GENETIC STRUCTURE OF *PLEUROBRANCHAEA MACULATA* IN NEW ZEALAND

A thesis presented in partial fulfilment of the
requirements for the degree of **Doctor of Philosophy** (**PhD**) in Genetics

The New Zealand Institute for Advanced Study
Massey University, Auckland, New Zealand

YEŞERİN YILDIRIM

2016

ACKNOWLEDEGEMENTS

I have a long list of people to acknowledge, as my PhD project would not have been possible without their support.

Firstly, I would like to thank my supervisor, Professor Paul B. Rainey (New Zealand Institute of Advanced Study, Massey University), for providing me with the opportunity to join his research group. He gave me constant support, insightful guidance, valuable input, and showed me how to think like a scientist. I would also like to thank my co-supervisor, Dr Craig D. Millar, who welcomed me into his genetics laboratory at the Department of Biological Sciences at the University of Auckland whenever I needed help. He guided me patiently right from the beginning of my PhD project, was generous with his time, encouraging, and taught me how to troubleshoot where necessary. I would like to extend a special mention to Selina Patel, a very talented technician at Dr Millar's Lab who shared her extensive technical and theoretical knowledge with me generously, but also allocated considerable time to help me progress with my research. My study was kicked off by Selina Patel and Dr Millar's willingness to sequence my study organism's genome in the facilities of the "Centre for Genomics, Proteomics and Metabolomics" at the at the University of Auckland. Thank you also to my co-supervisor, Professor Marti Anderson (Massey University), for her guidance, her helpful contributions, for not only teaching me how to analyse my data but also for being involved in the analysis herself, and patiently answering my very naive questions about statistical methods. Her explanations helped me to understand that these very sophisticated methods are not something to be scared of – they are actually fun. Many thanks to my other co-supervisor, Professor Nigel French, (Massey University), for supporting me whenever I asked for help, and encouraging and backing me in the Allan Wilson Centre meetings.

My project was very much dependent on the divers and researchers who collected *Pleurobranchaea maculata* samples for me. I am grateful to Dr Susanna Wood and Dr David Taylor (Cawthron Institute, Nelson) for supplying me with numerous sea slug samples, sharing their extensive knowledge of *P. maculata*, discussing my findings, providing access to unpublished manuscripts, and introducing me to the people who

helped me to obtain more samples. I would have been lost without them. I would like to acknowledge the following people and/or institutes who either collected samples, put me into contact with people who provided samples or helped with the shipment of the samples: Mike McMurtry (Auckland Regional Council); Lauren Salvitti, Dudley Bell, Warrick Powrie and David Culliford (University of Waikato, Hamilton); Richard Huges and Bakhti Patel (Leigh Marine Laboratory, the University of Auckland); Severine Hannam and Dr Wilma Blom (Auckland Museum); Shane Genage (Victoria University of Wellington), Steve Journee (The Dive Guys, Wellington); Don Morrisey, Matthew Smiths and Stephen Brown (The National Institute of Water and Atmospheric Research); Dr Mark Stevens (South Australian Museum) and Te Papa Museum (Wellington).

Dr Paul McNabb (Cawthron Institute) conducted the tetrodotoxin assay for some of the samples. Serena Khor (The University of Waikato) extracted tetrodotoxin from some samples for the toxicity assay. I thank them both.

I would like to extend my gratitude to everyone who helped me with the data analyses and interpretation. I need to mention the contribution of my supervisors and Selina Patel again in this area. Virginia Moreno-Puig (Massey University), who has just submitted her thesis in which she utilised a population genetics tool, was a fount of knowledge and took a great deal of time to reflect on my questions with me. I would like to thank Dr Peter Ritchie (Victoria University of Wellington) and my colleagues Dr Stephen Ritchie, Jenny Herzog, Christina Straub, Dr Katrin Hammerschmidt, Dr Libby Liggins, Dr Peter Deines, Oliver Hannaford and Luca Bütikofer for their help and suggestions.

Many people helped me while I was writing my thesis. Many thanks once more to my main supervisor Professor Rainey for his patience while reading my thesis, his guidance, his outstanding comments and his corrections. I am immensely grateful to Dr Honour McCann (NZIAS, Massey University) as she critically reviewed, proofread and edited Chapter 1. I thank Dr Gayle Ferguson (INMS, Massey University) as she edited Chapter 2, but also Jacque Mackenzie (Institute of Veterinary, Animal & Biomedical Sciences, MU) as she proofread an earlier version of the same chapter. Thank you to Adam Bedford, Tatyana Pichugina (the University of Auckland), and my colleagues

Chhavi Chawla and Elena Colombi for editing and proofreading some parts of my thesis. I am very grateful to Joanna Niederer as well for copy-editing my whole thesis.

On a more personal level, I thank my mother Çakır Ayşe Selçuk and my brother Övgün Yıldırım, who have never failed to support me from the other side of the world. I know that I will always have a warm home waiting for me. I am very grateful to Adam Bedford, who encouraged me and gave me sound advice whenever I fell into despair. He was the one who suffered my whims, but he was also at my side during the many happy days with lots of laughs and adventures. I would also like to thank Rashmi Ramesh Iyer and Setareh Mokhtari, who were my first flatmates and my first friends in New Zealand. I would not have been able to survive in New Zealand without their support. I am grateful to Jenny Herzog for her friendship and cheerfulness, her wise observations and encouragement. My special thanks goes to one of the most wonderful and sophisticated people I have ever met, Marleen Bailing. She has always been there without hesitation whenever I needed help. She also showed me many tricks to improve my climbing skills. Thank you to Chhavi Chawla for being such a kind-hearted, sincere, wise and supportive friend. I wish you had joined our lab earlier. Thank you to my office mates and friends Yuriy Pichugin and Yunhao Liu. They always listened to my complaints, understood me very well and offered me practical advice and valuable opinions on my work. Thank you to all the Rainey Lab members as well as to Claire-Marie Loudon, Tatyana and Luca for making my life enjoyable. I have to stop here otherwise this list will never end.

My study was funded by the Allan Wilson Centre, Institute for Advanced Study at Massey University, and the Auckland Regional Council. I am very grateful to these institutes, as this research could not have happened without their support.

ABSTRACT

AIMS

The grey side-gilled sea slug (*Pleurobranchaea maculata*), which is native to the western and south Pacific, is known to contain high concentrations of tetrodotoxin (TTX). *P. maculata* populations around New Zealand exhibit individual, spatial and temporal differences in TTX concentration, but the origin of TTX in *P. maculata* is not fully understood. The main goal of my PhD project was to examine the genetic structure and demographic history of *P. maculata* populations from different regions in New Zealand and to clarify whether there is a correlation between variability in TTX concentrations and genetic structure.

METHODS

A sample of 146 *P. maculata* individuals were collected from three populations from the north-eastern North Island (Ti Point-TP, Auckland-AKL and Tauranga-TR), one population from the southern North Island (Wellington-WL) and one population from the northern South Island (Nelson-NL). TP, AKL and TR were designated the "Northern cluster", whereas the WL and NL population were labelled as the "Southern cluster" due to the relative geographical locations of these clusters. Twelve nuclear microsatellite markers that were developed based on shotgun sequence were obtained from the genome of *P. maculata*. The markers were used to analyse the genetic structure of *P. maculata* populations. The mitochondrial cytochrome c oxidase subunit I (1153 bp) and cytochrome b (1060 bp) genes were also partially sequenced in *P. maculata* individuals.

RESULTS AND MAIN CONCLUSIONS

The microsatellite data reveal high genetic diversity and lead to the rejection of the hypothesis of panmixia: populations from the Northern cluster are highly connected but significantly differentiated from the Sothern cluster. A weak differentiation was also observed between the WL and NL populations. The two populations correlate with regional variations in TTX concentrations: the Northern cluster populations contain highly toxic individuals, whereas the Southern cluster (WL and NL) populations

harbour either slightly toxic or non-toxic populations. The disjunction between the Northern and Southern clusters can be explained by biogeographical barriers specific to New Zealand but also with a stepping stone model. The geographical gap between the sampling locations made it impossible to draw firm conclusions as to the origin of the disjunction. The mtDNA sequence data reveal high haplotype diversity, low nucleotide diversity and a star-shaped haplotype network. These data can be explained by a population expansion dating back to the Pleistocene era. All the sampling locations are significantly differentiated from each other according to mtDNA data. Given that microsatellite and mitochondrial sequences evolve at different rates, incomplete linkage sorting is expected to be completed for mtDNA before, which should be reflected in a more pronounced structure for mtDNA markers where members of the populations have diverged recently. Although this may explain the geographical conflict between the microsatellite and mtDNA data, it is necessary to consider the possibility that the discordance between microsatellite markers and mtDNA may be in part attributable to the relatively small sample size.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1: GENERAL INTRODUCTION

In late 2009, an otherwise little-known sea slug attracted attention after it was implicated in several dog deaths on beaches around Auckland, New Zealand (NZ). Forensic analyses revealed the deaths were a consequence of tetrodotoxin (TTX) poisoning associated with ingestion of *Pleurobranchaea maculata* (grey side-gilled sea slug, Figure 1.1) (McNabb et al., 2010), which is native to NZ, southern and south-eastern Australia, Tasmania (http://bie.ala.org.au/species/), China, Sri Lanka, and Japan (Willan, 1983). Figure 1.2 shows the global distribution of the species. This was the first time that TTX was reported in NZ and in a soft-bodied sea slug (Heterobranchia) (Willan, 1983).

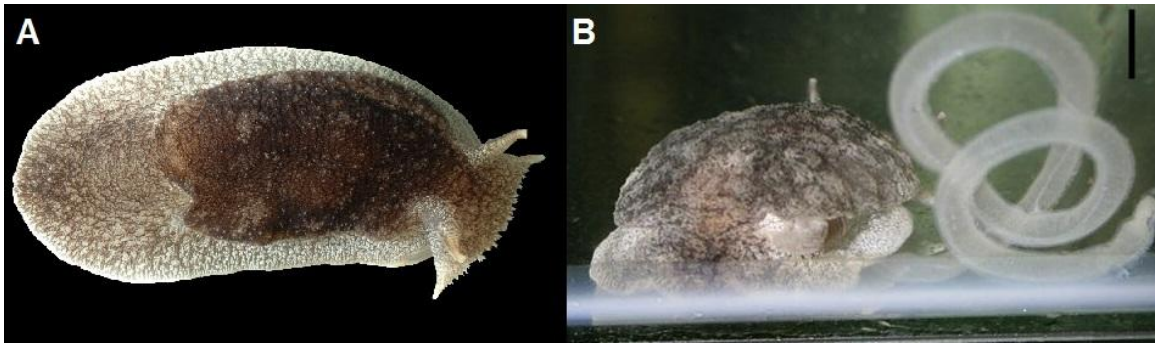


Figure 1.1 Images of *P. maculata*. A) An adult *P. maculata* individual in a tank (© Richard Taylor, the University of Auckland). B) An adult laying an egg mass. Scale bar = 1 cm. The image was adapted from McNabb et al. (2010).

Pleurobranchaea maculata is an opportunistic carnivore, feeding on various invertebrates such as sea anemones, marine worms and other molluscs (Willan, 1983), and it is commonly found in a diverse range of habitats; from shallow subtidal areas (Wood et al., 2012b) to intertidal areas, and from sandy sediments to rocky reefs (Willan, 1983). Additional studies have shown differences in concentrations of TTX-associated *P. maculata* populations from different regions of NZ (Wood et al., 2012b). Significant individual and seasonal variations in TTX have been observed for populations from various regions of the North Island, whereas no TTX was found in individuals from the South Island populations (McNabb et al., 2010; Wood et al., 2012b). However, the reason for these variations remains unclear.

Figure 1.2 The localities where *P. maculata* has been spotted. The location information was taken from Grove (2015); http://collections.tepapa.govt.nz/, http://collections.ala.org.au/; Wood et al., (2012b); Farias et al. (2015); Morley and Hayward (2015) and the sampling locations from this study. The records from Japan, China and Sri Lanka (Willan, 1983) were not included into the map, as these records need confirmation (Rudman, 1999). Samples from Argentina were recorded only recently, in 2009 (Farias et al., 2015).

TTX is a small, low molecular weight toxin with a complex molecular structure (Hanifin, 2010) (Figure 1.3). It inhibits action potential by binding to voltage-gated Na^+-channels on muscle and nerve cells, which results in paralysis and death due to respiratory failure (Zimmer and Ferrer, 2007). Ingestion of approximately 2 mg of TTX is lethal for humans (Noguchi and Ebesu, 2001).TTX is so named because it was first isolated and identified in puffer fish (or fugu) belonging to the order of Tetraodontiformes (Fuhrman, 1986). TTX and/or its derivatives were later reported in both marine and terrestrial animals from a variety of phyla such as flatworms, ribbon worms, bivalves, snails, blue-ringed octopus, crabs, arrow worms, starfish, newts, frogs and garter snakes (Chau et al., 2011; Noguchi

and Arakawa, 2008; Stokes et al., 2014). The structure of TTX was determined some years ago (Woodward, 1964), yet its biosynthesis remains unknown. There are three hypotheses regarding its origin: endogenous production, accumulation through the food chain or production by symbiotic/commensal microorganisms (Chau et al., 2011; Noguchi and Arakawa, 2008; Williams, 2010). However, the origin of TTX has remained inconclusive, despite attempts to uncover it.

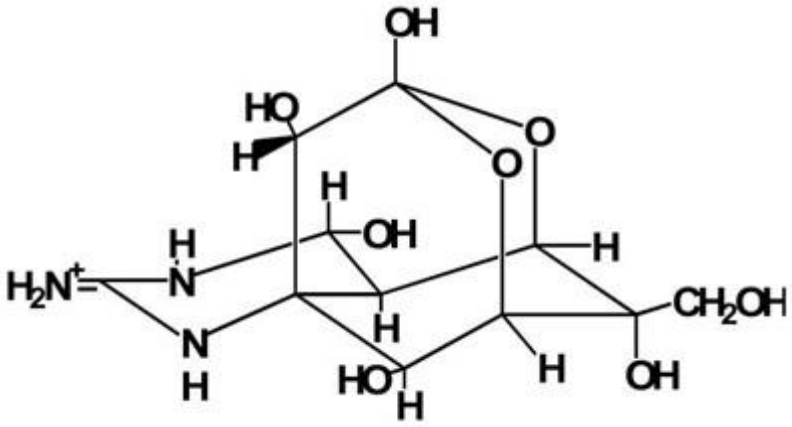


Figure 1.3 The structure of TTX . Figure taken from Noguchi and Arakawa (2008).

Understanding the genetic structure of *P. maculata* populations from different regions of NZ may provide one route to clarifying the origins and evolution of TTX-containing sea slugs. The central aim of my PhD is to reveal the population genetic structure of *P. maculata*.

The remaining sections of this introductory chapter address five key topics. The first section provides a detailed description of what is known about *P. maculata* toxicity. The second section introduces the principles and tools of population genetics. The third section is a comprehensive review of previous population genetic studies of NZ marine invertebrates, focusing in particular on factors that may also shape the distribution of genetic variation in *P. maculata*. The last part of the introduction presents the aims of this thesis, and describes how the further chapters are structured to achieve these aims.

1.1 *P. MACULATA* AND ITS ASSOCIATION WITH TTX

1.1.1 Variability in TTX levels

The recent discovery of TTX in *P. maculata* in late 2009 after it was implicated in dog poisoning incidences on Auckland beaches was surprising considering the fact that this species has been known in NZ for a long time (McNabb et al., 2010). It is an endemic species first described in 1832 (Quoy and Gaimard). On discovery of TTX in *P. maculata*, subsequent work has focussed on the distribution of TTX and the possible source of TTX in *P. maculata*. One study (Wood et al., 2012b) investigated spatial, individual and temporal differences in TTX concentration in *P. maculata* populations around the North and South Islands. The study found that *P. maculata* populations from the northern North Island regions (Whangarei, Auckland, Tauranga) had high TTX concentrations (the highest average being 368.7 mg/kg per individual), while the southern North Island region (Wellington) had very low concentrations (~2.2 mg/kg) (Wood et al., 2012b). The South Island samples (Nelson and Kaikoura) had either no, or trace amounts of TTX (McNabb et al., 2010; Wood et al., 2012b). Significant individual variations were observed in highly toxic populations. For example, the concentration differences between individuals could be as high as 62-fold. In addition, monthly sampling for six months from one region, Narrow Neck beach in Auckland, revealed that the average TTX concentration gradually decreased from June to December, indicating a seasonal variation in the levels of TTX (Wood et al., 2012b).

1.1.2 Ecological role of TTX in *P. maculata*

Some evidence suggests *P. maculata* utilises TTX as a chemical defence mechanism. Observation of a significant number of egg masses during the period when toxin levels peaked in TTX-associated *P. maculata* populations indicates a relationship between spawning time and a high toxicity level (Wood et al., 2012b). Secondly, adults can transfer TTX to their eggs, and the slowest rates of depuration of TTX take place in gonads when toxic *P. maculata* individuals are fed a non-toxic diet (Wood et al., 2012a). The second highest TTX concentrations (after the stomach) have been observed in the mantle in TTX-

fed sea slugs (Khor et al., 2014). Salvitti et al. (2015c) investigated intracellular distribution of TTX in several tissues of *P. maculata* using a monoclonal antibody-based immunohistochemical technique. The results showed that TTX is localised strongly in the mantle, oocytes and follicles of gonad tissue in *P. maculata* adults, but also in the eggs and larvae. All these findings suggest that TTX plays a defensive role in *P. maculata* (Salvitti et al., 2015c; Wood et al., 2012a; Wood et al., 2012b).

1.1.3 Is the source of TTX dietary?

Although the ecological role of TTX in *P. maculata* is generally understood, its origin is still not understood – as is the case for other TTX-containing organisms. One possibility is endogenous production by sea slugs (Wood et al., 2012a; Wood et al., 2012b). However, due to the complex molecular structure of TTX and its presence in organisms that belong to at least 14 different, environmentally diverse phyla, the convergent evolution of a common gene or gene clusters that could have been involved in biosynthesis of the toxin seems highly unlikely (Chau et al., 2013). The second possibility is accumulation through the food chain: there is some evidence supporting this hypothesis (Khor et al., 2014; Wood et al., 2012a). For example, Khor et al. (2014) fed *P. maculata* individuals from nontoxic regions on a diet containing a known amount of TTX that was prepared from toxic *P. maculata* individuals, and measured their TTX levels at different time-points for up to 39 days. The results showed that non-toxic *P. maculata* individuals could accumulate high concentrations of TTX similar to the levels that were observed in wild populations without any observable negative side effects on the slugs' health. TTX concentrations could be detected in all organs even as early as 1 hour after feeding, which is evidence of how rapidly the toxin is transported. Less variability in TTX concentrations among individuals that were harvested at the same time compared to wild toxic populations can explain the variability in TTX concentrations in natural populations: consumption of differential amounts of toxic food leads to variation in TTX concentrations (Khor et al., 2014). Similarly, when toxic *P. maculata* individuals were kept in captivity on a TTX-free diet, the average TTX concentrations decreased rapidly over time (Wood et al., 2012a). Additionally, the digestive tissues along with the gonads and mantle contain the highest TTX concentrations among

several investigated tissue types (Khor et al., 2014; Wood et al., 2012a). A monoclonal antibody-based immunohistochemical technique also confirmed the localisation of TTX in the digestive tissue of *P. maculata* (Salvitti et al., 2015c). These findings support the hypothesis that TTX in *P. maculata* originates from a dietary source.

Possible toxic food sources need to be identified to confirm that TTX in wild *P. maculata* populations is dietary in origin. Recently, Salvitti et al. (2015a) discovered that a flatworm *Stylochoplana* species, which cohabits with P. maculata in the Tauranga region, contains high concentrations of TTX (~380 mg/kg). Real time PCR assays of foregut specimens showed that *P. maculata* preys on this *Stylochoplana* sp. Seasonal variations in TTX concentrations in *Stylochoplana* sp accord with those of *P. maculata*, and mass calculations showed that *P. maculata* might accumulate the observed TTX concentrations by preying on *Stylochoplana* sp. However, the absence of this *Stylochoplana* sp. in geographical areas where other dense toxic *P. maculata* populations are observed implies that *Stylochoplana* sp. is not the only TTX source for *P. maculata* (Salvitti et al., 2015a). Environmental surveys at Narrow Neck and Illiomama Rock, which harbour *P. maculata* populations with high TTX concentrations, identified only very low TTX concentrations in the sand dollar *Arachnoides zelandiae* (0.25 mg/kg) and trace amounts in the gastropod *Turbo smaragdus*, crab *Macropthhalmus hitipes* and corraline turf algae *Carollina officinalis* (Khor et al., 2014). Additionally, bivalve *Paphies australis* samples from Whangapoua, which also harbours toxic sea slug populations, were found to contain very low TTX concentrations (0.80 mg/kg, McNabb et al., 2014). The TTX concentrations measured in these organisms were not high enough to cause the high concentrations of TTX observed in natural *P. maculata* populations considering the short lifespan (<1 year) of *P. maculata* (Khor et al., 2014; McNabb et al., 2010). There may be other as yet undiscovered toxic dietary resources (Khor et al., 2014). It is also curious how these other marine organisms picked up TTX. Given that they all belong to different taxa, it is very unlikely that they produce TTX endogenously. This suggests that the ultimate source of TTX is ubiquitous and possibly prokaryotic.

1.1.4 Bacterial origin of TTX

Another possibility regarding the origin of TTX in *P. maculata* is that commensal microorganisms are responsible for its biosynthesis (Chau et al., 2013; Salvitti et al., 2015b; Wood et al., 2012a; Wood et al., 2012b). The presence of TTX in organisms from taxonomically diverse ranges and combined with the toxin's complex molecular structure strengthens the hypothesis that the ultimate TTX source is biosynthesis by bacteria (Chau et al., 2013). If this is the case for *P. maculata*, the abundance of these microorganisms can vary with oceanic currents or with any other environmental factors such as temperature differences, possibly resulting in the clear toxic and non-toxic population cut-off between the North and South Island (Wood et al., 2012a; Wood et al., 2012b). It has been claimed that many bacterial species isolated from TTX-bearing animals or from deep-sea sediments produce TTX (Chau et al., 2011). However, the reliability of these studies is uncertain because the methods used to identify TTX are not specific to TTX (Chau et al., 2013; Matsumura, 1995; Matsumura, 2001; Williams, 2010). In a recent study, Chau et al. (2013) cultured numerous bacterial species from a TTX-associated blue-ringed octopus species (*Hapalochlaena* sp) and *P. maculata* showing that these species harboured diverse culturable microbial communities. However, they could not detect TTX in any of the bacterial isolates using a TTX-specific assay. Artefacts stemming from experimental design and shifting from natural environmental conditions could have led to the loss of TTX-production (Chau et al., 2011).

