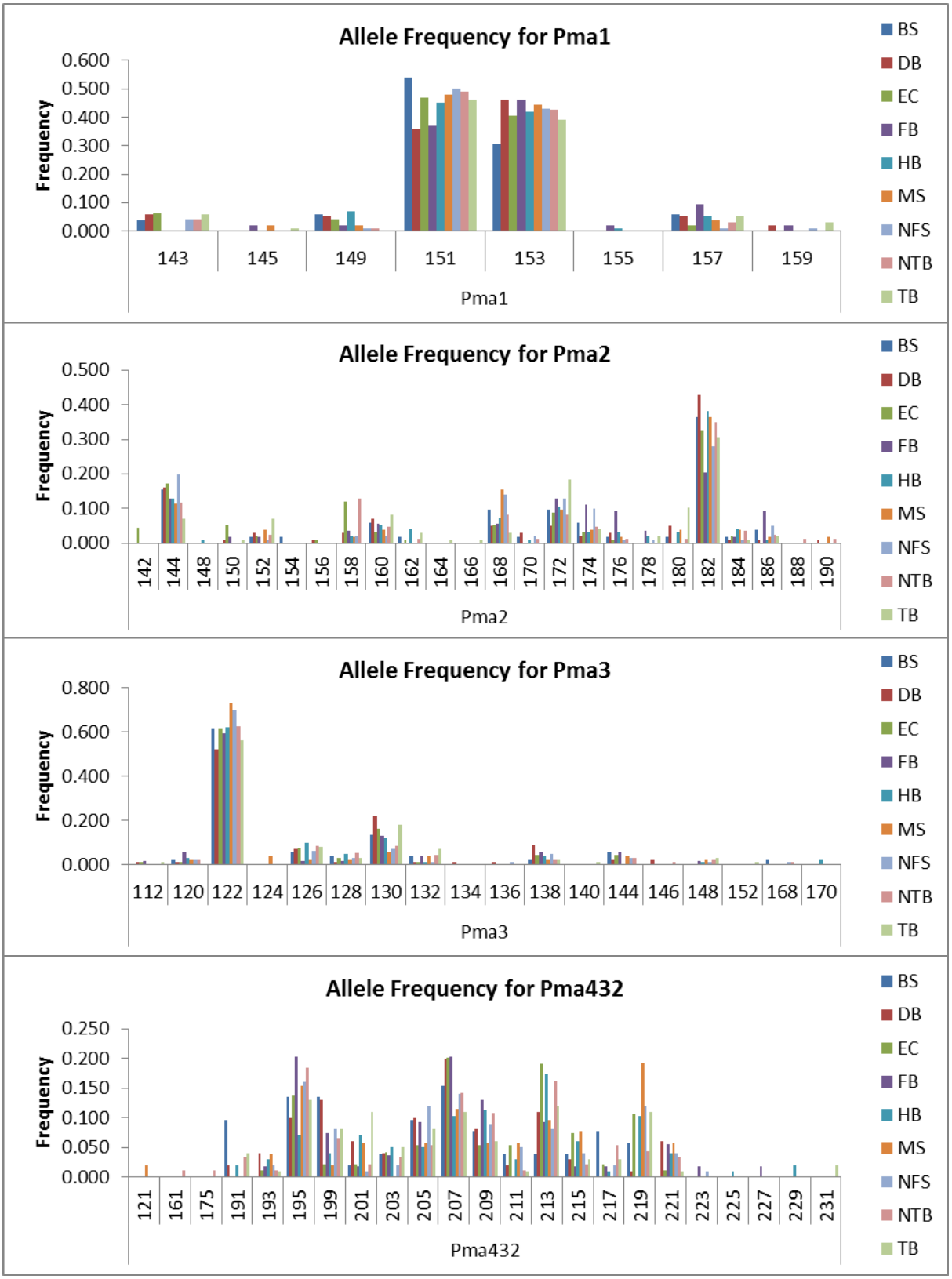
Wright, S. (1931) Evolution in mendelian populations. Genetics, 16, 97-259.

Wright, S. (1943). Isolation by Distance. Genetics, 28(2), 114-138.

Wright, S. (1951) The genetical structure of populations. Annals of Eugenics, 15, 323-354.

Zhang, Y., Wang, S., Li, J., Zhang, X., Jiang, L., Xu, P., Lu, C., Wan, Y. & Sun, X. (2012) Primary genome scan for complex body shape-related traits in the common carp *Cyprinus carpio*. Journal of Fish Biology, 1-16.

Appendix A: Microsatellite Allele Size Distributions



Appendix B: Microsatellite Allele Frequency Distributions

Broodstock
Doubtless Bay
East Coast
Foxton Beach
Hawkes Bay
Marlborough Sounds
North Taranaki Bight
Tasman Bay
North Farewell Spit
Prop. Alleles
Allele Frequency