Very recently, Salvitti et al. (2015b) isolated bacterial strains from three toxic *P. maculata* and *Stylochoplana* sp. individuals collected from Tauranga (NZ). They found 16 unique strains in *P. maculata* belonging to eight genera (*Pseudomonadales*, *Actinomycetales*, *Oceanospirillales*, *Thiotrichales*, *Rhodobacterales*, *Sphingomonadales*, *Bacillales*, and *Vibrionales*) and one strain from *Stylochoplana* sp. belonging to the *Vibrionales* strain. These strains were cultured in marine broth for four days and the TTX concentrations were measured using a C9-based LC-MS assay (Salvitti et al., 2015b), which can detect TTX precursor or degradation products that form the carbon backbone of TTX (C9: 2-amino-6-

(hydroxymethyl)quinazolin-8-ol) (McNabb et al., 2010). Similar to the results of Chau et al. (2013), the test did not identify potential TTX- or C9-based products in any isolates.

Although the studies on TXX production pathways have led to inconclusive results, some pathways have been proposed based on the predicated biosynthetic pathway of structurally similar toxins, including saxitoxin, phaseolotoxin, and cylindrospermopsin, which are produced by several bacteria and/or dinoflagellate species. These pathways propose that assembly of TTX is performed by a hybrid non-ribosomal peptide synthetase (NRPS) / polyketidesynthases (PKS) enzyme module that may also include an amidinotransferase (AMT) (Chau et al., 2011; Chau et al., 2013). Chau et al. (2013) investigated the presence of AMT, NRPS and PKS genes via PCR screening of bacterial isolates from *Hapalochlaena* sp and *P. maculata*. They found at least one of these targeted gene types in some *Nautellaitalica* and *Pseudoalteromonas* isolates from *Hapalochlaena* sp and some *Shwanella, Alteromonas* and *Pseudoalteromonas* isolates from *P. maculata*. It is possible that the pathways leading to TTX biosynthesis can be investigated by focusing on these bacterial species and the genes predicted to be responsible.

Studies that have reported isolation of TTX-producing bacterial strains noted that the single strains isolated produced significantly lower concentrations of TTX in vitro than were present in the host organisms (Cheng et al., 1995; Hwang et al., 1989; Lee et al., 2000; Wang et al., 2008; Yang et al., 2010). One suggestion is that this is because the host supplies inducing factors that are absent in vitro (Chau et al., 2011; Proksch et al., 2002). To address this, Salvitti et al.(Salvitti et al., 2015b) cultured bacterial communities obtained from toxic *P. maculata* individuals in marine broth with and without *P. maculata* tissue extracts for up to 14 days. TTX assays performed at ten time points showed no evidence of TTX or its backbone product, C9, after day zero. They also checked shifts in the bacterial community during the period of growth by rRNA intergenic spacer analysis, which is a PCR-based method utilised to estimate species richness and the composition of bacterial prokaryotic communities. However, they did not find a significant shift in the bacterial community structure over the first four days and yet such a change was expected had a

TTX-producing strain been present. Consequently, none of these studies found evidence to support the hypothesis that TTX in *P. maculata* or *Stylochoplana* sp. is bacterial in origin (Chau et al., 2013; Salvitti et al., 2015b). It is worth noting that very few marine microorganisms are culturable. Unculturable bacterial species might be direct producers of TTX or they might supply precursors or inducers of TTX-production to other bacterial species that have already been identified (Chau et al., 2011).

1.1.5 TTX resistance in animals

Voltage-gated sodium channels (Na_v), which are mainly expressed in nerve and muscle tissues, are responsible for creating an action potential along excitable membranes by selectively permitting Na^+ influx into the cell (Geffeney and Ruben, 2006). TTX binds to the Na_v channels and blocks the Na^+ influx (Jost et al., 2008). TTX-bearing animals can accumulate high concentrations thanks to their high resistance to the toxin (Noguchi and Arakawa, 2008). To date, many amino acid substitutions, which affect the binding affinity of TTX to the channels and render resistance, have been identified in the Na_v channels of TTX-resistant organisms (Jost et al., 2008). Proteins binding to TTX may also confer resistance by neutralising the toxin and/or play a role in transferring and transporting TTX (Hwang and Noguchi, 2007; Williams, 2010). Such TTX-binding proteins have been isolated from several organisms such as the shore crab *Hemigrapsus sanguineus* (Nagashima et al., 2002), the gastropods *Polinices didamy*, *Natica lineata*, *Oliva miniacea*, *O. mustelina* and *O. hirase* (Hwang and Noguchi, 2007) and the puffer fish species *Takifugu niphoble* (Matsui et al., 2000), *T. rubribes* (Williams, 2010) and *Fugu pardalis* (Yotsu-Yamashita et al., 2001).

There is little understanding of how *P. maculata* gained resistance to TTX as the genetic variations in the molecular structure of the Na_v channels which may confer TTX-resistance to *P. maculata* have not been studied. The ability of *P. maculata* to accumulate high concentrations of TTX and to localise it to specific tissues, such as the mantle and eggs, suggest that TTX-binding proteins are involved in transporting the toxin (Khor et al., 2014; Salvitti et al., 2015c). Furthermore, rapid depuration of TTX when *P. maculata* individuals

are fed on a non-toxic diet and differential depuration rates between organs also suggest the existence of TTX-binding protein(s) in *P. maculata* (Wood et al., 2012a).

In summary, the exact source of TTX in *P. maculata* remains unknown. There is strong evidence suggesting that the source is dietary in origin, yet accumulation via the food chain is not the only source of TTX for *P. maculata* (Khor et al., 2014; Salvitti et al., 2015a; Wood et al., 2012a). It is likely that an unidentified bacterial source is the main producer (Salvitti et al., 2015b).

1.2 POPULATION GENETICS: A TOOL TO STUDY POPULATIONS

1.2.1 Factors shaping populations

Population genetics encompasses theoretical frameworks for understanding the distribution of genetic variation within species, and investigates the evolutionary processes underlying the observed variations (Evanno et al., 2005). Population genetic analysis forms the backbone of this thesis as I used a range of tools to investigate the genetic structure of *P. maculata* populations in NZ. This section introduces the theory of population genetics, the evolutionary factors that shape populations and how genetic tools can help to understand the biology of *P. maculata*.

1.2.1.1 Mutation

A population can be defined as individuals that mate with each other and share a common gene pool that included all genetic variations in a defined location and time. Populations are not stable; their genetic composition and size are subject to changes over time due to ecological and evolutionary processes (Drenth, 1998). Micro-evolutionary processes, including mutation, selection, genetic drift, migration, recombination and non-random mating patterns, are the drivers of genetic variation in populations. Among these evolutionary processes, mutation is the ultimate source of genetic variation as it creates random variants (alleles) that can be transferred to offspring (Brinkman and Leipe, 2001). Mutations affect organism fitness beginning with those that are lethally detrimental,

through to strongly and slightly detrimental mutations, to neutral mutations that have no or very slight effects, and then to mildly or strongly advantageous mutations conferring adaptation to the environment (Eyre-Walker and Keightley, 2007).

1.2.1.2 Natural selection

Natural selection is the process that changes the frequency of different alleles in a population over generations. The effect of selection is dependent on the nature of the mutations (Eyre-Walker and Keightley, 2007). Detrimental mutations are generally subjected to negative selection pressure, which leads to their elimination. Favourable mutations are preserved due to positive selection pressure, whereas neutral mutations can be lost or fixed in populations by genetic drift or recombination events (Brinkman and Leipe, 2001; Hedrick, 2011).

1.2.1.3 Recombination and mating systems

Mating systems have evolutionary consequences. Asexual reproduction produces offspring that carry a copy of their parents' genome, whereas sexual reproduction produces offspring that carry a mixture of their parents' genomes. Apart from mixing the genomes of parents, recombination events during sexual reproduction also contribute to genetic variation in a population. Recombination is the genetic process by which DNA segments are broken and recombined to produce new combinations of alleles during meiosis; herewith, the alleles of the parents will be shuffled in their offspring (Lowe et al., 2009).

If every female gamete in a sexually reproducing organism has an equal chance of being fertilised by every male gamete, the population mates randomly. However, individuals generally choose their mates based on particular traits, which brings about non-random mating. One form of non-random mating is inbreeding, where individuals are more likely to mate with individuals that have similar genotypes than with ones with different genotypes for certain traits. A less common type is outbreeding, where individuals are more likely to mate with individuals that have different genotypes than with ones with similar genotypes. Inbreeding leads to a reduction in variation, whereas outbreeding increases it (Burley,

1983). The evolutionary consequences of the sexual reproduction mode can be observed in hermaphrodites. Hermaphrodite organisms have both male and female reproductive organs during their life span. A very rarely observed mode of hermaphroditism is autogamy or self-fertilisation mode (e.g. Ghiselin, 1969), which leads to inbreeding, may result in a decrease in fitness because it has a similar effect to asexual reproduction (Barrett, 2014). On the other hand, out-crossing hermaphroditism provides the opportunity for mixed genomes, thereby increasing the genetic variability and diversity (Hamrick et al., 1990)

Recombination can separate two physically close loci; however, the likelihood of a recombination event decreases as the distance between the loci decreases. This results in a phenomenon whereby two loci that are on the same chromosome and physically close to each other tend to be inherited together. Therefore, two particular alleles at such loci may be found together more frequently in different individuals than two loci segregating independently. This non-random association of alleles at different loci is termed linkage disequilibrium (LD). If the in-between distance is far enough, two loci will become unlinked for independent segregations. LD between two alleles will decrease in subsequent generations due to recombination events at the formation of each gamete (Hedrick, 2011; Slatkin, 2008).

1.2.1.4 Genetic drift

Genetic drift is a process that changes the allele frequencies in subsequent generations due to random sampling. This stems from the fact that only a small proportion of the produced gametes give rise to the next generation. Genetic drift leads to loss of genetic variation due to fixation or loss of alleles, especially for neutral mutations (Hartl and Clark, 2007). Drift has a stronger effect on small populations than on large ones because the changes due to chance events are more pronounced in small populations. Such populations only have a small proportion of the alleles of the original population, and particular alleles can be over-represented or under-represented just by chance. Furthermore, chance factors that affect mating and survival success, where selection has no effect, are other causes of genetic drift (Hartl and Clark, 2007; Lowe et al., 2009; Masel, 2011). Mutations and genetic drift will

continue to change the allele frequencies even in the absence of any selection pressure. Selection acts on gametes that were stochastically produced by genetic drift, along with mutation and recombination (Masel, 2011).

1.2.1.5 Migration

Migration changes allele frequencies by exchanging alleles between populations. Differentiation between populations becomes prominent in the absence of gene flow because the genetic connectivity of populations is only achieved through migration. So, migration homogenises the allele frequencies between populations (Lowe et al., 2009). These continuing evolutionary processes shape a species by affecting the distribution of genetic variation between and within populations. Population genetic models investigate the connection between these evolutionary processes and the distribution of molecular genetic variants. Measuring genetic variation and applying population genetic models can help us to make inferences about the biology of organisms (Evanno et al., 2005).

1.2.2 Genetic markers

In order to study genetic variation, genetic markers, which show heritable allelic variations between individuals, are needed (Sunnucks, 2000). A number of genetic markers have proven useful for population genetics. These markers include proteins, such as allozymes, or DNA markers including restriction fragment length polymorphisms, amplified fragment length polymorphisms, microsatellites and DNA sequences (Silva and Russo, 2000).

Microsatellites and mitochondrial DNA (mtDNA) were chosen from various markers as the primary tool to measure genetic variation during this study. Mitochondrial genome is a small (13–20 Mb) circular molecule, which is generally maternally inherited. Hundreds of copies of mtDNA are found in most cell types. Recombination is rarely present in mtDNA, and the mutation rate is high in non-coding regions. These properties make it a good candidate for population genetic and phylogenetic studies. However, mtDNA markers alone are not enough to reveal the genetic structure of populations because all the genes in the mitochondrial genome evolve as a single linkage unit. Therefore, it is generally preferable

to use mtDNA markers along with nuclear DNA markers for additional information. Because organellar and nuclear genes exhibit different evolutionary patterns, they will reveal different characteristics of the population history (Zink and Barrowclough, 2008). Microsatellite markers are becoming an increasingly popular choice of marker as recent developments in DNA sequencing technologies enable rapid and cost-effective isolation of species-specific microsatellites (Guichoux et al., 2011). They are short sequences (two to six bp) of DNA that are repeated ten to hundreds of times in tandem. Microsatellite markers, which are found in high frequency throughout the genomes of most taxa, enable the analysis of several unlinked loci at the same time. Being shaped by different evolutionary forces, each locus tells an independent history of the population. Therefore, it is important to use a large number of loci for population genetics analysis. All these properties make microsatellites powerful markers for population genetics (Oliveira et al., 2006).

Another usage of mtDNA in phylogenetic analysis focuses on DNA sequences of specific genes, especially cytochrome c oxidase 1 (COI), as a DNA barcoding marker to distinguish cryptic or not very well known species (Hebert et al., 2003). Cryptic species are the species that are genetically divergent but classified as a single species as they are morphologically identical (Bickford et al., 2006). mtDNA markers will be useful in this study as *P. maculata* is a poorly studied species.

1.2.3 Measurement of genetic variation

Investigation of genetic structure and connectivity with genetic tools is based on spatial variation in allele frequencies (Hedgecock et al., 2007). Such an investigation requires the measurement of genetic diversity, i.e. the amount of genetic variation within biological entities, in the first place. Diversity can be measured hierarchically at the individual, population and species level using different methods specific to the nature of the markers. For example, the mean number of alleles or amount of heterozygosity across loci for microsatellites, or the number of haplotypes and base pair differences between pairs of

sequences for DNA sequencing data can be used to estimate genetic diversity (Lowe et al., 2009).

A fundamental equation of population genetics is the Hardy-Weinberg Equilibrium (HWE), on which several genetic diversity as well as differentiation models are based (Hedrick, 2011). The model assumes that frequencies of alleles and associatively frequencies of heterozygotes do not change over time in an ideal population. An ideal population meets these criteria: an infinitely large population size, no selection on any genotype, no migration, and random mating (Weinberg, 1908). The expected heterozygosity generally deviates from the observed values because natural populations do not meet the demands of ideal populations. Deviations from expected heterozygosity values can reveal population structure hierarchically at the individual, population and total population levels as stated by Sewall Wright's F-statistics (Wright, 1951; Wright, 1965; Wright, 1978). Deviations may point out a variety of evolutionary forces. For example, high heterozygosity may indicate balancing selection, which favours heterozygous genotypes, outbreeding, or gamete migration. An excess of homozygotes may indicate the presence of selection on the locus, inbreeding, zygote migration or subpopulation structure (Hedrick, 2011).

The extent of genetic connectivity between populations is determined by migration (gene flow). When there is no migration between populations, alternate alleles proceed to fixation, which creates structured populations. Therefore, the rate of gene flow can be indirectly estimated using F-statistics, which reveal levels of population structure (Slatkin, 1987). Estimation of the migration rate may serve to predict the future distribution of species, especially in studies on endangered or invasive species.

1.3 COASTAL MARINE CONNECTIVITY IN NZ

Population genetic tools can provide information to reveal connectivity and gene flow between *P. maculata* populations, especially if the genetic information is evaluated in conjunction with other disciplines like oceanography, geography and ecology, to help us understand the factors that shape the genetic structure of this species (Selkoe and Toonen,

2011). Studying the population structure and the genetic discontinuities observed for other invertebrates that inhabit NZ's coasts, and factors that shape their distribution will improve our understanding of the genetic distribution of *P. maculata*. This section of the thesis focuses on the marine biogeography of NZ and reviews population genetic studies that have been conducted for other marine invertebrates inhabiting NZ's coasts.

Marine environments seem to be open habitats at first glance, but marine animals still encounter many biogeographical barriers that block gene flow between populations of species. Many marine organisms exhibit subpopulations even over short geographical distances (Gardner et al., 2010). The population structure and the evolutionary history of the species are shaped by biological traits and environmental factors that interact in a complex manner (Ross et al., 2009). NZ presents an interesting case study where we can observe the effect of different factors on its marine species' distribution and population structure. Geographically, NZ includes 700 islands extending from the subtropical Kermadec Islands (29°S) to sub Antarctic Campbell (52°S) (West and Thompson, 2013), and its coastline is estimated to cover about 15,000 km (Wallis and Trewick, 2009). Subtropical and sub Antarctic waters converge and interact with continents creating a complex oceanography with several currents and eddies. The coastal currents are also affected by topography and temporal winds, tides and waves. NZ's complex oceanographic characteristics create unpredictable effects on the genetic structure of coastal organisms (Ross et al., 2009). Phylogenetic and population genetic studies of NZ marine biota that utilised different types of genetic markers (including microsatellite markers, allozymes, mtDNA sequences and RAPD) have shown various patterns of genetic structure for different species (Gardner et al., 2010). Different physical and biological factors may influence the population connectivity in marine environments. These factors include the specific geography encountered by the marine organism in question, habitat requirements, the presence of competitors or invasive species, as well as the dispersal ability of the organism. These factors will be discussed in more detail in the following subsections.

1.3.1 Isolation by distance

Physical distance within or between populations, which is called isolation by distance (IBD), can decrease gene flow (Slatkin, 1993). According to the IBD model, there is a linear relationship between pairwise genetic and spatial distances between populations (Rousset, 1997). The IBD model or latitudinal gradient (which is a form of IBD) is observed in many NZ marine species from a variety of taxa (Gardner et al., 2010).

1.3.2 Hydrodynamics

Hydrodynamic conditions increase or decrease genetic connectivity of populations by affecting dispersal capacity of the species, and create population structures that cannot be simply explained by IBD (Riginos et al., 2011). For example, ocean currents play a role in both the extent and direction of dispersal; large-scale currents may increase gene flow between distant populations, but they can also act as a barrier in instances like upwelling areas or eddies (Muhlin et al., 2008). Figure 1.4 shows the coastal oceanic currents that are effective in NZ. Besides IBD, population genetic studies have shown significant genetic variations that are associated with a number of such phylogeographical boundaries specific to NZ. One of the most notable examples is differentiation between the northern and southern populations for a variety of taxa, such as Gastropoda, Bivalvia, Ophiuroidea, Asteroidea and Magnoliophyta (Gardner et al., 2010; Ross et al., 2009). Cook Strait, a body of water that separates the coast of the North and South Islands, and/or coastal upwelling regimes (the transport of deeper water to shallower depths) at 42°S are thought to be the source of this pronounced NZ-wide northern-southern disjunction (Gardner et al., 2010).

Another geographical barrier to gene flow was identified between the west and east coasts of the North and/or South Islands. The west-east differentiation can co-occur with the north-south differentiation (Gardner et al., 2010; Keeney et al., 2013; Ross et al., 2009). It has been postulated that a northerly offshore flow during summer (Brodie, 1960; Stanton, 1973) and the upwelling at the northern tip of the North Island (Ridgway, 1980; Roberts and Paul, 1978; Stanton, 1973) (Figure 1.4) are the sources of the discontinuity. Not only contemporary oceanic conditions but also historical events such as fluctuations in sea level

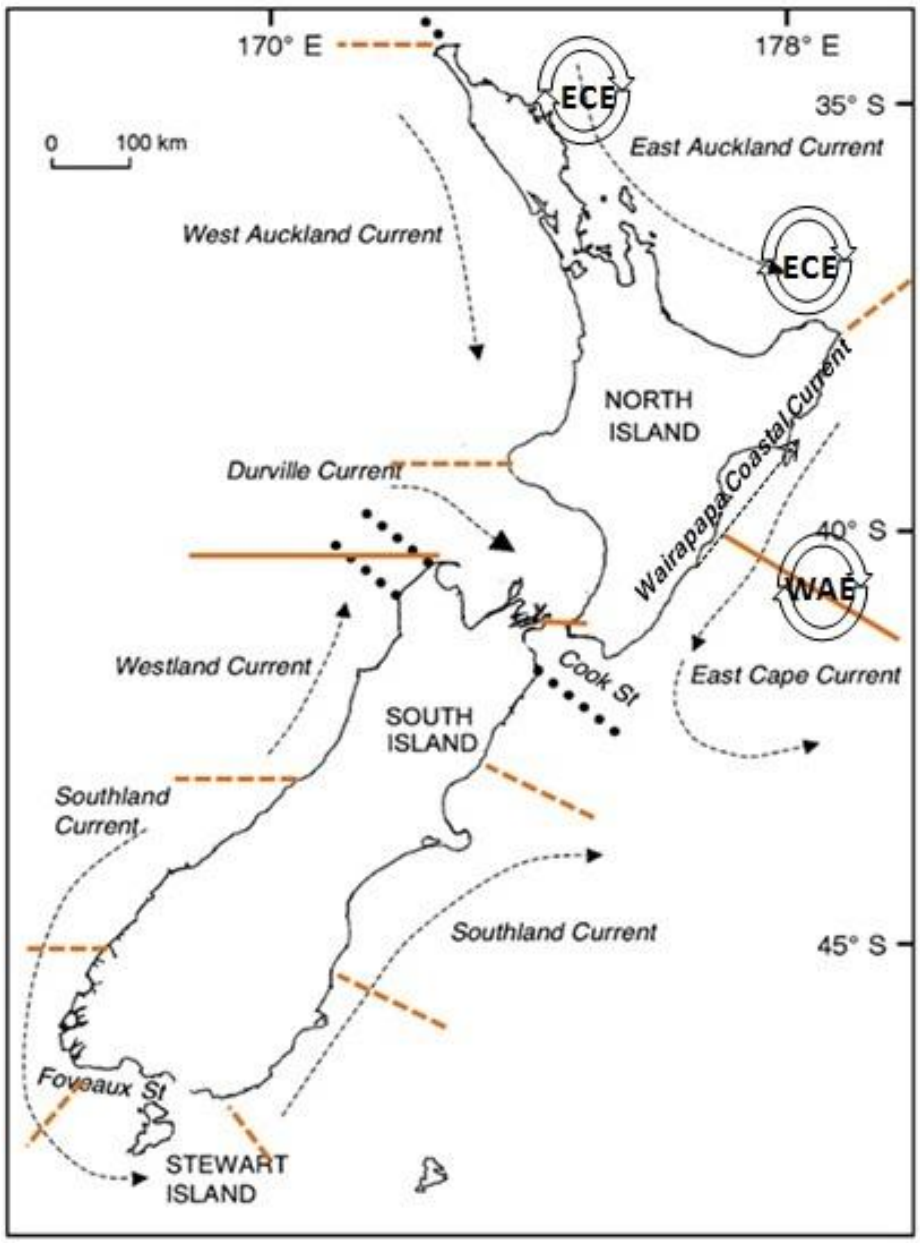


Figure 1.4 Map of NZ with major regional surface currents and biogeographic regions. The black dotted lines indicate upwelling regions in the northern South Island and Cape Reinga; the orange dashed lines represent the boundaries between the bioregions and the orange solid lines represent the boundaries between the bio-provinces that were defined by Shears et al. (2008). The figure was adapted from Ayers and Waters (2005) and redrawn based on Shears et al. (2008). ECE: East Cape Eddies, WAE: Wairapapa Eddy.

and temperature, and plate tectonics influence the biogeographic boundaries (Jacobs et al., 2011). The North and South Islands were connected during the Pleistocene glacial cycles

until the recent rise in sea levels created the present coastline (12–6 kya, Lewis et al., 1994). The historical connection between the land could be the reason for the observed discontinuity between the east and west populations (Keeney et al., 2013).

1.3.3 Panmixia or localised structures

Biogeographical barriers do not always have same the effects on the species inhabiting the same geography. For example, some NZ rock lobster, bivalve and echinoderm species do not exhibit any genetic structure that indicates high gene flow levels between populations of these species. On the contrary, highly structured populations throughout NZ have also been observed. Biogeographic breaks affecting the distribution of biodiversity (species) may also affect the genetic distribution of species. In this case, biogeographic barriers coincide with phylogeographic barriers affecting the genetic distribution of a particular species, creating structured populations (Ross et al., 2009). An interesting example is that the seven identified subpopulations of the Waratah anemone *A. tenebrosa* coincide with the bioregions (Figure 1.4) that were suggested based on the distribution of benthic invertebrate, microalgae and reef fish species (Ross et al., 2009; Shears et al., 2008; Veale, 2007).

1.3.4 Habitat requirement

An important biological trait affecting the distribution of a species is the habitat requirement (Keeney et al., 2013). Species requiring specific habitats, such as sheltered shores or an exposed rocky coast, show a higher level of genetic structure compared to species with reduced habitat specificity because such habitats do not exhibit continuous distribution (Ayre et al., 2009). Environmental factors such as the depth, salinity, and temperature of the water may vary from one region to another, thus shaping the distribution of the species depending on its sensitivity to such environmental factors (Riginos et al., 2011). Additionally, varying conditions could apply different selective forces on the populations resulting in differentiation of the populations. For example, Wei et al. (2013b) studied the population genetic structure of NZ green shell mussels around NZ using microsatellite markers, which revealed that varying sea surface temperatures exerted

varying selective forces that may explain the genetic variation observed within this species in NZ.

1.3.5 Developmental stage

Many marine organisms, including gastropods, show either direct development in which eggs hatch into crawl-away juveniles or they have planktonic larval development (Weersing and Toonen, 2009). As the adult stage is generally not highly mobile, dispersal is mainly achieved during the pelagic larval phase. Therefore, species with planktonic larval stages are expected to show less genetic structure compared to species without larval stages (Weersing and Toonen, 2009). Additionally, the type of larva, which determines the pelagic larval duration (PLD), may also define the extent of the dispersal ability of the larva and level of mixing between distant populations (Silva and Russo, 2000). There are two distinct larval types in invertebrates. Planktotrophic larvae that hatch from small eggs develop slowly because they must feed to complete the larval stage. Lecithotrophic larvae hatching from large eggs can complete the stage faster (McEdward, 1997). Despite some conflicting findings (Weersing and Toonen, 2009), there is a positive relationship between the PLD and the strength of the genetic connectivity between populations (Selkoe and Toonen, 2011). When the PLD is short, other biological or environmental factors have a stronger effect on the genetic structure of the organism compared to species with long PLDs (Ross et al., 2009). It is worth noting that the effect of the PLD is attributed more to passive movements due to tidal and wind movements than the active movement of the larvae (Gardner et al., 2010). *P. maculata* lays multiple small eggs (100 μm) encapsulated in a gelatinous cylindrical mass (Figure 1.1B). The eggs hatch into larvae within ten days (Gibson, 2003; Wood et al., 2012b). The planktotrophic larvae feed on plankton for approximately 3 weeks before settling on bio-filmed surfaces as juveniles (Gibson, 2003; Wood et al., 2012b). This relatively long PLD can increase genetic homogeneity among *P. maculata* populations.

1.3.6 Patterns of sexual reproduction

Several patterns of sexual reproduction are observed in invertebrates, and the pattern of sexual reproduction determines the structure of the populations to some extent. These

patterns are biparental reproduction, parthenogenesis and hermaphroditism. Hermaphroditism is the focus in this study because it is the reproduction pattern of *P. maculata* (Willan, 1983). Hermaphroditic organisms have both male and female reproductive organs during their life span. Hermaphroditic organisms can reproduce by autogamy or self-fertilisation, which will lead to structured populations (Charlesworth and Wright, 2001). Conversely, outcrossing simultaneous hermaphroditic invertebrates have a large effective population size and show high levels of genetic variation (Silva and Russo, 2000). *P. maculata* is a simultaneous hermaphrodite in which reciprocal copulation is often observed (Willan, 1983). In this case, high genetic variation can be expected for this species.

To summarise, this chapter has identified several biogeographic barriers that affect genetic connectivity in populations of NZ coastal organisms. The effect of these barriers can vary from species to species based on the life history and evolutionary responses of the respective species to various geological changes and other environmental factors (Ross et al., 2009). Therefore, the potential divergent effects of these barriers should be taken into consideration in a species-specific manner for studies that address the issue of population structure for any NZ marine species, including *P. maculata.*

1.4 THESIS SUMMARY

The origins and evolution of toxic *P. maculata* individuals have not been clarified. Identification of distinct geographical regions that contain either toxic or non-toxic individuals raises certain questions. For example, has a toxic population recently invaded NZ's coasts? Are individuals from toxic and non-toxic regions genetically isolated, or are they even cryptic species? What is the degree of genetic connectivity between distinct populations, if there is any? Population genetic tools can address these questions when combined with the current knowledge of phylogeographic barriers that are effective in several NZ marine species. It will also help to understand the underlying evolutionary forces that shape the population structure of *P. maculata*. My PhD project had three overall aims: 1) to understand the spatial distribution of genetic diversity across *P. maculata*

populations from various regions in NZ utilising microsatellite and mitochondrial data; 2) to understand the genetic connectivity of different *P. maculata* populations, and determine the biogeographical factors shaping its population structure; 3) to determine whether there is a correlation between variability in TTX concentrations and genetic structure, or not. With these goals in mind, I developed microsatellite markers, which was a significant first step with this intractable organism. The development and validation of these markers is described in **Chapter 2**. **Chapter 3** deals with the materials and methods used in this study. I used microsatellite markers that I developed to investigate the diversity and genetic structure of the *P. maculata* populations around NZ, the results of which are presented in **Chapter 4**. **Chapter 5** presents the analysis of the population structure of *P. maculata* again, but this time utilising sequences from mitochondrial DNA genes, including COI and Cytb, and discusses the results found from both marker types (microsatellite and mitochondrial). The species-specific markers employed for this step were also developed. The final **Chapter (6)** contains concluding remarks and future perspectives. The toxicity assay was performed by our collaborators (Susanna Wood and Paul McNabb) at Cawthron Institute (Nelson, New Zealand) for not all, but only a proportion of the samples. However, the general trend of toxicity levels at each locality was known based on previous studies ((McNabb et al., 2014; Salvitti et al., 2015a; Wood et al., 2012b) or personal communications (Susanna Wood). Samples from the northern North Island (Ti Point, Coromandel, Auckland and Tauranga) were considered as toxic whereas relatively southern regions (Wellington and Nelson) were considered as slightly toxic or non-toxic. One interpretation of these data suggests that there is an association between toxicity and population structure, however this is a correlative association and does not constitute proof. Nonetheless, the correlation raises important hypothesis that can be tested directed in future studies.

CHAPTER 2: DEVELOPMENT OF THE MICROSATELLITE MARKERS

This chapter is a modified version of the article published in Biochemical Systematics and Ecology; Yildirim et al., 2014; DOI: 10.1016/j.bse.2014.04.001, © 2014, Elsevier. The marker names have been changed to be compatible with the rest of the thesis, and each section of the original article has been extended. The original publication is attached to Appendix One.

2.1 ABSTRACT

Using 454 pyrosequencing data, 24 polymorphic microsatellite markers were identified for the grey side-gilled sea slug, *Pleurobranchaea maculata.* The grey side-gilled sea slug is found throughout the western and south Pacific and is known to contain high concentrations of tetrodotoxin. Polymorphism was assessed in 20 individuals obtained from geographically distinct locations within New Zealand. Between 2 and 15 alleles were identified at each locus. The observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0.10 – 1 and 0.10 – 0.94, respectively. No significant linkage disequilibrium between pairs of loci or deviations from the Hardy–Weinberg proportions was observed. The markers are central to understanding the population biology and genetic structure of *P. maculata*.

2.2 INTRODUCTION

Microsatellites, or simple sequence repeats (SSRs), are tandem repeats of the short DNA sequence motifs that are one to six base pair in length (Chambers and MacAvoy, 2000). These co-dominantly inherited markers are abundant and distributed almost randomly throughout the genome of most taxa, which enable the use of many loci at the same time (Oliveira et al., 2006). As microsatellites are generally located within non-coding regions,

they are selectively neutral (Jarne and Lagoda, 1996). Neutral loci tend to be polymorphic because they are free to accumulate mutations (Oliveira et al., 2006). It has been proposed that some mutational mechanisms including DNA polymerase slippage during replication (Ellegren, 2004), errors during recombination-dependent repair and unequal crossing-over contribute to high mutation rates and high polymorphism in microsatellites (Bhargava and Fuentes, 2010). All these properties make microsatellites powerful markers for genetic analysis, including genome mapping, paternity testing, and conservation or population genetics (Oliveira et al., 2006).

As *P. maculata* is a non-model organism whose genome sequence is unknown, no nuclear molecular markers including microsatellites were available to study the population structure of this species. De novo development of microsatellite markers in non-model organisms with traditional techniques requires the construction of a genomic library enriched for repeated motifs using genetically modified bacteria, and hybridisation to identify microsatellite positive clones, plasmid extraction and Sanger Sequencing. Most of these steps are laborious, time-consuming and costly (Abdelkrim et al., 2009), which used to hinder the utilisation of microsatellite markers for non-model organisms. Recent developments in DNA sequencing technologies now make it possible to obtain large numbers of sequences more rapidly and cost-effectively compared to Sanger sequencing methods, where the cloning, clone scanning and plasmid extraction steps of the traditional methods are skipped. This enabled fast and cost-effective identification of microsatellites from genomic DNA sequences for non-model species (Csencsics et al., 2010). Therefore, population genetic studies using microsatellites have become popular over the last few years.

Next generation sequencing techniques require the fragmentation of the whole genome, ligation of each fragment to an adapter oligonucleotide and capture of each fragment on a single bead. Then the fragments are enriched by an emulsion PCR, which is capable of amplifying a single single-stranded DNA fragment and producing millions of copies of each fragment. The beads, each of which contains the clone of a single DNA fragment, are

subjected to massively parallel phosphate sequencing (Margulies et al., 2005). After obtaining the low-coverage whole genome sequences of the organism from these randomly amplified shotgun sequences, it is possible to explore microsatellite regions and to design primers to amplify the putative microsatellites using bioinformatics tools (Abdelkrim et al., 2009). It is also necessary to be able to amplify candidate loci by PCR and to establish a suitable genotyping method in order to distinguish allelic variants. Finally, microsatellite markers that exhibit polymorphism should be identified. This is possible by investigating multiple individuals (Abdelkrim et al., 2009). We followed this pipeline to develop microsatellite markers for *P. maculata*; this chapter describes this process and reports on the markers' identity and associated properties.

2.3 MATERIALS AND METHODS

2.3.1 Sampling, DNA extraction and genome sequencing

The genome of *P. maculata* was partially sequenced using a single sea slug collected from Narrow Neck, Auckland, New Zealand in June 2010. The slug was placed at -80 ºC upon collection. DNA extraction included the following steps: grinding of tissue with liquid nitrogen grinding, lysis with an acetyltrimethylammonium bromide buffer followed by phenol/chloroform extraction (Winnepenninckx et al., 1993). The resulting aqueous phase was put through a Qiagen DNeasy Blood & Tissue Kit.

Preparation of the DNA library, emulsion-based clonal amplification of the library and sequencing were performed on Roche 454 GS Junior instrumentation using Titanium technology (Roche, NJ, USA) according to the manufacturer's protocols. The utility of microsatellite markers was determined by assessing the levels of polymorphism in a sample of twenty slugs collected from Ti-Point Wharf (Leigh, Auckland) and Tasman Bay (Nelson) in August and November 2012, respectively. Slug DNA was extracted using the OMEGA Biotek E.Z.N.A.^{TM} Mollusc DNA Kit after grinding the freeze-dried samples with a mortar and pestle.

2.3.2 Isolation and development of the primers

The bioinformatics software QDD (Meglecz et al., 2010) was used to analyse the 454 sequences to search for microsatellite motifs, and to design primers. Sequences with significant similarity were discarded first (default settings) in order to avoid redundant sequences, and consensus sequences were generated. Sequences >80 bp were screened for perfect and compound microsatellites with a minimum of four repeats of di-hexanucleotide motifs. Primer 3 software, built into QDD, was used to design primers using default settings. 90–350 bp of amplicon size was selected. The 5' ends of the forward primers of candidate loci were tagged with a 50 -fluorescent dye (6-FAM, VIC, NED and PET) or with a universal M13(-21) sequence (18bp) (Schuelke, 2000). For the latter loci, a third primer with the M13(-21)-sequence and a 50 -fluorescent dye (6-FAM, HEX, NED and PET) was used to screen the primer pairs. PCRs for each locus were carried out in a total volume of 12.5 mL, including 1X PCR buffer (Invitrogen Kit), 2 mM MgCl2, 0.2 mM M13 primer and reverse primers, 0.05 mM forward primer, 0.2mM dNTP, 0.25 U/mL Taq polymerase and 10–25 ng genomic DNA. Either standard or touchdown PCR regimes were applied. Standard conditions were as follows: 94 ºC (3 min), then 30 cycles at 94 ºC (30 s)/Tm (30 s)/72 ºC (30-60 s). The touchdown conditions were as follows: 94 ºC (3 min), then 20 cycles at 94 ºC (30 s)/63 ºC (decreasing by 0.5 ºC/cycle, 30 s)/72 ºC (20–60 s) followed by 20 cycles at 94 ºC (30 s)/53 ºC (30 s)/72 ºC (20–60 s). Both regimes were followed by 8 cycles at 94 ºC (30 s)/53 ºC (30 s)/72 ºC (20–60 s) if the M13(-21) sequence is used for fluorescent labelling. A final extension at 72 ºC was applied for all reactions for 30 minutes. The PCR regime, annealing temperature, and elongation time were estimated according to the primer pair properties, and the aforementioned are presented in Table 2.1. The multiplexed amplicons were electrophoresed on a commercial ABI 3730 Genetic Analyzer (Applied Biosystems, USA) along with a size standard GeneScanTM-500 LIZTM (Applied Biosystems). The resulting genotyping data were analysed with the Geneious Microsatellite Plugin 1.2 on Geneious Pro 5.5.6.

Table 2.1 PCR annealing temperature and elongation time for each primer pair.

Locus ID	Annealing temperature (˚C)	Elongation time (sec)
[a]1, [a]15, [a]26, [a]27, 38, 51, 52, [a]61, 65	57	45
[a]2, 7, 12, [a]41, 42, 43, 46, [a]49, [a]67, [a]89	63 to 53[b]	45
[a]48, [a]22,	63 to 53[b]	30
[a]32	49	60
4, 11	55	45

[a]forward primers directly tagged with a fluorescent dye, [b]touchdown regime

2.3.3 Data analysis

The genotyping data were tested for genotyping errors, null alleles, large allele dropout and departure from the Hardy-Weinberg equilibrium (HWE) using MICRO-CHECKER v.2.2.3 (Van Oosterhout et al., 2004). GenoDive v2.0b24 was used to calculate allelic richness (Meirmans and Van Tienderen, 2004). The observed heterozygosity (H_o), expected heterozygosity (H_e) and linkage disequilibrium (LD) for each pair of loci were estimated using ARLEQUIN v3.5 (Excoffier et al., 2005). The number of permutations and initial conditions for the EM algorithm were set to 20,000 and 5, respectively. An R package (Team, 2013) was used to adjust the LD results for multiple simultaneous comparisons based on Bonferroni correction.

2.4 RESULTS AND DISCUSSION

The 454 sequencing runs yielded 115,579 reads, giving 27.456 Mbp of sequence data. (NCBI SRA accession number SRR1168412).The length of the sequences varied between 40 to 1030 bp with an average length of 241 bp. 30,130 (23%) of these sequences contained candidate microsatellite loci with ≥ 4 di to hexa repeats. The size of the loci ranged from 80 to 866 bp in length. The total amount of detected SSR motifs was 70,235. The majority of the motifs were dinucleotides (45,572; 64.9%), followed by tri- (19,659; 28%), tetra- (4,447; 6.3%), hexa (320; 0.5%) and pentanucleotides (237; 0.3%). Among these sequences, Primer 3 created putative primer sets for 639 and 201 loci, having perfect and interrupted/compound repeats, respectively, which yielded a minimum of 90 bp of PCR product. Generally, several primer pairs were suggested for each locus. Among them,

primer pairs with the "best" score, which corresponds to pairs with the lowest penalty, were considered as candidate pairs. Among the designed primers, 65 loci (8 di-, 34 tri-, 16 tetra-, 1 penta- and 6 hexanucleotide repeats) were selected for polymorphism analysis, and a numerical code was given for each primer pair (For the list of primers, see Supplementary Table 1 in Appendix Two). Generally, dinucleotides are the most frequently used repeats due to their high polymorphism level. However, PCR amplification of dinucleotide repeats often gives rise to stutter bands due to DNA polymerase slippage during PCR reaction. This activity may result in misidentification of the actual alleles (Guichoux et al., 2011). Therefore, in this study, tri- and tetranucleotides were the preferred repeats. This is because they are more likely to be polymorphic compared to penta- and hexanucleotides. Additionally, the numbers of detected penta- and hexanucleotides were low. Primer pairs that produced an amplicon size smaller than 100 bp were generally avoided because of potential genotyping problems. During genotyping, products of primer dimerisation can result in a strong false-positive signal, which can interfere with genotyping data of small amplicons.

After the amplification test, thirty of the 65 loci yielded successful amplicons of the expected size while the remaining loci yielded no products, non-specific products or smear products. These thirty loci were examined further for consistency of amplification and polymorphism with a panel of 20 slugs from two different geographic regions (Leigh and Nelson, New Zealand). Six loci did not amplify consistently. Genotyping of the remaining 24 loci showed that two loci had limited polymorphism (two alleles), but 22 loci yielded polymorphism in the range of 3 to 15 alleles (Table 2.2). The average number of alleles per locus was 5.13. The H_o ranged from 0.10 – 1, and the H_e ranged from 0.10 – 0.94 (Table 2).

MICROCHECKER analysis showed no evidence of large allele drop-out or scoring error due to stuttering; however, manual analysis of the band patterns revealed evidence of stuttering at Loci 67 and 90. Deviations from HWE were observed at Loci 15, 67 and 90 for both populations. These are most likely a consequence of chance effects due to the small

sample size. No significant linkage disequilibrium was found after Bonferroni correction ($P<0.05$).

The results are evidence of a robust set of microsatellite markers suitable for bio-geographic and population structure analysis of *P. maculata*.

Table 2.2 Characteristics of 24 polymorphic microsatellite loci in *P. maculata* isolated using 454 high-throughput sequencing.

LocusID& GenBank No	Primer Sequence (5' to 3')	Dye	Repeat motif	*Na	Allele size (bp)	Ti Point (n:10)		Nelson (n:10)	
						Ho	He	Ho	He
1 KJ510176	F: TGGTGATCACGACACAAATG R: CGCAAAATAAAATAGCAAGCC	NED	$(ATTC)_{15}$	15	108–208	0.70	0.77	0.70	0.93
2 KJ510177	F: TGTAGGAACAAATAGGCAGTGAGA R: TTCGAAGAGGAATGGAATGG	PET	$(CAGA)_{12}$	7	105–133	0.70	0.79	0.60	0.74
4 KJ510178	F: ACCACTTTGCAGGTCAAACC R: GTTCACACTGCATCATTGCC	PET	$(GCGTGT)_{7}$	5	140–170	0.30	0.52	0.50	0.67
7 KJ510179	F: AACGATCTAATCGGTGGCTG R: GAAATTTGTACCTCGCGCTC	PET	$(CTGC)_{7}$	3	246–254	0.40	0.57	0.40	0.54
11 KJ510180	F: AAAGCAATGATAACAGCGCC R: CGGGTGAAAGTTGACAGAC	FAM	$(CAGG)_{6}$	3	119–127	0.20	0.19	0.10	0.10
15 KJ510182	F: TTATCTCTGTTTCCCCACCG R: TGAGATAGTGTGCCATGTCCTT	PET	$(GCAG)_{6}$	5	144–160	0.80	0.66	0.70	0.56
20 KJ510183	F: ACCGACATGCAAACAGACAG R: TTGTCTCCAAGTGGCCAAG	NED	$(CAGG)_{7}$	5	132–148	0.50	0.73	0.80	0.72
26 KJ510184	F: GGGACAGACAAATGGCTTGA R: GAAAAGAAAGAATGAAAGAAAGAATGA	VIC	$(CTTT)_{11}$	5	140–164	0.60	0.70	0.80	0.67
27 KJ510185	F: GGCATAGGATGAAAAGGCAA R: CAGGCTCCAGATCGGTAAAG	FAM	$(ATG)_{11}$	9	103–136	0.90	0.78	0.80	0.77
32 KJ510186	F: GGAATTGGGCGCTATACAAT R: AGCCGCTTGGTACTCAAAGA	PET	$(TTA)_{12}$	5	110–122	0.70	0.70	0.60	0.66
38 KJ510187	F: CAAAGACCCGATATTCACGG R: GAAACAATAGCAACACAACGG	VIC	$(GTT)_{9}$	8	160–181	1.00	0.89	0.60	0.82
41 KJ510188	F: TCAAGGTTAGTTTCACCATCACC R: TAGTGATAGTGGCAGTGGCG	PET	$(CAT)_{9}$	3	122–128	0.70	0.62	0.60	0.66
42 KJ510189	F: TCAACTGGGAACAAACATCG R: ATGTCATTGGTTGACGTTGG	FAM	$(GAT)_{10}$	4	121–133	0.50	0.54	0.50	0.49
43 KJ510190	F: TGCTACCACTGCTACTATGCTATTG R: CATCCACGATCCACAGTCAA	HEX	$(CTGCTA)_{10}$	4	164–194	0.50	0.54	0.60	0.64
46	F: AGGTCGAACCCTTAACCTCG	PET	$(CAC)_{8}$	4	241–250	0.40	0.53	0.90	0.63

KJ510191	R: AATGGTTTTAGCCTTGGCCT								
49	F: GCAGCCAACAAGGAAGGATA								
KJ510192	R: GGTAACCATGGATTTGACGC								
51	F: ACTGAATGGGGTCGAGTACG								
KJ510193	R: TTTGTGCTAAAACCTTCCCG	VIC	$(GTTT)_5$	4	167–179	0.50	0.57	0.30	0.27
52	F: GCTCACTCATCGCATTCATC								
KJ510194	R: ATGTGCCTTTGTGCTGTCAC	NED	$(CTA)_8$	2	175–178	NA	NA	0.10	0.10
61	F: TGTTCTTGATTCGTTATGGGG								
KJ510195	R: TGTCCGATGGTGAGCAGTAA	NED	$(GTT)_8$	5	169–181	0.20	0.19	0.80	0.59
65	F: AGTCGATATTATGGCGTCGG								
KJ510196	R: TTGTTTAAACTTGTTTGAGAAAAGG	FAM	$(CTT)_9(CTC)_6$	2	120–126	0.60	0.87	0.10	0.10
66	F: TTTAACTGGCTGTAGGCCAAA								
KJ510197	R: CGAAGAGAAAGCCGGTAGTG	FAM	$(CTT)_9(CTC)_6$	4	198–207	0.60	0.63	0.60	0.50
67	F: CTGTCTTGGCCATAGGCTCT								
KJ510198	R: GGCATAGGATGAAAAGGCAA	VIC	$(ATC)_{18}$	8	101–122	0.40	0.76	0.60	0.67
89	F: CAATAACAAAGGCCTGGGAA								
KJ510199	R: GGCATGTCGATTGCACTAAA	NED	$(AG)_{14}$	7	156–170	0.70	0.77	0.60	0.63
90	F: TGGAAGGTTGTGTATCAAGCA								
KJ510200	R: ATCCATCAAATGGTGGCAAT	PET	$(AC)_{14}$	6	179–191	0.70	0.79	0.60	0.74

Na: total number of alleles, Ta: annealing temperature (63 to 53 means touchdown regime), n: total number of samples, NA: not available due to only one allele, * data from the individual used for 454 sequencing was included in the calculations.

CHAPTER 3: MATERIALS AND METHODS

The previous chapter, Chapter 2, described the development of microsatellite markers for *Pleurobranchaea maculata*. A subset of the developed microsatellite markers and DNA sequences from mtDNA genes, the cytochrome b (Cytb) and cytochrome c oxidase subunit I (COI) were used to study the population genetics of *P. maculata.* This chapter, Chapter 3, outlines the materials and methods used to study the population structure of *P. maculata*.

3.1 Sampling, DNA extraction

A total of 157 samples were collected from Whangarei (WH), Ti-Point (TP), Auckland (AKL), Coromandel (CR), Tauranga (TR), Wellington (WL), Picton (PC), Nelson (NL) and Kaikoura (KK) in New Zealand from subtidal regions by the divers by hand. The single specimen from KK was collected from 290 m depth using a fish trap (Wood et al., 2012a). The sampling was performed between winter 2009 to October 2013. All samples were placed at -80 ºC upon arrival in the laboratory. The sampling locations are shown in Table 3.1 and on the map in Figure 3.1.

Table 3.1 Sampling locations and abbreviations for the NZ *P. maculata* populations. The number of individuals sampled per population for each marker type was indicated.

# of samples (mtDNA/microsats)	Location	Longitude	Latitude	Sampling date
5 /1	Urquharts Bay, Whangarei (WH)	-35.842675	174.527764	Aug 2010
20 /20	Whangateau Harbour, Ti Point (TP)	-36.3175	174.784444	Aug 2012
3 /1	Whangapoua Harbour, Coromandel (CR)	-36.741281	175.627905	Sep 2012
19 /19	Tamaki Strait, Auckland (AKL)	-36.781333	174.824473	Jul 2011/Oct 2012
1 /1	Tamaki Strait, Auckland (AKL)	-36.81104	174.800399	Winter 2009
7 /7	Waitemata Harbour, Auckland (AKL)	-36.82888	174.76458	Oct 2012
4 /4	Waitemata Harbour, Auckland (AKL)	-36.853077	174.712449	Oct 2012
20 /20	Tauranga Harbour, Tauranga (TR)	-37.634465	176.176455	Jun 2011/Jun 2012
6 /6	Tauranga Harbour, Tauranga (TR)	-37.677623	176.1766	Jun 2013
7 /7	Tauranga Harbour, Tauranga (TR)	-37.643948	176.148448	Oct 2013
18 /18	Lambton Harbour, Wellington (WL)	-41.288433	174.779983	Oct 2012
45 /45	Tasman Bay, Nelson (NL)	-41.058056	173.091111	Nov 2010-Aug 2013
1 /1	Picton Harbour, Picton (PC)	-41.272714	174.013433	Dec 2012
1 /0	Kaikoura (KK)	-42.465567	173.781517	Nov 2011

DNA was extracted from tissue using an E.Z.N.A.™ Mollusc DNA Kit (Omega Bio-Tek, Inc., USA) after grinding the freeze-dried samples with a mortar and pestle or grinding frozen samples in Eppendorf tubes using a sterile plastic or metal stick.

Alternatively, DNA extraction included the following steps: grinding of tissue with liquid nitrogen, lysis with an acetyltrimethylammonium bromide buffer followed by phenol/chloroform extraction (Winnepenninckx et al., 1993). DNA from the resulting aqueous phase was further purified using a Qiagen DNeasy Blood & Tissue Kit (Qiagen, USA).

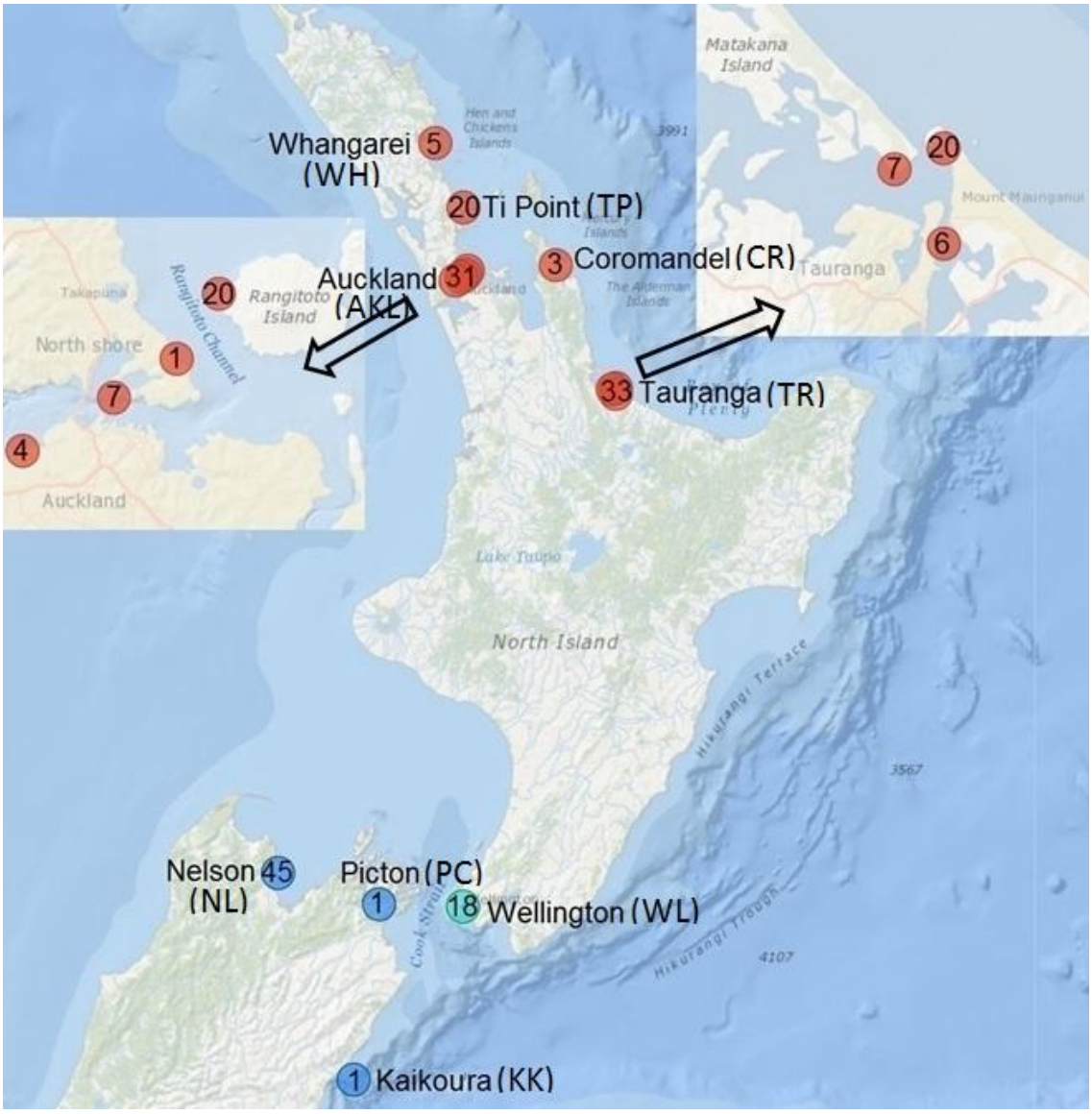


Figure 3.1 Sampling locations for the *P. maculata* individuals. The circles show the sampling locations, and the numbers within indicate the sampling size. The arrows show magnified maps of Auckland and Tauranga. Red: *P. maculata* individuals with high TTX concentrations were observed; cyan: location contained individuals with low TTX concentrations; blue: non-toxic samples with trace amounts of TTX were observed.

3.2 Tetrodotoxin assay

I had limited knowledge of the toxicity levels in *P. maculata* individuals from the Wellington region as only one individual had been tested so far (Wood et al., 2012b). To obtain a general idea about this region, eight individuals collected from WL in October 2012 were tested for their TTX concentrations. A subsample (50–70 mg) from six of the eight individuals was homogenised in Eppendorf tubes using an electric drill with a plastic drill bit. 0.1% acetic acid that was prepared with Millipore water was added so that the total volume became 10 times the sample weight. The homogenised samples were centrifuged at 3000 × g for 10 minutes. Then, 100 µl of the supernatant was mixed with freshly prepared 900 µL 0.1% acetic acid in methanol. The mixture was kept at -20°C for a minimum of one hour and then centrifuged at 16,000 rpm for 10 minutes. The supernatant was diluted four times with 0.1% acetic acid in methanol. The extraction method was adopted from Khor et al. (2014). Serena Khor (The University of Waikato) extracted TTX from the remaining two individuals using the same methodology (Khor et al., 2014). The TTX assay was performed by Paul McNabb at the Cawthron Institute (Nelson) using a liquid chromatography-mass spectrophotometry method that is described in McNabb et al. (2010).

3.3 Genotyping

3.3.1 Microsatellite markers

Fifteen of the microsatellite markers (loc *1, 2, 15, 20, 26, 27, 32, 38, 42, 51, 61, 65, 97, 89* and *90*), whose development process was described in Chapter 2 (Yildirim et al., 2014), were genotyped for 149 samples. The microsatellite markers from four samples from Whangarei, two from Coromandel, one from Nelson and one from Kaikoura could not be amplified due to DNA degradation.

The 5’ ends of the forward primers of fifteen loci were tagged with a 5’-fluorescent dye (6-FAM, VIC, NED and PET) or with a universal M13(-21) sequence (18bp) (Schuelke, 2000). For the loci tagged with a universal M13(-21) sequence, a third primer with the M13(-21)-sequence and a 5’-fluorescent dye (6-FAM, and NED) was used to screen the primer pairs. The primers for the former loci were pooled into three multiplex sets; polymerase chain reaction (PCR) was used to amplify the DNA for each individual. For directly labelled primer pairs, PCRs were carried with 2X Qiagen Multiplex PCR

Master Mix (containing 3 mM $MgCl_2$, the rest of the kit components are not disclosed by Qiagen Inc.), 5X Qiagen Q-buffer, RNase-free water, primer mix (loc *27* and *32* at 0.025 μM, loc *65* at 0.05 μM, the rest of the primers at 0.1 μM), and 10–25 ng genomic DNA. When a third primer with an M13(-21) sequence was used, the PCR was carried out with a 1X PCR buffer (10X Invitrogen™ *Taq* DNA Polymerase buffer contains 200 mM Tris-HCl pH 8.4, 500 mM KCl), 2 mM $MgCl_2$, 0.4 μM M13 primer and reverse primers, 0.1 μM forward primer, 0.2 mM dNTP, 0.02 U/mL Invitrogen™ *Taq* DNA Polymerase and 10–25 ng genomic DNA. Either standard or touch-down PCR regimes were applied (Table 3.2). Standard conditions were as follows: 94°C for 3 minutes when an M13(-21) sequence was used or 95°C for 15 minutes for the directly labelled primers, then 30 cycles at 94°C (30 s)/ 57°C (30 s)/72°C (30–60 s). Touch-down conditions were as follows: 94°C (3 min), then 20 cycles at 94°C (30 s)/63°C (decreasing by 0.5°C/cycle, 30 s)/72°C (20–60 s) followed by 20 cycles at 94°C (30 s)/53°C (30 s)/72°C (20–60 s). Both regimes were followed by 8 cycles at 94°C (30 s)/53°C (30 s)/72°C (20–60 s) when an M13(-21) sequence was used for fluorescent labelling. A final extension at 72°C was applied for all reactions for 30 minutes. The PCR regime, annealing temperature and elongation time were estimated according to the primer pair properties, and these thermocycling conditions are shown in Table 3.2. The multiplexed amplicons were electrophoresed on a commercial ABI 3730 Genetic Analyzer along with a size standard GeneScan™-500 LIZ® (Applied Biosystems). The resulting genotyping data were analysed with a Microsatellite Plug-in in Geneious Pro 5.5.6. The genotyping data were imported into GenAlex 6.5 (Peakall and Smouse, 2012), which is an Excel Add-In to convert the data into the necessary format for further analysis. Alternatively, PGDSPIDER V2.0.7.2 (Lischer and Excoffier, 2012) was used for the data conversions.

Table 3.2 PCR annealing temperature and elongation time for microsatellite markers.

Locus ID	Annealing temperature (˚C)	Elongation time (sec)
†*1*, †26, †*27*, †*32*, *61*, †*90*	57	60
†*2*, †*20*, †*42*, †51	63 to 53*	60
†*65*, †*67*, †*89*	63 to 53*	30
15	63 to 53*	45

* Touchdown regime, †forward primers directly tagged with a fluorescent dye

3.3.2 mtDNA markers

Mitochondrial DNA sequences for Cytb and cytochrome COI genes of three Nudibranch species, *Notodoris gardineri*, *Chromodis magnifica* and *Berthellina* sp.

TLT-2006 (Medina et al., 2011), and partial COI sequences of *P. maculata* (Wood et al., 2012b) were obtained from Genbank. The sequences were aligned using the Geneious multiple alignment tool, and several primer pairs were designed using the most conserved regions to cover as much of the genes as possible with Primer in Geneious 5.6.5. The list of the mtDNA primers used during this study can be found in Table 3.3. In addition, universal COI (Folmer et al., 1994), 16S rRNA (Palumbi, 1991) and universal Cytb primers from (Merritt et al., 1998) which were modified in this study based on Nudibranch alignment were used to amplify fragments of the genes. Universal COI primer pair did not work. Therefore, only universal reverse COI primer from Folmer et al. (1994) could be used. To check if these genes are polymorphic for *P. maculata,* the gene fragments were amplified first in seven individuals sampled from various locations: two from AKL and one from each of the following locations: TP, TR, WL, NL and PC. The sequence data were aligned to identify the polymorphic regions for *P. maculata,* and species-specific primer pairs were designed to amplify the polymorphic regions of the COI and Cytb genes. 16S rRNA gene was not used for further amplifications due to the lack of polymorphism among the seven individuals. As the variable regions for both genes were too long to be sequenced in a single reaction, two primer pairs were designed to amplify the 5' halves and the 3' halves separately (Table 3.3). These primer pairs were used to genotype the remaining 149 individuals. The PCR for each pair of primers was carried out in a total volume of 50 µL, including 1X Invitrogen™ *Taq* DNA Polymerase buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 4 mM tetramethylammonium chloride, 1X bovine serum albumin, 0.4 µM from each of the forward and reverse primers, 0.02 U Invitrogen™ Taq DNA Polymerase, and 30–90 ng genomic DNA. Either standard or touchdown PCR regimes were applied. Standard conditions were as follows: 94°C (3 min), then 30 cycles at 94°C (30 s)/Tm (30 s)/72°C (80 s). Touch-down conditions were as follows: 94°C (3 min), then 15 cycles at 94°C (30 s)/ 61.5°C (decreasing by 0.5°C/cycle, 30s)/72°C (80 s) followed by 20 cycles at 94°C (30s)/53°C (30s)/72°C (80 s). A final extension at 72°C was applied for all reactions for 6 minutes (Please see Table 3.3 for information on the thermocycling conditions). The PCR products were checked on 2–3 per cent agarose gels that were freshly prepared with 1X TBE buffer and 1X SYBR® Safe DNA Gel Stain per 1 ml of agarose solution. The products were visualised under UV light. 0.1 U/µL Alkaline Phosphatase-Calf Intestinal and 0.05 U/µL Exonuclease I (New England, BioLabs Inc., USA) were used to purify the PCR products. Purification was performed at 37°C for 1

Table 3.3 Primers used to amplify the mtDNA genes.

Gene	Primer ID	Sequence (5' to 3')	Annealing	Extension	Reference	Taxa
16S rRNA	¥16S-F	CGCCTGTTTATCAAAAACAT	57 to 50*	60	Palumbi, 1991	Universal
	¥16S-R	CCGGTCTGAACTCAGATCACGT			Palumbi, 1991	Universal
Cytb	¥7044-F	ACYDGTGGCCTTCAARGCC	50		This study	Nudipleura
	¥7686-R	ACCATTCAGGTATGAGGATG			This study	*P. maculata*
	¥151F-MOD	TGRGGWGCWACAGTAATTACTAA	45	80	Merritt et al., 1998	Nudipleura
	¥270R-MOD	AAYAARAARTATCAYTCAGGTTG			Merritt et al., 1998	Nudipleura
	¥Cytb F	CATGCTAATGGGGCTTCYTTRT	53	80	This study	*P. maculata*
	¥Cytb R	ATWGASCGWAAAATRGCATAAGC			This study	*P. maculata*
	¥8077-F	CTTTCCGACATTACTTGGAGATCC	53	80	This study	*P. maculata*
	¥8926-R	CCWAVTAGAAMTACTATAGCATG			This study	Nudipleura
	†a var268-F	TCGTCAAACTAGACCAGCGG	61.5 to 53*	90	This study	*P. maculata*
	†a var1128-R	ACAAGAGCAATAACTCCCCC			This study	*P. maculata*
	†b 8077-F	CTTTCCGACATTACTTGGAGATCC	61.5 to 53*	90	This study	*P. maculata*
	†b var1584-R	TGCCTTCAAAATCCCACTGA			This study	*P. maculata*
COI	¥dsgn-F	GCGWTGAYTATTTTCDACAAAYC	50	80	This study	Nudipleura
	¥universal-R	TAAACTTCAGGGTGACCAAAAATCA			Folmer et al., 1994	Universal
	¥cons2203-F	GGAATAGATGATGATACTCG	53	60	This study	*P. maculata*
	¥cons2834-R	GTATTACTGTGAAAATCTAAAGG			This study	*P. maculata*
	¥cons1877-F	GATCTGTTTTAGTGACTGC	53	60	This study	*P. maculata*
	¥cons2834-R	GTATTACTGTGAAAATCTAAAGG			This study	*P. maculata*
	†a dsgn-F	GCGWTGAYTATTTTCDACAAAYC	53	80	This study	*P. maculata*
	†a cons688-R	GTTGGTAAAGAATAGGATCACCTC			This study	*P. maculata*
	†b var422-F	CGGGGGCGTCTTCTCTACTA	61.5 to 53*	90	This study	*P. maculata*
	†b var1278-R	CCGTCGAGGCATACCAGAAA			This study	*P. maculata*

*Touchdown thermocycling conditions. † Primer pairs used to genotype all samples. ¥ Primer pairs used to amplify the16S rRNA, Cytb and COI first in seven *P. maculata* individual from six localities. Primer pairs used to amplify [a] 5' half and [b] 3' half of the gene locus.

hour and at 87°C for 15 minutes. Then they were sequenced from both 5' and 3' ends using Sanger Sequencing techniques by Macrogen Inc. (Korea). Geneious 6.1.6 was used to trim, assemble, align and concatenate the obtained DNA sequences, and to convert the aligned sequences into the required formats for further analysis.

3.4 MegaBlast analysis of the mtDNA sequences

The partial Cytb (1060 bp) and COI (1153 bp) sequences obtained from all the individuals were used to interrogate GenBank (NCBI) databases using the MegaBlast algorithm that is implemented in Geneious. Additionally, each of the COI and Cytb sequences from a single individual were partitioned into 100 bp intervals and, searches were repeated for each fragment. The length of the interval was chosen randomly. The first 100 matches were recruited and the matching taxa were evaluated. The aim of this procedure was to check whether the amplified regions matched the sequences from related taxa to ensure that any cross contamination, especially from microorganisms, had not occurred. Such a concern arose as a high amount of haplotype diversity was identified for *P. maculata* samples for COI and Cytb genes.

3.5 Statistical analysis

Samples from Tamaki Strait and Waitemata Harbour were grouped into a single Auckland population for statistical analysis as the sampling locations were in close proximity. Similarly, all the samples from Tauranga were also grouped into a single population. The samples from WH, CR, PC and KK were not included in the population-based statistical analysis due to the small sample sizes from these locations. Population-level calculations were performed only for the five main populations: TP, AKL, TR, WL and NL.

3.5.1 Genetic diversity

3.5.1.1 Microsatellite data

The genotyping data were tested for scoring errors due to stuttering, null alleles, and large allele dropout using MICRO-CHECKER v.2.2 (Van Oosterhout et al., 2004). The total number of alleles (*Na*), allele frequencies, observed heterozygosity (H_o), unbiased expected heterozygosity (H_e) corrected for the small sampling size (Nei, 1978), and private alleles (*PA*) per locus and population were calculated using GenAlex 6.5

(Peakall and Smouse, 2012). Allelic richness (A_R; i.e. the distinct number of alleles observed) and private allelic richness (PA_R; i.e. the number of alleles that are specific to a particular population) per locus and population were calculated using the rarefaction method implemented in ADZE v.1.0 (Kalinowski, 2005) and HP-Rare (Kalinowski, 2005). H_e, A_R and PA_R were used to compare the amount of genetic diversity across the populations. The statistical significance of differences in average A_R, H_e and PA_R between populations was estimated using the one-way analysis of variance test ANOVA (JMP® v5.0.1.2, SAS Institute Inc., Cary, NC, USA).

Departures from Hardy-Weinberg equilibrium (HWE) were estimated using Nei's (1987) heterozygosity-based inbreeding coefficient G_{IS} with 100,000 permutations in GenoDive v2.0b25 (Meirmans and Van Tienderen, 2004). G_{IS} describes the decline in heterozygosity within individuals relative to their populations due to non-random mating. G_{IS} may range from -1.0 (all individuals heterozygous) to 1.0 (all individuals homozygous). A value of zero implies that the population is randomly mating.

Linkage disequilibrium between pairs of loci was estimated in FSTAT v2.9.3.2 (Goudet, 2001) within each population and across all five populations. Correction for multiple testing was performed via the Bonferroni method (Bonferroni, 1936) implemented in the same software to avoid type I errors.

A common problem with microsatellites is genotyping errors due to null alleles, i.e. alleles that are undetectable because of not being amplified during the PCR. Null alleles may interfere with the results of population structure analysis as they distort the allele frequencies and heterozygosity estimations (Kalinowski and Taper, 2006). In order to estimate the potential interference of null alleles in the statistical analysis, null allele frequencies for each population and locus were estimated as in Dempster et al. (1977) using the software FreeNA (Chapuis and Estoup, 2007). FreeNA was also used to estimate the interference of putative null alleles on genetic differentiation between sampling sites by calculating global and pair-wise F_{ST} values (Weir, 1996) for each locus and overall loci either with or without the ENA (exclusion of null alleles) method. The method includes the following steps: estimation of visible and null allele frequencies, allocation of a single-sized allele as a representative of null alleles that are not present in the datasets, correction of allele frequencies and calculation of F_{ST} values

using observed and corrected allele frequencies (Chapuis and Estoup, 2007). The 95% CI was calculated using a bootstrapping method with 10,000 permutations.

3.5.1.2 mtDNA data

Several estimates of genetic diversity including the number of singletons (*Sin*), which are the mutations found only in one individual, the number of haplotypes (*Hap*), the number of segregating (polymorphic) sites (*S*), the average number of nucleotide differences between sequences (*k*) (Tajima, 1983), haplotype diversity (*Hd*) and nucleotide diversity (π) (Nei, 1987) were calculated for the Cytb, COI and concatenated sequences for each sampling location in DNASP 5.10.0.1 (Librado and Rozas, 2009). Hd is the probability that two randomly chosen haplotypes are different (Tajima, 1983), whereas π is the average number of nucleotide differences between each pair of sequences divided by the total length of the sequences for standardisation (Nei, 1987).

The pairwise individual genetic distances were calculated by dividing the total number of base pair differences between individuals by the total number of sites that were compared. This distance matrix (D_{SEQ}) was used to calculate the operational taxonomic units (OTUs) so as to cluster genetically similar individuals. The OTUs were calculated either by clustering individuals with identical sequences, or by clustering individuals that varied by a distance of 0.003 into a single OTU. Then, using the OTU numbers obtained after the clustering process as a diversity index, rarefaction analysis was performed for the five main populations to estimate whether a reasonable amount of diversity had been sampled or not. Rarefaction analysis was conducted by calculating the OTU number for a random subsample of size *n* that is drawn from the population by permutating the data 1000 times. Then *n* (starting from one) is increased one by one at each step until all the samples within the populations are included in the calculation. Preparation of the distance matrix, OTU calculations and rarefaction analysis were performed in Mothur v1.33.3 (Schloss et al., 2009).

3.5.2 Genetic differentiation and population structure

3.5.2.1 Microsatellite data

Exact differentiation test

Allele frequencies in the different populations were compared overall loci in GENEPOP v4.0 (Rousset, 2008) by using the genic differentiation test described by Raymond and

Rousset (1995). The null hypothesis, which is that all populations have the same allele distribution, was tested with the G log likelihood ratio.

Fixation index and related estimators

Global genetic differentiation between the populations and pairwise differentiation between each pair of populations were investigated with the fixation index F_{ST} (Weir and Cockerham, 1984), standardised fixation index G''_{ST} (Hedrick, 2005; Meirmans and Hedrick, 2011), Jost's (2008) differentiation (*Dest*) and R_{ST} (Slatkin, 1995) per locus and across loci in GenAlEx and GenoDive. The significance of the differentiation for all the estimators was estimated with 9,999 permutations.

Power test

The statistical power to detect true population differentiation and α-error probability (type I; probability to detect genetic differentiation while there is none) at various true levels of divergence was estimated for the empirical sample sizes, number of loci and allele frequencies using the programme POWSIM v4.1 (Ryman and Palm, 2006). Powsim simulations mimic sampling of genes from populations at various levels of a predefined expected level of divergence under the assumption of no migration and mutation (Ryman and Palm, 2006). The simulation was performed for effective population size N_e=1000 with several numbers of generations (t) to obtain a specific F_{ST} (t=10 for F_{ST} =0.001, t=25 for F_{ST}=0.0025, t=51 for F_{ST}=0.005, t=102 for F_{ST}=0.01, t=10 for F_{ST}=0.001, and finally t=0 for F_{ST}=0 to detect α error probability). One thousand replications were run for each level of differentiation. The analyses were conducted using 1000 dememorisations, 100 batches and 1000 iterations per batch.

Multivariate approach

In addition to fixation indices and the *Dest* method, genetic differentiation between the populations was estimated using the Manhattan distance approach (Czekanowski, 1909) (*DM*), which calculates the mean character differences between individuals and Clonal (*DCL*) (Meirmans and Van Tienderen, 2004) distances. For *DM*, firstly the Manhattan distances over genotyped loci were calculated for each pair of individuals using R (Team, 2013). The resulting matrix provided the number of mismatching alleles per locus divided by two. *DCL*, which is specific to microsatellite markers, calculates the number of mutation steps needed to transfer the genotype of the first individual to that

of the second assuming the SMM model (Meirmans and Van Tienderen, 2004). *DCL* was calculated in GenoDive (Meirmans and Van Tienderen, 2004). Further statistical analysis based on two distance matrices was performed in the statistical software package PRIMER 6 (Clarke and Gorley, 2006) with PERMANOVA+ (Anderson et al., 2008) (Primer-E Ltd., Plymouth, United Kingdom) using a conventional multivariate approach.

In the multivariate approach, the samples can be imagined as objects in multi-dimensional space where the distance matrix represents the relationship between them. Firstly, the *DM* and *DCL* matrices were subjected to 2-dimensional (2D) non-metric multi-dimensional scaling (MDS), which is an ordination technique for visualising the dissimilarities between objects (Kruskal, 1964). How well the plot represents the original matrix is calculated and outlined by a "stress" value – where "0" indicates a perfect fit and "1" a total mismatch. Permutational Multivariate Analysis of Variance (PERMANOVA) with Type III (partial) sums of squares was conducted to test the null hypothesis that there is no difference between sampling locations. Again, if we imagine the samples as a data cloud in multivariate space, PERMANOVA analysis measures the location of the centroids of each group and overall data, and statistically tests whether the locations of the group centroids are significantly different from the overall centroid, or not. The statistic for PERMANOVA is the pseudo F-ratio. As the pseudo F-ratio gets larger, support for the null hypothesis weakens. The significance of the pseudo F-ratio (*F*) was estimated by rearranging the labels of the samples within and between the populations arbitrarily by unrestricted permutations of raw data for 9,999 iterations, and calculating the pseudo F-ratio (*F*') for each permutation. The *P*-value is the ratio of $F' \geq F$ (Anderson, 2001; McArdle and Anderson, 2001).

Canonical Analysis of Principal coordinates (CAP) (Anderson and Willis, 2003) was also performed to visualise patterns of genetic differences between the sampling locations. CAP, which is a constrained ordination method, is designed to visualise multivariate objects by projecting them to the axes that maximise the variation between the *a priori* groups. Additionally, a cross-validation test was performed with CAP by following the one-leave-out approach. This test evaluates how well the model explains the observations and how different the groups are by calculating the percentage of correct classification of the samples to their predefined group. This is achieved by

leaving out a particular individual, calculating the centroid of each group within the multivariate space, and assigning the sample to the group whose centroid is closest after inclusion of the sample. For CAP analysis, it is important to choose an appropriate number of axes (*m*) that include the highest possible variation with some restrictions. *m* should be a number that maximises the success of allocation by the one-leave-out approach and/or minimises the residual sum of squares (SSres), i.e. the sum of squared deviations of samples from the new axes. CAP was also used to test the null hypothesis of no difference between the centroids of the groups with 9,999 permutations (Anderson et al., 2008; Anderson and Willis, 2003).

Population assignment with the Bayesian approach

STRUCTURE 2.3.4 (Pritchard et al., 2000; Pritchard et al., 2010) was used to determine the probable number of distinct populations (*K*), to assign the individuals to the populations and to identify the admixed individuals using the Bayesian assignment approach. Analysis was conducted with and without introducing an *a priori* sampling location assuming an admixture model and correlated allele frequencies across the populations. The lambda (λ) value, which is a parameter that specifies the prior distribution of allele frequencies (Evanno et al., 2005), was first estimated by running the programme with a burn in of 50,000 runs and a further 5000,000 Markov Chain repetitions for K=1. The same burn in and Markov chain values were used for further calculations. Using the estimated λ value (λ=0.57739), 10 iterations each for each *K* value, where *K* was assumed to range from one to ten, were performed. The most likely number of groups was resolved using the ΔK method (Evanno et al., 2005) with the Structure Harvester Web v0.6.93 (Earl and vonHoldt, 2012). The results from the Structure Harvester were introduced to the CLUMPP v1.1.2 software (Jakobsson and Rosenberg, 2007), which produces the average permutated individual and population clustering outputs throughout the ten replicates for each *K* value. Then Destruct v1.1 (Rosenberg, 2004) was used to visualise the results.

Analysis of molecular variance

AMOVA (Excoffier et al., 1992) was performed in ARLEQUIN to determine the hierarchical genetic structure based on SMM (R_{ST}-based). To identify the significance level, 20,000 permutations were used. Two separate calculations were performed: not introducing any structure, grouping the TP, AKL and TR populations in a Northern

cluster and grouping the WL and NL populations in a Southern cluster. This nesting design was decided based on the population structure that was suggested by F-statistics, multivariate and STRUCTURE analyses.

3.5.2.2 mtDNA data

Haplotype networks

POPART v1 (http://www.leigh.net.nz/software.shtml) was used to create a median joining haplotype network (MJN) (Bandelt et al., 1999) for the Cytb, COI and concatenated sequences to visualise the mutational relationship between the haplotypes and their geographic distribution. As the COI and Cytb genes were amplified by two sets of primers, the fragments from the wrong individuals could have been assembled erroneously. MJN analysis was performed separately for 5' and 3' halves of the Cytb sequence (each are 530 bp in size) that are amplified by two different primer sets. The 3'-half included overlapping sequences from the 5'amplicon. Lastly, MJN analysis was conducted for COI sequences excluding the third positions from the data in order to test whether homoplasy obscured any present population structure. I hypothesised that the COI sequences would have been affected by homoplasy more severely than the Cytb sequences as the former contained a higher degree of polymorphism. Therefore, only the COI sequences were included in the final analysis.

Saturation test was performed also using DAMBE v6.2.9 (Xia, 2000; Xia and Xie, 2001) for both COI and Cytb sequences using "Test by Xia et al." (Xia et al., 2003). The proportion of invariant sites (P_{inv}) were estimated by Jmodeltest v0.1.1 (Posada, 2008) with the Akaike information criteria (AIC). The P_{inv} values (0.844 and 0.789 for Cytb and COI, respectively) obtained from the most likely models (HKY+I and TrN+I for Cytb and COI, respectively) suggested by the software and default settings for other parameters were used for the calculations.

Multivariate analysis

D_{SEQ} matrixes obtained from both the COI and Cytb sequences were subjected to MDS, PERMANOVA and CAP analyses, as described in subsection 3.5.2.1 under the heading "Multivariate approach", to reveal any differences between the sampling locations.

F-statistics

The Akaike information criterion (AIC) implemented in Jmodeltest v0.1.1 (Posada, 2008) was used to choose a mutation model for nucleotide substitutions observed in the Cytb and COI sequences. Out of 88 models, the algorithm suggested TrN+I and HKY+I as the most likely models to explain the COI and Cytb data, respectively (Hasegawa et al., 1985). Two different F-statistics were calculated in ARLEQUIN to estimate population differentiation: F_{ST}, which is based on haplotype frequencies and does not take nucleotide distances between the haplotypes into account (Nei, 1987), and Φ_{ST} (Excoffier et al., 1992), which is based on nucleotide diversity. For Φ_{ST}, the Tamura-Nei (TrN) model (Tamura and Nei, 1993) was used for both genes as a mutational model to calculate the distances between the haplotypes because it was the closest model to HKY+I. The HKY model assumes variable base frequencies, and different rates of transitions and transversions (Hasegawa et al., 1985). The TrN model is similar to HKY, but it also distinguishes between purines and pyrimidines (Tamura and Nei, 1993).

Analysis of molecular variance

The hierarchical structure was investigated by AMOVA in ARLEQUIN using both F-statistics and Φ-statistics, and introducing the same structure pattern used for the microsatellite data analysis: grouping TP, AKL and TR into a Northern cluster, while grouping WL and NL into a Southern cluster (please refer heading "Analysis of molecular variance" under subsection 3.5.2.1) and not introducing any population structure. Significant levels of differentiation were estimated with 20,000 permutations. PERMANOVA analysis was performed using the D_{SEQ} matrix by contrasting the TP, AKL and TR populations (north) against WL and NL (south).

3.5.3 Migration

The microsatellite data were analysed with GeneClass2 (Piry et al., 2004) to identify the first generation migrants, i.e. individuals born in a population other than the one in which they were sampled, using the Bayesian criterion of Rannala & Mountain (1997). The likelihood computation was based on Lhome/Lmax, where Lhome is the likelihood of sampling an individual's genotype from its sampled location, and Lmax is the ratio of Lhome to the highest likelihood value among all the available populations. This likelihood estimation was preferred over Lhome due to its higher power of detecting the migrants. A Monte-Carlo resampling algorithm was used as in Paetkau et al. (2004) with

a Type I error threshold (alpha) of 0.01 and 10,000 simulated individuals to calculate the probability of membership of each individual in five sampling locations.

3.5.4 Spatial analysis

To test for isolation by distance (Slatkin, 1993), pair-wise values of $F_{ST}/(1-F_{ST})$ (Rousset, 1997) for both the microsatellite and Cytb data, which revealed a more pronounced population structure compared to COI data, were plotted against the minimum coastal distances (km) between the sampling sites. The coastal distances were calculated using Google maps (http://www.daftlogic.com/projects-google-maps-distance-calculator.htm). Mantel tests for isolation by distance (IBD) were performed to investigate any correlation between the genetic and geographical distances. These analyses were conducted using IBDWS (the isolation by distance web service at http://ibdws.sdsu.edu/~ibdws/) (Jensen et al., 2005).

3.5.5 Demographic analysis, neutrality test and genealogy

3.5.5.1 Microsatellite data

Population expansion was investigated for each population using the *k* and *g*-tests (Reich et al., 1999; Reich and Goldstein, 1998) implemented in the KGTESTS Excel macro (Bilgin, 2007). These approaches use the distribution of allele sizes to detect demographic events. This macro was used to calculate *P* values of k-statistics using a binomial distribution, and *g* values (Bilgin, 2007).The significance of the *g* values was determined according to 5% percentile cut-off values for a given sample size and loci. These theoretical values were established by Reich et al. (1999) by simulating the *g* values under constant population size.

The possibility of recent population reduction was tested using the programme BOTTLENECK v1.2 (Cornuet and Luikart, 1996; Piry et al., 1999). BOTTLENECK calculates the probability of a recent bottleneck by testing whether a significant number of loci show heterozygosity excess or not (Luikart and Cornuet, 1998). Three different microsatellite mutational models were used for the calculations: IAM and SMM, which are extreme models, and an intermediate two-phase model (TPM). TPM is an extension of the SMM model, where most mutations result in an increase or loss of one repeat unit, but changes in large number of units are possible (Bhargava and Fuentes, 2010). The analysis was conducted with 10,000 simulation iterations. 70% SMM with a

variance of 30 was assumed for TPM, which was the default setting. The Wilcoxon signed rank test was used to determine the significance of heterozygosity excess, which was suggested for studies with fewer than 20 loci (Piry et al., 1999).

A possible sign of a recent bottleneck was investigated also by a mode-shift analysis implemented in BOTTLENECK. The analysis is based on observed deviations from the L-shaped distribution pattern that is expected for non-bottlenecked populations (Luikart and Cornuet, 1998).

3.5.5.2 mtDNA data

The population parameter θ and two tests of neutrality, Tajima's *D* (Tajima, 1989) and Fu' *Fs* (Fu, 1997), were calculated for the COI and Cytb sequences in ARLEQUIN to test deviations from neutrality and demographic changes within and across the populations. Confidence intervals were determined by 20,000 permutations. Historical population expansion events within and across the *P. maculata* populations were investigated by mismatch distribution analysis in ARLEQUIN with 20,000 bootstrap replicates for the COI, Cytb and concatenated sequences. The null hypothesis of expansion was statistically tested with the sum of squared deviations (*SSD*) from the expected values (Schneider and Excoffier, 1999) and Harpending's *raggedness* index (Harpending, 1994). Significant *raggedness* and *SSD* values ($P<0.05$) are taken as evidence for rejecting sudden population expansion (Schneider and Excoffier, 1999). Mismatch distribution analysis also estimated tau (τ), which is the age of the population expansion. The approximate date of population expansion across all samples was calculated for the COI gene at the website http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss, 2011) by converting τ to time since expansion (t) in years using the formula $t = \tau /2\mu k$, where k is the sequence length (Excoffier et al., 2005). One year of generation time and μ of 5.3% divergence/mya (average μ for marine invertebrate COI sequences) (Crandall et al., 2012) were assumed for the calculations. The significance of the above tests was assessed with 20,000 bootstrap replicates.

CHAPTER 4: GENETIC DIVERSITY AND STRUCTURE OF *P. MACULATA* POPULATIONS AS REVEALED BY MICROSATELLITE DATA

4.1 INTRODUCTION

Marine species that are distributed across a large range of environmental conditions may exhibit significant phenotypic variations partitioned between populations inhabiting distinct geographical regions (e.g. Vergeer and Kunin, 2013). The underlying mechanisms of phenotypic variations among locations have always intrigued evolutionary biologists and population geneticists (Endler, 1977). The variations are often genetically determined and arise from a combination of genetic drift and local adaptation (e.g. Vergeer and Kunin, 2013). In some instances, phenotypic variations are determined by environmental factors because the phenotypic responses of organisms to variable environmental factors are affected by mechanisms underpinning phenotypic plasticity (Trussel and Eter, 2001).

One example of geographical variability has been observed in the tetrodotoxin (TTX) levels in *Pleurobranchaea maculata* populations around New Zealand (NZ). Significant individual and seasonal variations in TTX concentrations have been observed for *P. maculata* populations from various regions of the North Island, whereas no TTX was found in 21 individuals sampled from the South Island populations (Wood et al., 2012b). The variability of toxicity in TTX-bearing organisms is not unique to *P. maculata*. Significant variation within and between populations has been observed for other organisms, such as the newt *Taricha granulosa* (Hanifin et al., 2008; Hanifin et al., 1999), the red-spotted newt *Notophthalmus viridescens* (Yotsu-Yamashita et al., 2012), the gastropods *Rapana rapiformis* and *R. venosa venosa* (Hwang et al., 1991), and the horseshoe crab *Carcinoscorpius rotundicauda* (Dao et al., 2009). Significant seasonal differences have also been observed in other species, such as *Cephalothrix* sp. ribbon worms (Asakawa et al., 2003), the starfish *Astropecten polyacanthus* (Miyazawa et al., 1985) and the puffer fish *Lagocephalus sceleratus* (Sabrah et al., 2006).

Although the cause of variation in TTX levels in TTX-containing taxa is not known, the geographical variability in *Taricha* newts is attributed to its co-evolutionary arms race

with its TTX-resistant predator, the garter snake *Thamnophids* sp (Brodie Jr et al., 2002). The variability in TTX levels found in puffer fish has been attributed to exposure to variable amounts of toxic food sources (Noguchi et al., 2006a; Noguchi et al., 2006b; Noguchi et al., 2004; Yu et al., 2011). However, toxin sources for many taxa are still not entirely understood. For *P. maculata*, there is evidence supporting the possibility that toxin accumulation may occur through feeding (Khor et al., 2014; Salvitti et al., 2015c; Wood et al., 2012a). One such TTX source in the Tauranga region is a flatworm *Stylochoplana* species that contains high TTX concentrations. Additionally, very low or trace amounts of TTX have been identified in the sand dollar *Arachnoides zelandiae*, the gastropod *Turbo smaragdus*, the crab *Macropthhalmus hitipes*, the corraline turf algae *Carollina officinalis* (Khor et al., 2014) and the bivalve *Paphies australis* (McNabb et al., 2014), but the concentrations detected in these species are too low to account for the accumulation of high TTX concentrations recorded in *P. maculata* individuals (Khor et al., 2014). Further surveys are required in order to identify common and abundant toxic dietary sources in other regions harbouring tetrodotoxin-associated *P. maculata* individuals. It is possible that there exist yet to be discovered toxin sources of TTX (Khor et al., 2014). A further possibility is that the slugs themselves synthesise TTX (Wood et al., 2012a; Wood et al., 2012b), but the complex chemical structure of TTX makes this unlikely (Chau et al., 2013). In addition, if *P. maculata* themselves are the source of TTX, it is difficult to understand the existence of non-TTX-producing populations. An alternate possibility is that TTX arises from commensal or symbiotic microorganisms that are associated with *P. maculata*. Despite attempts, culturable bacterial species producing tetrodotoxin have not yet been identified. However, a bacterial origin has not been ruled out (Chau et al., 2013; Salvitti et al., 2015b; Wood et al., 2012a; Wood et al., 2012b). In summary, neither the sources of TTX nor the mechanisms underpinning the variations in TTX-containing organisms, including *P. maculata*, are understood.

Understanding the genetic structure of *P. maculata* populations from different regions of New Zealand will reveal genetic diversity within populations, the distribution of diversity between populations, the level of genetic connectivity between individuals, as well as the nature of the populations of this rarely studied organism. Such analyses may also provide some insights into the origins and evolution of TTX-producing sea slugs, indirectly. More specifically, population genetics will reveal whether the individuals

from distinct toxic and non-toxic regions are genetically isolated. Any such genetic correlation may imply that differences in the genetic background of different NZ *P. maculata* populations are behind the variation in TTX concentrations. However, such a correlation may also imply that some additional environmental factor underpins the observed variation. Population structure and migration analysis of the TTX-bearing *T. granulosa* newt from various localities in western North America (Hanifin et al., 2008; Kuchta and Tan, 2005; Ridenhour et al., 2007) and the red-spotted newt *Notophthalmus viridescens* (Yotsu-Yamashita et al., 2012) showed that highly toxic and non-toxic populations were genetically connected. It is not clear whether *P. maculata* also lacks such a correlation between toxicity and population structure.

Barriers to dispersal such as oceanographic and tectonic processes can exert a strong influence on the genetic structure of populations of marine organisms (Grosberg and Cunningham, 2001; Riginos et al., 2011). One such barrier exists in central NZ. It is associated with coastal upwelling regimes at 42°S and/or Cook Strait, which separates the coasts of the North and South Islands. This barrier often results in a significant disjunction between the North and South Island populations of species from a variety of taxa, such as sea grass (Jones et al., 2008), bivalves (Smith et al., 1989; Wei et al., 2013a), teleosts (Bernal-Ramírez et al., 2003; Hickey et al., 2009), polyplacophores (Veale, 2007), gastropods (Waters et al., 2005; Will and Gemmel, 2008),echinoderms (Ayers and Waters, 2005; Sponer et al., 2002) and arthropods (Schnabel et al., 2000).

The way a species responds to geographical barriers is connected to its life history traits, which determine its dispersal capacity (Galarza et al., 2009; Keeney et al., 2013; Pelc et al., 2009). For example, marine organisms with a planktonic (pelagic) larval stage show greater dispersal capacity and less population structure compared to direct-developing ones (Bradbury et al., 2008; Weersing and Toonen, 2009). In addition, there is a positive correlation between pelagic larval duration (PLD) and the dispersal capacity – and consequently, connectivity of populations – to some extent (Shanks et al., 2003a; Shanks and Brink, 2005; Shanks et al., 2003b), although conflicting findings based on literature surveys revealed only a very weak correlation, if any (Riginos et al., 2011; Weersing and Toonen, 2009). Another trait that affects the dispersal capacity of a species is the habitat requirement. Habitat generalists have a better chance of having extended gene flow than habitat specialists (Pelc et al., 2009). Recognising that *P.*

maculata has a moderate-length planktonic larval stage that lasts about three weeks, and that it can be found in a wide range of habitats ranging from sandy sediment to rocky reefs, and from shallow subtidal areas to intertidal areas deeper than 300 m (Willan, 1983; Wood et al., 2012b), decreased population structure and a moderately high level of genetic connectivity between different populations can be expected. The results of this study will show whether these expectations will be met for the NZ *P. maculata* populations.

In the present chapter, I analysed the genetic diversity and structure of 146 *P. maculata* individuals from five main localities representing four regions in the North Island (Ti-Point-TP, Auckland-AKL, Tauranga-TR and Wellington-WL) and one region from the South Island (Nelson-NL) using twelve nuclear microsatellite markers (Yildirim et al., 2014). Regarding TTX concentrations, TP, AKL and TR contain highly toxic individuals, the WL individuals are slightly toxic, whereas the NL populations contain either no, or trace amounts of TTX (McNabb et al., 2010; Wood et al., 2012b) based on partial toxicity analysis. In addition, Kaikoura (KL) populations are considered to be non-toxic based on one sample assayed (Wood et al., 2012b). Although no toxicity assay has been performed for samples from Picton (PC), PC was assumed not to harbour toxin-producing populations given the general trend of low toxicity levels in the South Island (McNabb et al., 2010; Salvitti et al., 2015a; Wood et al., 2012b). I also tested for a correlation between geographical variability in TTX concentration and population structure. Additionally, the sampling locations allowed me to investigate whether the known central NZ disjunction might also affect the genetic structure of the *P. maculata* populations.

4.2 RESULTS

4.2.1 Tetrodotoxin concentrations in the Wellington samples

Individuals of *P. maculata* from WL are thought to contain “low concentrations of TTX” (Wood et al., 2012b), but this generalization is based on a single sample. This very low sample size was not enough to gain a general idea of toxicity levels in slugs from the WL region. Therefore, eight individuals collected from WL in October 2012 were tested for their TTX concentrations by Paul McNabb using the Liquid Chromatography-Mass Spectrophotometry technique (McNabb et al., 2014). Only three individuals tested positive for TTX with very low concentrations (0.12, 0.16 and

0.5mg/kg). The other five samples were diluted 50-fold before the assay, which could have prevented identification of very low levels of TTX because the limitation of the assay is approximately 0.1 mg/kg (McNabb et al., 2014). The results were consistent with the previous finding that reported low TTX levels in WL (2.2 mg/kg) (Wood et al., 2012b). Based on these findings, WL individuals were classified as "slightly toxic". NL samples were classified as "non-toxic" because only one sample was reported to contain trace amounts of TTX (0.45 mg/kg) (McNabb et al., 2010). All the other NL samples tested (25 individuals) were non-toxic (Khor et al., 2014; McNabb et al., 2010; Wood et al., 2012b).

4.2.2 Genetic diversity

Fifteen microsatellite markers were initially used for genotypic analyses (see section 3.2.1); however, null alleles were observed for loc *90* in all populations. Additionally, stutter alleles strongly interfered with the allele scoring for loc *67* and for loc *90*. In addition, background noise in the genotyping data affected the allele scoring for locus *15*. These three loci were therefore excluded from all the subsequent analyses.

4.2.2.1 Genetic diversity estimators

One hundred and twenty-one alleles were detected across 12 microsatellite loci in 146 individuals (summary statistics in Table 4.1, detailed statistics for each locus and population in Supplementary Table 2 in Appendix Two). The genetic diversity across the loci was high. The total number of alleles per locus across the populations ranged from five (loc *61* and *65*) to 23 (loc *1*). The average number of alleles at specific loci ranged from 2.6 at loc *65* to 13.6 at loc *1* across the populations, with a grand average of 6.67±0.43 alleles over the loci and populations. The expected heterozygosity (H_e) and observed heterozygosity (H_o) ranged from 0.407 (loc *65*) to 0.843 (loc *1*) and 0.407 (loc *65*) to 0.842 (loc *1*), respectively. Global H_o and H_e across the loci and populations were 0.661±0.019 and 655±0.019, respectively.

The allelic richness (A_R; i.e. the distinct number of alleles observed), H_o and H_e values across the 12 loci per population are presented in Table 4.2. A_R per population was calculated using the rarefaction method as the sample size varied across the sampling locations. Rarefaction curves for A_R across the 12 loci reach the exponential phase for each sampling location, which indicates that a reasonable amount of diversity was sampled for each location (Figure 4.1A). The one-way analysis of variance (ANOVA)

test was performed to test the null hypothesis that populations do not exhibit significant differences in genetic diversity. The test shows no significant difference between the populations in A_R ($F_{4,146}$=0.0048, P=1.000), H_o ($F_{4,146}$=0.6290, P=0.6438) and H_e ($F_{4,146}$=0.4210, P=0.7928). Each of these statistical tests shows that the populations do not differ in terms of genetic diversity.

Table 4.1 Summary of the genetic diversity statistics at 12 microsatellite loci across five *P. maculata* populations (n=146).

Locus	***Na***	***Size-Range (bp)***	H_O	H_e	G_{IS}	***P-HWE***	F_{ST}	G''_{ST}	*Dest*	R_{ST}
1	23	108–208	0.842	0.843	0.000	0.484	0.057[c]	0.328[c]	0.291[c]	0.247[c]
2	9	105–137	0.671	0.742	0.090[a]	0.02	0.026	0.044	0.033	–0.010
20	10	128–164	0.737	0.720	-0.024	0.315	0.014	-0.007	-0.005	-0.002
26	6	141–165	0.710	0.660	-0.079	0.09	0.115[c]	0.365[c]	0.274[c]	0.024[a]
27	16	91–142	0.733	0.736	0.004	0.453	0.045[c]	0.142[c]	0.108[c]	-0.012
32	6	107–122	0.737	0.699	-0.055	0.154	0.071[c]	0.233[c]	0.175[c]	0.040[b]
38	11	157–187	0.838	0.813	-0.032	0.205	0.035[b]	0.137[b]	0.114[b]	0.019
42	8	103–136	0.572	0.576	0.007	0.442	0.007	-0.024	-0.014	-0.007
51	8	155–187	0.457	0.452	-0.011	0.444	0.058[b]	0.097[b]	0.046[b]	0.112[c]
65	5	114–132	0.407	0.407	-0.001	0.583	0.181[c]	0.340[c]	0.174[c]	0.203[c]
89	14	156–184	0.703	0.692	-0.016	0.436	0.132[c]	0.467[c]	0.378[c]	0.253[c]
61	5	169–181	0.523	0.519	-0.008	0.518	0.041[b]	0.067[b]	0.036[b]	0.103[c]
Ave	6.667		0.661	0.655	-0.009	0.32	0.064[c]	0.175[c]	0.122[c]	0.136[c]
SE	0.428		0.020	0.019		-	0.014	0.046	0.0352	0.088

Na number of alleles, H_O observed heterozygosity, H_e unbiased heterozygosity, G_{IS} Nei's (1987) inbreeding coefficient, F_{ST} Weir and Cockerham's (1984) fixation index, [a] significant deviation from HWE (P<0.05), G''_{ST} fixation index corrected for the multi-allelic loci and sampling bias (Hedrick, 2005; Meirmans and Hedrick, 2011), R_{ST} Slatkin's (1995) fixation index for microsatellite markers under the SMM model. Significant genetic differentiation: [b] (P<0.01), [c] (P<0.001).

Private alleles were assessed to understand the distinctiveness of each population. Twenty-six alleles are unique to a single population, ranging from one in TP to sixteen in NL. The mean number of private alleles per population after rarefaction analysis (PA_R) ranges from 0.21 (in TR) to 0.69 (in NL) (Table 4.2), and ANOVA shows no significant differentiation between the sampling loci ($F_{4, 146}$=1.4616, P=0.2264). Since the PA_R values of WL and NL are high compared to those of TP, AKL and TR (Table 4.2), ANOVA was performed to test for differences after nesting populations into two clusters: TP-AKL-TR and NL-WL. TP, AKL and TR were designated the "Northern cluster". Similarly, the WL and NL populations were labelled as the "Southern cluster" due to their relative geographical location compared to the Northern cluster. These

cluster names will be used in subsequent sections. The Southern cluster has a significantly higher PA_R than the Northern cluster ($F_{1,146}$=4.4976, P=0.0385), which indicates that WL and NL are distinct from the rest of the populations.

Table 4.2 Summary of the genetic diversity statistics for 146 *P. maculata* individuals sampled from five locations. Twelve microsatellite markers were used.

Pop	*N*	*Na*	A_R	H_O	uH_E	G_{IS}	*PA*	PA_R
Ti Point (TP)	20	6.083	6.02 ±0.55	0.679	0.680	0.001	1	0.16
Auckland (AKL)	30	6.667	5.88±0.68	0.681	0.667	-0.021	4	0.46
Tauranga (TR)	33	6.883	5.89	0.697	0.676	-0.032	2	0.20
Wellington (WL)	18	6.00	6.00	0.606	0.636	0.047	3	0.47[a]
Nelson (NL)	45	7.75	5.93	0.641	0.616	-0.041	16	0.77[a]

N: sample size, *Na* number of alleles, A_R: mean allelic richness based on 18 diploid individuals, H_o observed heterozygosity, H_e: unbiased expected heterozygosity, G_{IS}: Nei's (1978) inbreeding coefficient index, *PA*: total private allele numbers, PA_R: mean private allele numbers based on 18 diploid individuals, The NL-WL populations nested in one group had a significantly higher PA_R compared to the rest when nested within one group.

4.2.2.2 Analysis of the Hardy-Weinberg equilibrium

Departures from the Hardy-Weinberg equilibrium (HWE) were tested with the inbreeding coefficient G_{IS} per locus. G_{IS} across the populations range from -0.079 (loc *26)* to 0.098 (loc *2*), and are significantly different from zero at loc *2* (P=0.02) (Table 4.1). G_{IS} values across the loci for each population range from -0.041 (NL) to 0.047 (WL) (Table 4.2). NL is not in HWE at loc *20* and loc *61* (P<0.05), however; across the loci, all the populations including NL meet the Hardy-Weinberg expectations (Supplementary Table 2).

4.2.2.3 Investigation of null alleles

MICROCHECKER suggests homozygote excess and possibly null alleles at loc *2*, *65* and *61* for the TP population, at loc *20* and *51* for the AKL population, at loc *51* and *65* for the TR population, at loc *2* and *20* at the WL population, and at loc *42* for the NL population. Among these loci, scoring errors due to stuttering are suggested for loc *61*, which is indicated by the highly significant shortage of heterozygote genotypes with alleles of one repeat unit difference. A stuttering pattern was previously observed for the allele calling for loc *42* and *65*, but for not *61*. However, the stuttering pattern observed for these two markers (loc *42* and *65*) did not strongly interfere with the allele calling. Preliminary tests for differentiation were therefore performed both with and without these two markers. The markers were included in the further analysis because their exclusion was not found to significantly impact on the results.

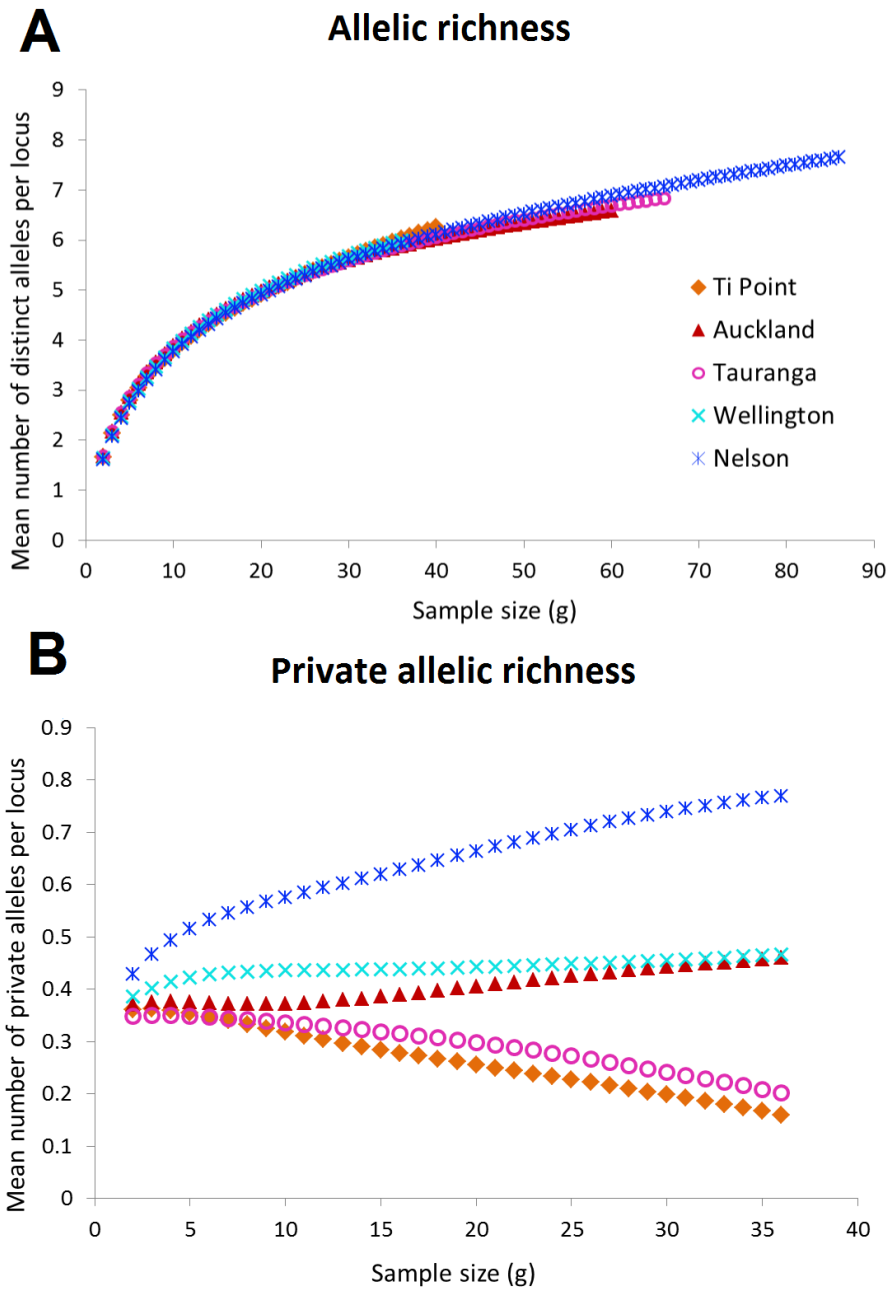


Figure 4.1 Rarefaction curves for the microsatellite data. The mean number of **A**) distinct alleles per locus and **B**) private alleles per locus as a function of a standardised sample size for the five main sampling locations. Population abbreviations within the text are indicated in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL), and Nelson (NL).

In order to investigate the putative negative effects of null alleles on structure analysis, the frequencies of the null alleles were estimated for each locus and populations using FreeNA (Chapuis and Estoup, 2007), and the results are presented in Supplementary Table 3. The likelihood of null allele frequencies per locus per population varies from zero to a maximum of 0.096 (loc *61* in TR). There is no pattern of null frequency as a function of population, although the algorithm implemented by FreeNA suggests null

alleles for all sampling locations at loc *2*. However, the frequency for null alleles does not exceed 10% for any population at any locus, which suggests a low level of null allele frequencies.

As a further step to investigate the possible effect of null alleles on the population structure analysis, FreeNA was used to calculate the global and pair-wise F_{ST} values (for each locus and overall loci) using the ENA (excluding null alleles) method that is described in Chapuis and Estoup (2007). Global F_{ST} values over all the loci were calculated as 0.0665 and 0.0669, with and without null alleles, respectively. Similarly, only slight changes were found for each global F_{ST} value as well as pairwise F_{ST} values between populations for each locus (Supplementary Table 4A and B) after accounting for null alleles, suggesting that the measurement of genetic differentiation between sampling sites was only slightly affected due to the existence of putative null alleles. Therefore, the original allele frequencies were used for the further analysis.

4.2.3 Linkage disequilibrium

The extent of LD was determined with genotype randomisations between 12 pairs of loci at each population as well as across populations using the log-likelihood ratio *G*-statistic (Goudet et al., 2002). No significant linkage disequilibrium was detected between the 12 pairs of loci across the five populations or within each population after Bonferroni correction (P<0.05, Supplementary Table 5).

4.3 Genetic differentiation and structure

4.3.1 Exact differentiation test

Allele frequencies vary across the *P. maculata* populations for most of the loci (Supplementary Table 6). Graphical presentation of the frequencies (Supplementary Figure 1) shows that TP, AKL and TR have closer frequencies of common alleles at each locus than those of WL and NL. Similarly, the frequencies in WL and NL are similar to each other.

The allele frequencies between the populations were compared by pooling the frequencies over all the loci using the probability test for genic differentiation described by Raymond and Rousset (1995). The test found that WL and NL are significantly differentiated from TP, AKL and TR ($\chi 2$=infinity, d.f.=24, P<0.00000) (Table 4.3).

There is significant differentiation also between WL and NL ($\chi 2$=infinity, d.f.=53.20, P=0.0006), whereas no significant differentiation was found between the TP, AKL and TR populations (P<0.05).

4.3.2 Fixation index and related estimators

Genetic differentiation between the populations was estimated using various fixation indices and the related estimators F_{ST}, G_{ST} and R_{ST}, as well as *Dest*. F_{ST} is the mean reduction of heterozygosity in a population relative to the total heterozygosity. The F_{ST} value varies between zero and one. Zero means that allele composition of two compared populations is identical, whereas one means that there is no shared allele (Wright, 1951; Wright, 1965). The standardised fixation index G''_{ST} (Hedrick, 2005; Meirmans and Hedrick, 2011), R_{ST} (Slatkin, 1995) and Jost's (2008) differentiation (*Dest*), all of which are independent of within-population diversity, were also calculated to estimate population differentiation. R_{ST} assumes a stepwise mutational model (SMM) for allele evolution unlike the other estimators, which assume an infinite allele model (IAM). *Dest* is different from F-statistic based measurements because it takes allelic diversity into account as a diversity index instead of heterozygosity and variance in allele frequencies, and it partitions diversity in a multiplicative manner rather than in an additive one (Jost, 2008).

The values of all the differentiation estimators vary between the loci across all the sampling locations from -0.009 to 0.181 for F_{ST}, from -0.007 to 0.467 for G''_{ST}, from -0.014 to 0.378 for *Dest* and from -0.012 to 0.253 for R_{ST}. Differentiation is significant at eight loci for F_{ST}, G''_{ST} and *Dest* (P<0.01 at three loci and P<0.001 at six loci) and at seven loci for R_{ST} (P< 0.05 at two loci and P<0.001 at five loci) (Table 4.1). Global differentiation between the populations across the 12 loci is 0.064 (±0.014), 0.175 (±0.046), 0.122 (±0.035) and 0.136 (±0.088) for F_{ST}, G''_{ST}, *Dest* and R_{ST}, respectively, and the differentiation is highly significant (P≤0.0001) with all indices.

Pairwise F_{ST}, G''_{ST}, *Dest* and R_{ST} values between the populations across the twelve loci range from -0.009 to 0.122, -0.286 to 0.3367, -0.0191 to 0.246 and -0.016 to 0.237, respectively (Table 4.3). The pairwise *Dest* values between the pairs of populations at each locus are presented in Supplementary Figure 7. The pairwise R_{ST} distances between the populations are represented as a multi-dimensional scaling (MDS) graph in Figure 4.2. None of the estimators suggests significant differentiation between TP, AKL and TR

(*P*>0.05). Negative differentiation values for each pairwise comparison within the Northern cluster show that they are highly connected. On the other hand, all the indices show that these populations are significantly differentiated from the WL and NL populations (Table 4.3, *P*=0.0001). The most divergent populations are AKL and NL. F_{ST} comparison suggests weak but significant differentiation between the WL and NL populations (F_{ST}=0.008, *P*=0.046). Other indices do not support this differentiation even though the *P* values are close to 0.05, ranging from 0.0509 (*Dest*) to 0.0639 (R_{ST}). When the pairwise F_{ST} values at each locus are taken into consideration, the WL and NL populations are significantly differentiated at only two loci (loc *1*, F_{ST}=0.031, *P*=0.003; loc *2*, F_{ST}=0.06, *P*=0.009).

Table 4.3 Pairwise population differentiation estimates based on the microsatellite data. Genic: $\chi 2$ values for the homogeneity of allele frequencies in pairwise comparisons tested with the exact *G*-test, Inf=infinity, d.f.=24. F-statistics and related distances: F_{ST}, *Dest*, G''_{ST} and R_{ST}. *t*: Pseudo-*t* statistic values obtained through pairwise PERMANOVA analysis of Manhattan (*DM*) and Clonal distances (*DCL*) between individuals.

Groups	Genic	F_{ST}	G''_{ST}	*Dest*	R_{ST}	*t-DM*	*t-DCL*
Ti Point-Auckland	17.26	-0.009	-0.027	-0.019	-0.016	0.699	0.766
Ti Point-Tauranga	29.67	-0.004	-0.012	-0.008	-0.016	0.727	0.79
Ti Point-Wellington	Inf[c]	0.087[c]	0.256[c]	0.185[c]	0.123[c]	2.851[c]	2.711[c]
Ti Point-Nelson	Inf[c]	0.118[c]	0.330[c]	0.241[c]	0.211[c]	4.160[c]	4.083[c]
Auckland-Tauranga	20.48	-0.004	-0.011	-0.007	-0.008	0.983	0.655
Auckland-Wellington	Inf[c]	0.095[c]	0.274[c]	0.198[c]	0.143[c]	2.548[c]	2.958[c]
Auckland-Nelson	Inf[c]	0.122[c]	0.337[c]	0.246[c]	0.237[c]	3.677[c]	4.693[c]
Tauranga-Wellington	Inf[c]	0.074[c]	0.216[c]	0.153[c]	0.115[c]	2.541[c]	3.035[c]
Tauranga-Nelson	Inf[c]	0.097[c]	0.272[c]	0.194[c]	0.202[c]	3.725[c]	4.661[c]
Wellington-Nelson	53.20[b]	0.008[a]	0.021	0.013	0.023	1.235[a]	1.593[a]

Significant differentiation: [a] P<0.05, [b] *P*<0.0005, *P*≤0.0001. Population abbreviations within the text are indicated in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL), and Nelson (NL).

Recognising that all these estimators react differently to varying levels of shared and unshared diversity, it was important to use different estimators. Considering that F_{ST} estimates indicate "how close populations are to alternate fixation" (Bird et al., 2011), I can conclude that the process of fixation has begun for alternative alleles in the Southern cluster. The global and pairwise F_{ST} values are systematically low and most probably underestimate the real differentiation due to limitations surrounding the use of

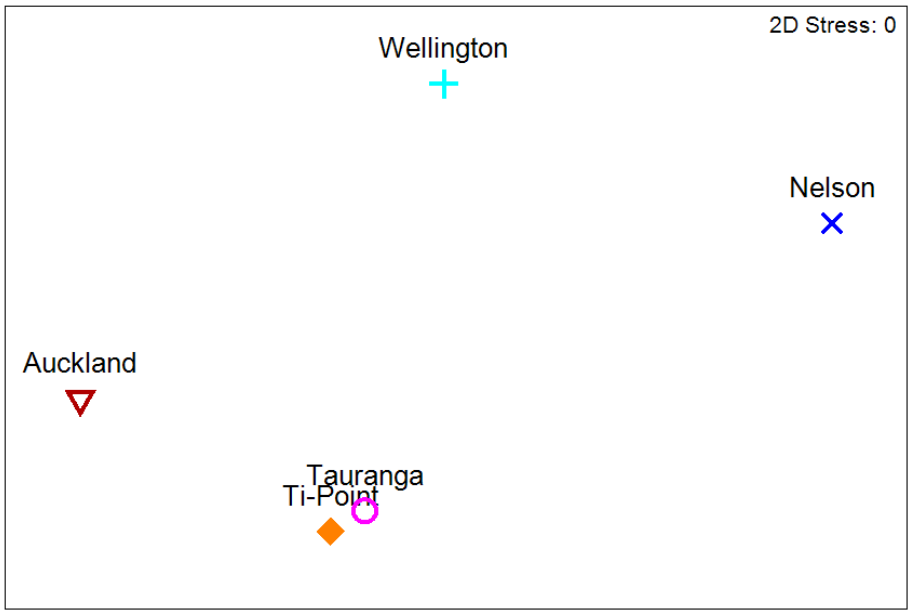


Figure 4.2 MDS graph representing the pairwise R_{ST} distances between the five sampling locations. Abbreviations of the locations within the text are indicated in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL) and Nelson (NL).

multi-allelic markers. The first limitation is that F_{ST} – which was originally proposed for bi-allelic markers – cannot reach its maximum value for multi-allelic markers because fixation of alternate alleles in different populations is very unlikely when multiple alleles are present. Secondly, F_{ST} suffers from high within-population diversity by approaching zero even in those instances where populations do not share any alleles (Bird et al., 2011; Hedrick, 2005; Jost, 2008). As a consequence, three other estimators, G''_{ST}, *Dest* and R_{ST}, which were developed for multi-allelic markers (Bird et al., 2011), were included in this study. These estimators can accommodate high levels of differentiation, and the patterns of differentiation suggested are similar in each instance. The IAM model, which is assumed for R_{ST}, G''_{ST} and *Dest*, treats all alleles as being equally and maximally distant from each other regardless of the number of repeats that they contain (Bird et al., 2011). On the other hand, R_{ST} estimate assumes the SMM model for microsatellite evolution where any mutation results in a gain or loss of one repeat (Slatkin, 1995). This approach makes it possible to incorporate distances between alleles into differentiation calculations, and consequently to establish an evolutionary relationship between alleles. Therefore, it has been suggested that R_{ST} is better at identifying longer historical separations compared to non-distance based indices: F_{ST}, G''_{ST} and *Dest* (Balloux and Lugon-Moulin, 2002). However, in this study, R_{ST} does not improve the resolution of differentiation between the Northern and Southern clusters

compared to G''_{ST} and *Dest*. This finding may imply that the differentiation is too recent to allow fixation of distant alleles carrying very different repeat numbers in the Northern and Southern clusters. In summary, all the differentiation estimators used in this study support the existence of highly significant differentiation between the Northern and Southern clusters.

4.3.3 Analysis of molecular variance

Analysis of molecular variance (AMOVA) (Excoffier et al., 1992) of the 12 loci was performed to determine the hierarchical genetic structure based on a stepwise mutational model (R_{ST}-based). The aim was to investigate the contribution of each hierarchical level to total variation. Two separate calculations were performed: not introducing any structure, grouping the TP, AKL and TR populations under the Northern cluster and grouping the WL and NL populations under the Southern cluster. This grouping design was decided based on the population structure that was suggested by F-statistics and related analyses.

Table 4.4 AMOVA results for the *P. maculata* populations based on the microsatellite data. Analysis was performed with SMM based on R_{ST}-statistics. The hierarchical distribution of variation was determined between the sampling locations without introducing any structure, and nesting the TP-AKL and TR populations in the Northern cluster and the WL and NL populations in the Southern cluster.

	F-stat	Source of variation	Nested in	%var	SS	var	*F*-value	*P*-value
No structure	R_{IT}	Within Individual	-	88.3	2013	51.53	0.117	0.0136 [a]
	R_{IS}	Between Individual	Population	-2.0	6930	-1.16	-0.023	0.7005
	R_{ST}	Between Population	-	13.7	2013	8.02	0.137	0.0000 [b]
Clusters	R_{IT}	Within Individual	-	81.7	7515	51.53	0.183	0.0123 [a]
	R_{IS}	Between Individual	Population	-1.8	6930	-1.16	-0.023	0.7075
	R_{SC}	Between Population	Clusters	0.1	158	0.06	0.001	0.3410
	R_{CT}	Between Clusters	-	20.0	1856	12.63	0.201	0.0000 [b]

Significant contribution to the total variation: a P<0.05, b P<0.0001, var: variance.

The AMOVA results are shown in Table 4.4. Analysis without introducing any structure revealed that the highest proportion of the total variation stems from variation within individuals (88.3%) with a fixation index (R_{IT}) value of 0.117. Variation between the populations (R_{ST}=0.137) constitutes 8.02% of the total variation with high significance (P=0.0000). Variation between individuals within populations does not contribute to genetic variation (-1.16 %) (R_{IS}=-0.020, P=0.700). AMOVA analysis was repeated by

clustering the populations into the Northern and Southern clusters. The results show that the variation between the Northern and Southern clusters explains 20.0% (R_{CT}=0.201) of the total variation with high significance (P=0.0000). The variation between the populations within the clusters explains only 0.1% of the total variation with no significance (R_{SC}=0.001, P=0.3418). These results show that the total variation for the NZ *P. maculata* populations can be explained best by the differences in distribution of variance between the Northern and Southern clusters.

4.3.4 Power analysis

The analysis of statistical power by POWSIM suggests that the microsatellite data are able to detect F_{ST} values as low as 0.005 at least 90 per cent of the time for 1000 replications with both the chi-square and Fisher's exact test. 100% probability of detecting population differentiation was observed at an F_{ST} value of 0.01. The probability of α error is lower than 0.05 with the chi-square approach (P=0.04) and only slightly above with the Fisher method (P=0.057), suggesting a low risk for Type I error. I conclude that the differentiation pattern found in this study is real because the F_{ST} values between the significantly differentiated populations are larger than 0.01.

4.3.5 Individual distance matrices and the multivariate approach

Fixation indices and *Dest* react differently to varying levels of shared and unshared diversity (Bird et al., 2011). Differentiation between the populations was investigated based on distances between individuals to analyse the differentiation pattern at the level of individuals. Multivariate analysis was used for this purpose (Please refer to the heading "Multivariate analysis" in subsection 3.5.2.1 for information on the analysis). The distances between individuals were calculated using two different methods: Manhattan (*DM*) distances (Czekanowski, 1909) and Clonal (*DCL*) distances (Meirmans and Van Tienderen, 2004). The *DM* matrix provides the number of mismatching alleles per locus between individuals. The *DCL* matrix calculates the number of mutation steps needed to transfer the genotype of the first individual to that of the second, assuming an SMM model. Non-metric Multi-Dimensional Scaling (MDS) analysis was used to visualise any dissimilarities between individuals. The resulting 2D graphs of both the *DM* and *DCL* matrices (Figure 4.3A and B, respectively) show that individuals from TP, AKL and TR (the Northern cluster) cluster together, whereas individuals from WL and NL (the Southern cluster) form another cluster. The

single individual collected from CR is located along with the other North Island individuals, whereas the single individual from WH, which is the northern-most of the other sampling locations, was located between the Northern and Southern clusters according to MDS results. One individual from Picton was located in the Southern cluster, as expected. However, a single individual is not sufficient to draw a general idea about the trend at these specific locations.

The results of the MDS analysis of centroids for the *DM* and *DCL* distances are shown in Figure 4.3C and D, respectively. Both measurements exhibit similar pairwise differentiation patterns, even though the order of the distances between the populations differs slightly. The subsequent statistical analysis with PERMANOVA shows (Table 4.5) that the sampling location has a significant effect on population structure for both distance measurements (*DM*: $F_{4, 146}$=7.3914, *P*=0.0001; *DCL*: $F_{4, 146}$=9.8256, *P*=0.0001). The pairwise PERMANOVA analysis between the sampling locations supports the differentiation pattern that is obtained by the F_{ST} and related indices. There is no significant differentiation between the TP, AKL and TR populations (*DM*: *t*=0.655-0.983, *P*=0.51–0.94), whereas each of these populations is significantly different from the WL and NL populations (*t*=2.54–4.69, *P*=0.0001). As the pseudo-t value, which represents the distance between the populations, increases, the support for the null hypothesis of no differentiation between populations weakens. Both increased *t* and significant *P*-values support the differentiation patterns. The *DCL* distances also suggest weak but significant differentiation between WL and NL (*t*=1.59, *P*=0.0108).

Any significant difference found between the populations identified by PERMANOVA could be the result of differences in the location or the dispersion of samples, or a mixture of both. The sample sizes of the populations were variable in this study, which could have led to dispersion differences between the populations. PERMDISP analysis was performed to test the null hypothesis of homogenous dispersion of samples between the populations (Anderson, 2006). The results of the test did not reveal significant differences in dispersion of the variations for either *DM* ($F_{4,141}$=0.1243, *P*=0.1243) or *DCL* ($F_{4,141}$= 0.4856, *P*=0.4856) matrix, meaning that the distances between the individuals are homogeneously distributed between the populations. PERMANOVA results can be clearly attributed to differences in the fragmentation of genetic variation based on the PERMDISP results.

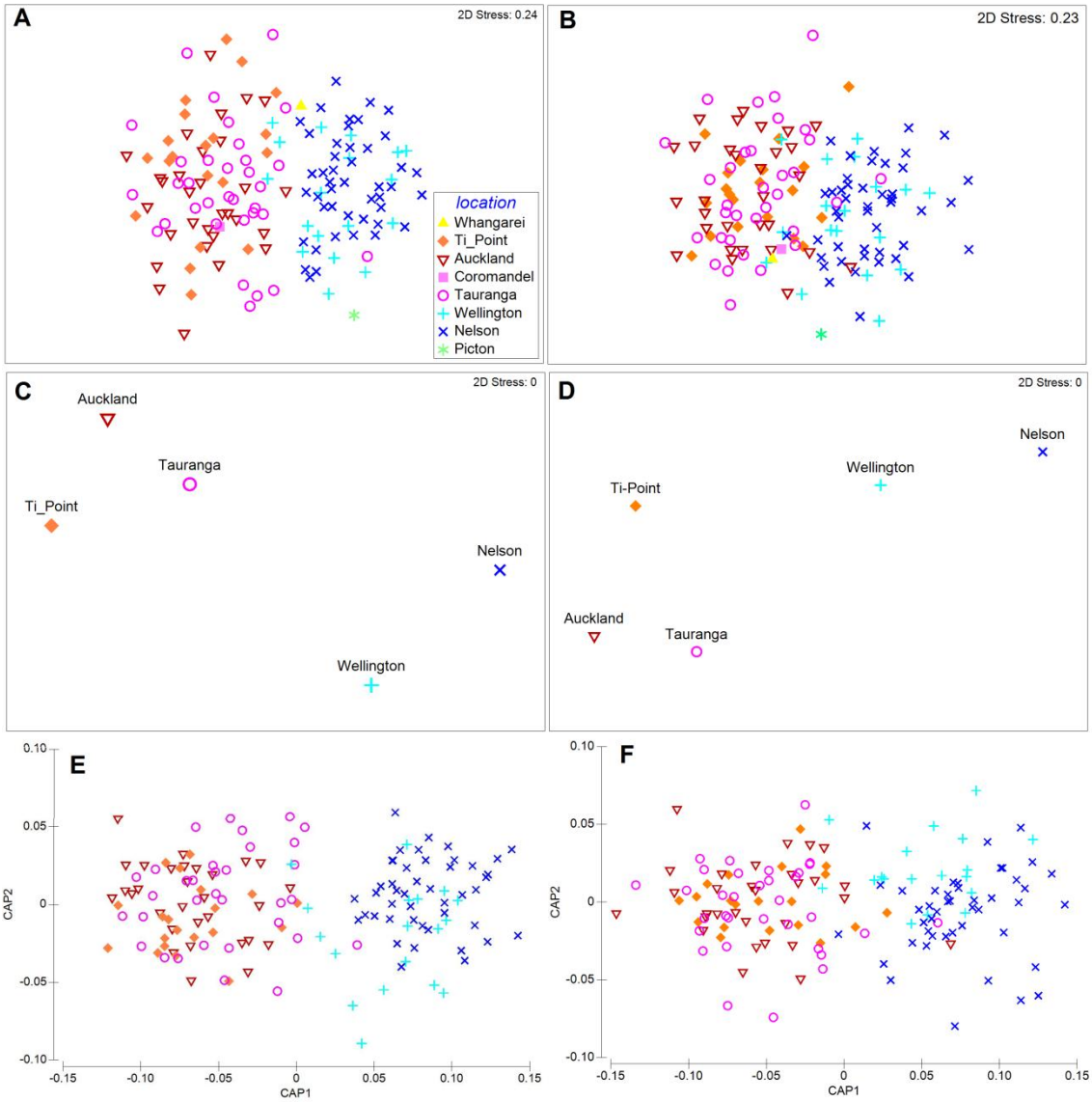


Figure 4.3 Ordination analyses of the multivariate genetic distances between the *P. maculata* individuals obtained from 12 microsatellite markers. 2-D non-metric multi-dimensional scaling (MDS) graph of pairwise A) Manhattan (*DM*), B) Clonal (*DCL*) distances between 149 *P. maculata* individuals from eight sampling locations. MDS output of centroids of the five main sampling locations obtained from C) *DM* and D) *DCL* distances. Output graphs of Canonical Analysis of Principal Component (CAP) analysis of E) *DM* (the number of axes-*m*=13) and F) *DCL* (*m*=6) distances. The axes represent the amount of variation between the populations that is explained by the two axes. Each symbol represents a different sampling location. Population abbreviations within the text are indicated in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL) and Nelson (NL).

MDS analysis examines the data from the perspective of the dimension that contains the highest amount of variation. However, the actual differences between the groups can be hidden under another dimension with less variation. In addition, the stress, i.e. the resolution, of the MDS graphs are slightly too high (Figure 4.3A: 0.23; Figure 4.3B: 0.24) to represent the actual matrix for both *DM* and *DCL* matrices. Therefore,

Canonical Analysis of Principle Coordinates (CAP) was performed to visualise the samples through the axes that maximise the variation between the samples relating them to their pre-defined populations. CAP also evaluates how well the model (sampling sites) explains the observations and how different the groups are by calculating the percentage of correct classification of the samples to their predefined group with the "leave-one-out" procedure. CAP analysis shows that the first 13 and 6 axes (*m*) can explain 80.6% and 72.5% of the total variation in the *DM* and *DCL* matrices, respectively. Inclusion of more axes did not lead to a large gain in allocation success rates; therefore, these *m* values were used for further assessments. CAP analysis yielded four canonical axes to distinguish the populations for both matrices. For *DM*, the CAP1 and CAP2 axes explain 28.2% and 20.0% of the variation between the groups, respectively (Figure 4.3E). The first canonical axis clearly separates the data with a high squared canonical correlation coefficient (δ_1^2=0.83). On the other hand, the second axis has a smaller eigenvalue (δ_2^2=0.11) and the groups are not clearly separated on this axis. The δ^2 varies between -1 and 1, in which -1 means total negative correlation, 0 means no correlation and 1 means total positive correlation.

Table 4.5 Results of the PERMANOVA analysis of the microsatellite data. PERMANOVA tested the effect of the sampling location on genetic distances between 146 - *P. maculata* individuals from five locations. Abbreviations: *DM*=Manhattan distance *DCL*=clonal distance, d.f.=degrees of freedom, SS=sum of squares, MS=mean sum of squares, Pseudo-*F*=*F* value calculated by permutation, *P* (perm)=statistical significance.

Distance Method	***Source***	***df***	***SS***	***MS***	***Pseudo-F***	***P***
DM	Sampling location	4	17.901	4.4751	7.3914	0.0001
	Residuals	141	85.369	0.60545		
	Total	145	103.27			
DCL	Sampling location	4	34233	8558.3	9.8256	0.0001
	Residuals	141	1.23×10^{-5}	871.02		
	Total	145	1.57×10^{-5}			

For the *DCL* matrix, the first two CAP axes explain 41.8% and 28.6% of the variation between the groups (Figure 4.3F). Similar to the *DM* data, the first axis separates the groups with high correlation (δ_1^2=0.73), whereas the second axis does not distinguish them clearly (δ_2^2 = 0.10). Membership of each individual to its sampling location was tested by CAP analysis. Both *DM* and *DCL* show high overall misclassification rates of 56.8% (83 out of 146 individuals) and 58.2% (85 out of 146), respectively. The misclassification rate within the populations varies with different distance

measurements (Table 4.6), and it is mainly high (max 90.9% for the TR population with the *DCL* measurement. For example, with the *DM* distance method, 16 out of 20 individuals from the TP population were assigned to the AKL or TR populations, and 9 out of 18 individuals from the WL population were assigned to the NL population. The highest correct assignment rate is observed for NL (60% and 80%) for both distance measurements. Similar to the output for the F-statistics and related estimators, TP-AKL and TR (the Northern Cluster) are not distinguishable from each other as a general trend, and they form a distinct cluster. Most of the misclassified individuals from this cluster are assigned to another population within the same cluster. Similarly, the WL and NL populations (the Southern cluster) form another distinct group by being indistinguishable from each other. Misclassifications within this cluster mostly occur between WL and NL. Even so, three and six individuals in total from the Northern cluster are assigned to the Southern cluster with the *DM* and *DC* methods, respectively; whereas, three individuals from the Southern cluster in total are assigned to the Northern cluster with both distance measurements. Between-cluster misclassifications were especially noted in the TR and WL populations. Trace statistics (*tr)*, which is the sum of the squared canonical correlations, support a significant differentiation between the centroids of sampling (*DM*: *tr*=1.019, *P*=0.0001; *DM*: *tr*=0.850, *P*=0.0001). The rate of overall misclassification dramatically decreased when CAP analysis was repeated after pooling the individuals according to the Northern and Southern clusters. For example, the total misclassification rates were 2.7% and 4.11% with the *DM* and *DCL* distances, respectively (data not shown), which implies that a two-clustered model explains the population data best. This evidence also shows that there is high genetic connectivity within the clusters, whereas connectivity between the clusters is evident but low.

4.3.6 Bayesian population assignment

Bayesian clustering of individuals based on allele frequencies as implemented by STRUCTURE (Pritchard et al., 2000; Pritchard et al., 2010) shows a ΔK value and mean log probabilities of data (Ln*P* (x/*K*) that are maximal at *K*=2 (based on ten replicate simulations for each *K* value (*K*=1 to 10) (Figure 4.4A). This finding was not affected by including sampling locations as priors (Supplementary Figure 2). These results further support the presence of two distinct clusters. As the F-statistic and related estimators as well as multivariate analyses previously suggested, STRUCTURE assigns

individuals from the North Island, except WL, (the Northern cluster) to one population, whereas WL and the South Island individuals (the Southern cluster) are assigned to the other (Figure 4.4B).

Table 4.6 Assignment of 146 individuals to the a priori groups by CAP analysis using the microsatellite data. Two different distance methods; *DM*: Manhattan distances and *DC*L: Clonal distance specific to microsatellites. Five groups, which were also the main sampling locations, were introduced as a priori groups.

Distance Method	Original group	Classified					Correct in total	%correct
		TP	AKL	TR	WL	NL		
DM	Ti-Point	10	11	9	0	0	10/20	33.333
	Auckland	6	5	9	0	0	5/30	25
	Tauranga	8	8	14	3	0	14/33	42.424
	Wellington	0	0	3	7	8	7/18	38.889
	Nelson	0	0	0	18	27	27/45	60
DCL	Ti-Point	3	10	5	2	0	3/20	15
	Auckland	7	11	10	1	1	11/30	36.667
	Tauranga	14	14	3	1	1	3/33	9.091
	Wellington	2	0	0	8	8	8/18	44.444
	Nelson	1	0	0	8	36	36/45	80

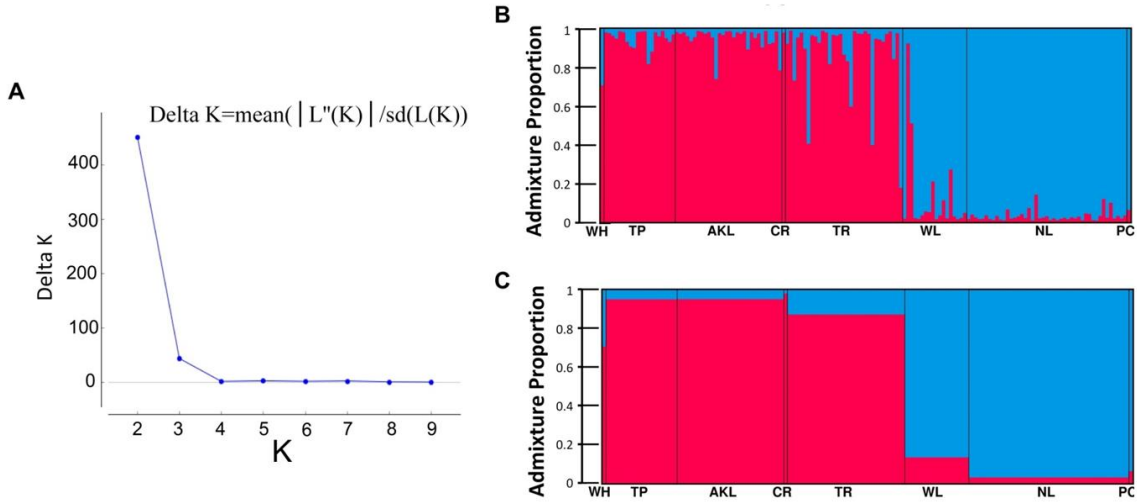


Figure 4.4 Results of Bayesian clustering analysis where the sampling location was introduced for the calculations. A) Plot of ΔK versus K indicating that the data is explained best by clusters K=2. B) Population structure in 149 individuals of *P. maculata* from 8 different sampling sites at K=2. Each individual is represented by a vertical line divided into two segments, which indicate proportional membership in the two clusters. C) Group assignments, indicating proportional membership in K=2 clusters. Population abbreviations in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL) and Nelson (NL).

Inclusion of sampling locations in the calculations increases the admixture proportions for individuals, especially the ones from the TR and WL populations (Supplementary Figure 2B and C). Four individuals from TR have a probability higher than 0.5, which assigns them to the Southern cluster. Similarly, two individuals from WL have a probability higher than 0.5 of being from the Northern cluster.

4.3.7 Migration

Migration analysis with GeneClass2 using Lhome/Lmax detected four first generation migrants (P=0.01); one individual sampled from TP is a migrant from AKL, one individual sampled from TR is from AKL, and two individuals sampled from NL are migrants from WL (Figure 4.5). These migration events happen within the clusters defined by the F_{ST} and related estimators, STRUCTURE and CAP analyses. According to the GeneClass2 analysis, generally, individuals from each cluster had a relatively higher probability of belonging to the populations from the same cluster (Supplementary Table 8). A lack of first generation migrants between the clusters shows that these clusters are genetically not well connected.

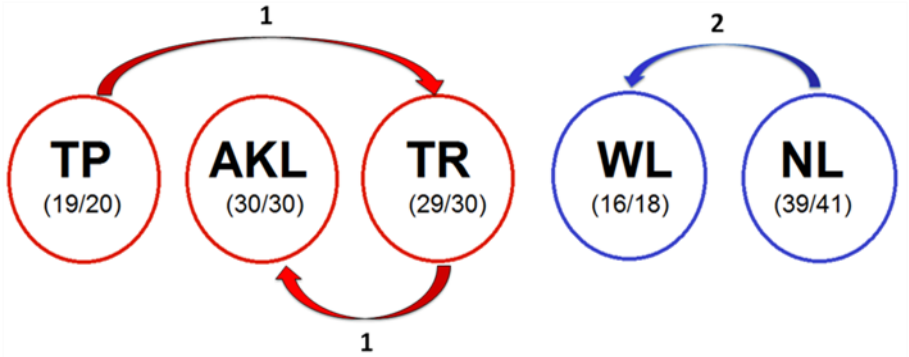


Figure 4.5 First-generation migrants in the *P. maculata* individuals from five sampling locations. *P. maculata* inferred through GeneClass2 analysis). The circles indicate the sampling locations; the numbers in the bracelets indicate the proportion of individuals that were assigned to their sampling locations. The red and blue colour corresponds to the two populations identified by STRUCTURE and other differentiation analysis. Arrows indicate the inferred origin, and the numbers above indicate the number of migrants.

GeneClass2 identifies first generation migrants, i.e., individuals that were not born in the population from which they were sampled (Paetkau et al., 2004). Population assignment performed by CAP and STRUCTURE may also be used to identify the level of gene flow between different populations. In this study, both approaches show that populations within the clusters are highly connected, and that different populations are not easily distinguishable from each other (Figure 4.3E and F, Table 4.6, Figure 4.4B

and C). On the other hand, connectivity between the clusters is low, but admixed individuals are still observed in both the Northern and Southern clusters. The highest admixture levels are in the TR and WL populations. For example, STRUCTURE analysis suggests that four individuals from TR have a probability higher than 0.5 of belonging to the Southern cluster. Similarly, two individuals from WL have a probability higher than 0.5 of belonging to the Northern cluster (Figure 4.4B and C). In addition, the highest misclassifications between the clusters detected by CAP analysis were observed in TR and WL (Table 4.6). This evidence suggests that the TR and WL populations are the bridges between the clusters.

4.3.8 Spatial analysis

The isolation by distance model was tested by the Mantel test (1967) which calculates the correlation between pairwise genetic and geographical distances. The test revealed that the pairwise F_{ST} values are significantly correlated with the geographical distance between the populations (r=0.992, P=0.0088, Figure 4.6).

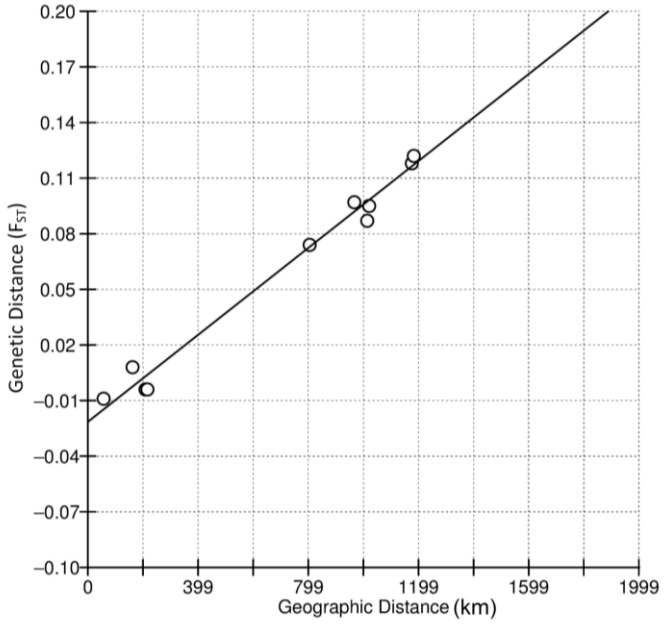


Figure 4.6 Test of isolation by distance, based on the microsatellite data. Relationship between the pairwise genetic distances (F_{ST}) and geographical distances (km). Significant isolation by distance is evident (r=0.9882, P=0.0088).

4.3.9 Demographic analysis

Population expansion for the five main populations was tested with two tests: *k* and *g* tests. Alleles are expected to be distributed into a few distinct groups in constant populations, whereas distinct allele groups are not expected in expanding populations

(Reich et al., 1999).The *kg*-tests are based on these assumptions (Luikart and Cornuet, 1998). The *k*-test is a within-locus test, and it assumes that the pattern of distribution of allele-lengths follows a bimodal shape in constant populations, while this distribution is uni-modal and peaked in expanding populations. The *k* statistic, which measures the peakedness of the distribution, tends to acquire negative values in expanding populations (Reich et al., 1999). The proportion of loci that provides negative *k* values is evaluated with a binomial distribution (Bilgin, 2007). For the *P. maculata* dataset, *k*-tests suggested negative values at the majority of loci ranging from eight to ten out of the twelve loci when the populations were pooled into a single population. This can be a sign of an expansion. However, the test returned significant results for only the TP population (P=0.0156) when the analysis was performed for each population separately. No significance was observed for the other populations (P>0.0596). The *g*-test compares the observed and expected variance of allele length distribution across loci assuming that the variance of the widths of the distribution will be lower in a recently expanding population. The *g* value is the ratio of the observed to the expected variation under assumptions of constant population size. The significance of the *g* value is determined according to theoretical 5% percentile cut-off values for a given sample size and loci (Reich et al., 1999). The *g* values calculated in this study do not support a history involving population expansion for the *P. maculata* data. It is worth noting that both tests, but especially the *g*-test, have a low power to detect population expansion with a small number of samples and loci (Bilgin, 2007).

In a bottlenecked population, the number of alleles decreases more dramatically than heterozygosity (Cornuet and Luikart, 1996). Consequently, the heterozygosity calculated from allele frequencies (H_e) becomes higher than the heterozygosity calculated from allele number (H_{eq}) in a bottlenecked population (Cornuet and Luikart, 1996). In this concept, BOTTLENECK (Piry et al., 1999) was used to test whether a significant number of loci show heterozygosity excess ($H_{e>}H_{eq}$) (Luikart and Cornuet, 1998) in order to investigate the possibility of recent population reduction in each population using three different mutational models: IAM, SMM and the two-phase model (TPM), which is a mixture of IAM and SMM. The Wilcoxon test that was used to evaluate the significance of the heterozygosity excess did not detect recent bottlenecks in any population under the TPM and SMM models (Table 4.7). However, the TR population was found to have experienced recent bottleneck (P=0.0031) under the IAM

model. There is no information available on the mutation model of *P. maculata* microsatellites; however, IAM seems to be inappropriate for microsatellite evolution based on empirical data. TPM and SMM, which take size homoplasy into account, are said to provide a more realistic explanation of allele evolution in microsatellites (Bhargava and Fuentes, 2010). Therefore, the bottleneck in TP suggested by IAM may not be realistic. When the distribution of allele frequencies is taken into consideration, nonbottlenecked populations that are under mutation-drift equilibrium are expected to show an L-shaped pattern in which there are high proportions of low frequency alleles. On the other hand, a shift is observed in the mode of allele frequency distribution in bottlenecked populations where intermediate frequencies become more abundant than low frequency alleles (Luikart and Cornuet, 1998). Analysis of mode-shift in the distribution of allele frequencies for the *P. maculata* dataset with BOTTLENECK suggests that all the populations exhibit a normal L-shaped pattern. All this evidence suggests that none of the populations is likely to have experienced a recent bottleneck.

Table 4.7 Summary results of population expansion and decline analysis for the NZ *P. maculata* populations based on microsatellite data. Expansion and decline were assessed with kg-tests and BOTTLENECK, respectively, using data obtained from the 12 microsatellite loci.

	Ti Point (TP)	Auckland (AKL)	Tauranga (TR)	Wellington (WL)	Nelson (NL)
***k*-test (# of negative loci)**	10	8	9	9	9
***k*-test (P value)**	0.0156	0.1661	0.0596	0.0596	0. 0596
***g*-values**	1.229	1.5642	1.7533	2.0630	3.7239
IAM	0.0756	0.1167	0.0031	0.0881	0.3386
TPM	0.7407	0.7153	0.6890	0.6333	0.9539
SMM	0.9451	0.9933	0.9948	0.0523	0.9994

4.4 DISCUSSION

4.4.1 Genetic diversity

In this study, I investigated the diversity and genetic structure of 146 samples from five main localities representing four regions of the North Island (Ti-Point, Auckland, Tauranga and Wellington) and one region of the South Island (Nelson) using twelve nuclear microsatellite markers. High genetic diversity was observed for all five populations. All the populations are in HWE. Allelic richness and heterozygosity, which are two important diversity estimators, are high for all five populations, indicating that all the populations contain significant amounts of genetic diversity. According to

population genetics theory, there is positive correlation between effective population size and expected genetic diversity at a neutral locus that is under mutation-drift equilibrium. This is due to the fact that the effect of genetic drift is less dramatic in large populations compared to small ones (Hartl and Clark, 2007). High diversity in marine invertebrates is attributable to their large population size (Zhan et al., 2009), which may explain the high diversity observed in *P. maculata*. There are other factors affecting genetic diversity, such as the geographical range of the organism (Frankham, 1996), selection and the mating system (Bazin et al., 2006). Selection affects variation in an increasing or decreasing manner depending on the nature of the selection. However, microsatellites are assumed to be neutral (Oliveira et al., 2006). The possibility that they are hitchhiking with the regions under selection cannot be ruled out. However, the microsatellites used in this study are in linkage equilibrium as revealed by the LD test. All these unlinked loci are unlikely to be under the same type of selection. *P. maculata* is an outcrossing simultaneous hermaphrodite (Willan, 1983). This mating type increases the effective population size, and consequently the genetic diversity (Silva and Russo, 2000). Genetic variation is expected to be high in species with wide geographical ranges (Frankham, 1996). This effect may also contribute to genetic variation in *P. maculata* as it has a wide range of geographical distribution in the South-Eastern Pacific (Willan, 1983).

4.4.2 Population differentiation

All population differentiation estimators, including the exact differentiation tests based on allele frequencies, fixation index-related measures, and multivariate analysis based on pairwise individual distances consistently show highly significant differentiation between the populations from the different sampling locations. Differentiation is evident at the majority of the loci. The data show the existence of two clusters: the northern North Island samples from TP, AKL and TR (the Northern cluster) are significantly differentiated from WL and the South Island population from NL (the Southern cluster). The Southern cluster has a significantly higher level of private allelic richness (PA_R) compared to the populations sampled from the Northern locations. Their high PA_R content also shows that the Southern populations are genetically differentiated from the Northern populations. Weak differentiation between WL and NL was observed with some of the differentiation tests. AMOVA testing shows that differentiation between populations within the Northern and Southern clusters does not contribute to the total

variation, which shows that there is no further structure within the clusters. Population assignment tests support this conclusion, but they also suggest the possibility of some admixture within the sampling locations. TR and WL have the highest level of average admixing proportions, and individuals assigned between the Northern and Southern clusters are mainly from these locations, although the estimated level of gene flow was dependent on the particular distance method used for analysis. The admixture in the WL population may explain the weak differentiation between the WL and NL populations. First generation migrant analysis does not support the existence of any migrants between the Northern and Southern clusters.

4.4.3 Biogeographical barriers

The north-south differentiation observed in the NZ *P. maculata* populations is not unique, as it has been observed in many other marine organisms around NZ. However, the location of the disjunction varies in different organisms (Gardner et al., 2010). Studies with a reasonably detailed sampling regime capable of locating the disjunction have shown that NZ marine species such as the green-lipped mussel *Perna canaliculus* (Apte and Gardner, 2002; Wei et al., 2013a) and the sea urchin *Evechinus chloroticus* (Nagel et al., 2015) exhibit north-south differentiation at 42°S along the southern edge of Cook Strait (Ross et al., 2009), where upwelling regimes (Chapter 1,Figure 1.4) are thought to form a barrier to gene flow between the North Island and northern coast of the South Island, and the south of the South Island (Veale and Lavery, 2011). However, the divergence for *P. maculata* may correlate with the biogeographical barrier on the east coast of the North Island between 37–39°S, originating from the East Cape current and the Wairapapa Eddy (Chapter 1, Figure 1.4), (Ross et al., 2012; Stevens and Hogg, 2004; Will et al., 2011). Such a pattern has been observed for some organisms such as the amphipods *Paracorophium excavatum* and *P. lucasi* (Stevens and Hogg, 2004) and the gastropod *Diloma subrostrata* (Donald et al., 2005). The disjunction can also be explained by the fact that the North and the South Islands were connected during the Pleistocene glacial cycles including the last glacial maxima (LGM, 26.5–18 kya) as the sea levels dropped as low as 120 m along NZ coasts until the recent rise in sea levels separated the islands and created the present coastline (12–6 kya, Keeney et al., 2013; Lewis et al., 1994; Trewick and Bland, 2012). This land connectivity has lead to east-west disjunction and connectivity of the southern/western North Island and the northern South Island populations for some NZ marine organisms such as the greenshell mussel

Perna canaliculus (Star et al., 2003), 2003) and the amphipods *Paracorophium excavatum* and *P. lucasi* (Stevens and Hogg, 2004). Weak differentiation between WL and NL can perhaps be explained by the body of water separating the North and the South Islands, the Cook Strait, although there is not enough evidence supporting barrier effect of the Strait (Ross et al., 2009).

The highly connected WL and NL populations might be due to the D'Urville Current between the North and South Islands, which flows from the west into Cook Strait (Heath, 1982). In addition, assignment tests show that there is no direct migration between the NL and the North Island populations, whereas some limited admixture was observed between the WL and the North Island populations. This gene flow from the North Island to the WL population might explain the weak differentiation between the WL and the NL populations.

Although the structure pattern can be explained by biogeographical barriers, it can be explained also with an isolation by distance model (IBD). IBD assumes the stepping-stone model where the genetic distance between the populations increases with geographical distance and gene flow occurs only between neighbouring regions (Kimura, 1953; Kimura and Weiss, 1964). In this model, there is no physical barrier between populations, and predictions are based on the steps between pairs of populations (Kimura and Weiss, 1964). For *P. maculata*, the geographical regions between the Northern and Southern clusters could not be included in the sampling process. If the missing populations have ongoing gene flow, the missing data might have led to an incorrect conclusion of distinct populations. Moreover, the limited admixture observed between the WL and the North Island populations may support an IBD model. Therefore, the geographical gap between the sampling locations made it impossible to draw firm conclusions as to the origin of the disjunction, whether it is solely due to biogeographical barriers.

Wei et al. (2013a) identified a north-south discontinuity at 42°S for the NZ greenshell mussel, *Perna canaliculus*. The same group (Wei et al., 2013b) performed seascape analysis on various environmental and geospatial variables, including the amount of solar radiation, sea surface temperatures, orbital water velocity, tidal currents and freshwater fraction, to assess whether these variables correlate with or explain the genetic variation between the populations. They found that genetic divergences between

the populations are best explained by local adaptation to varying sea surface temperatures. The possibility of temperature-dependant local adaptation for *P. maculata* is yet to be clarified.

4.4.4 Pelagic larval duration

Egg type is a significant factor affecting the genetic connectivity of populations of marine organisms. Species with benthic eggs that hatch into adults consistently have more structured populations than ones with pelagic eggs (Riginos et al., 2011; Weersing and Toonen, 2009). Additionally, a general inverse relationship between pelagic larval duration and genetic structure has been suggested by some studies (Ross et al., 2009; Shanks, 2009), although the extent of genetic differentiation varies between species with similar PLD depending on larval behaviour, the size of the adults, post-settlement dispersal and biogeographical barriers (Bradbury et al., 2008; Ross et al., 2009). Therefore, PLD cannot directly explain genetic differentiation patterns (Riginos et al., 2011; Weersing and Toonen, 2009).

Pleurobranchaea maculata has pelagic eggs, and settlement of *P. maculata* larvae on bio-filmed surfaces as juveniles takes approximately 3 weeks (Gibson, 2003; Wood et al., 2012b) in a laboratory environment. In their comparative analysis of NZ pelagic marine species, Ross et al. (2009) demonstrated a significant negative correlation between PLD and genetic differentiation (r^2=0.39, P<0.001). When NZ-wide sampling regimes are considered, NZ organisms with PLD durations similar to *P. maculata* (2–4 weeks) exhibit various structural patterns including no structure, north-south disjunction, IBD or differentiation within and between sampling locations (Ross et al., 2009). My PhD study revealed that *P. maculata* exhibits a structure related to some biogeographical breaks in NZ, or an IBD pattern. As the origin of the structure is vague for *P. maculata*, it was not possible to interpret the effect of the egg type of the species on its population structure.

4.4.5 Correlation with toxicity

One interpretation of the data suggests that the population genetic structure of *P. maculata* correlates with the regional variations in TTX concentration: the northern North Island populations contain highly toxic individuals and are significantly different from the WL and NL populations, which harbour either slightly toxic or non-toxic populations, respectively. In addition, weak differentiation was observed between the

WL and NL populations. Population structure and migration analysis based on allozymes, mtDNA (Kuchta and Tan, 2005) and microsatellites (Ridenhour et al., 2007) for the TTX-bearing *Taricha granulosa* newt from various localities in western North America shows little differentiation and high gene flow between populations that exhibit large phenotypic variations. Similarly, phylogenetic analysis based on mtDNA markers for the TTX-containing red-spotted newt *Notophthalmus viridescens* shows that highly toxic and non-toxic populations (Yotsu-Yamashita et al., 2012) are genetically identical, and the authors suggested an exogenous source for TTX. *P. maculata*'s population structure correlating with TTX concentrations does not resemble these examples.

The correlation with toxicity and population genetic structure in *P. maculata* could be interpreted as indicative of a genetic mechanism underlying the phenotypic variation. However, the capacity of *P. maculata* individuals from non-toxic regions to accumulate high concentrations of TTX (Khor et al., 2014; Wood et al., 2012a) and rapid depuration of TTX when the individuals are fed on non-toxic diet (Wood et al., 2012a) suggests an external TTX source. This implies that geographical variations in toxicity are more likely associated with environmental differences at distinct geographical locations. One possible explanation is the geographical variations in the distribution and/or abundance of *P. maculata*'s toxic food sources. For example, a flatworm, *Stylochoplana* species, which is also a prey for *P. maculata* in the Tauranga region, was also found to contain high concentrations of TTX (Salvitti et al., 2015a). The toxicity level between individuals of this species also shows significant differentiation as found in *P. maculata*. Seasonal variations in TTX concentrations in *Stylochoplana* sp accord with those of *P. maculata*, and mass calculations show that *P. maculata* might accumulate the observed TTX concentrations by preying on *Stylochoplana* sp. However, this *Stylochoplana* sp. has not been found in geographical areas where other toxic *P. maculata* individuals are observed, implying that *Stylochoplana* sp. is not the only TTX source for *P. maculata* (Salvitti et al., 2015a). Additionally, no other common source that has enough TTX concentration has been identified in other regions associated with toxic *P. maculata* individuals. However, the possibility that other toxic dietary resources have not been discovered still exists (Khor et al., 2014).

Another possibility – if the ultimate source of TTX is biosynthesis by one or more commensal microorganisms – is that certain environmental factors may affect the differential distribution of the commensal or symbiont (Wood et al., 2012b). Ultimately, a bacterial origin of TTX may explain the presence of TTX in other organisms such as *Stylochoplana* sp. The population structure of *P. maculata* revealed by microsatellites seemingly correlates with geographical barriers that result in the north-south disjunction although my limited sampling did not allow me to study the possible effects of the seascape features around NZ on the structure of *P. maculata* populations. If this is the case, the geographical barriers may also affect the distribution of such microorganisms, which would indirectly result in a correlation with toxicity and population genetic structure in *P. maculata.*

Episodes of regionally confined TTX poisoning such as dog poisoning events that occurred on Auckland beaches in late 2009 (McNabb et al., 2010) are not unique to NZ. In 2007, TTX-poisoning events occurred due to consumption of *Charonia lampas*, which is a gastropod species native to Mediterranean and Atlantic waters (Rodriguez et al., 2008). Silva et al. (2012) investigated TTX levels in several gastropod, bivalve and starfish species sampled from the Portuguese coasts between July 2009 and November 2010. They detected low concentrations of TTX or its derivatives in two native edible gastropod species, *Gibbula umbilicalis* and *Monodonta lineata*. Based on the identification of putatively TTX-producing *Vibrio* species in puffer fish only at warm temperatures (20–29˚C) (Sugita et al., 1989), Silva et al. (2012) argued that increased sea temperatures in the Mediterranean Sea may facilitate the growth of TTX-producing bacteria, and result in TTX-production in native gastropod species (Silva et al., 2012). Differences in sea surface temperatures between the North and South Islands might have also led to the clear toxic and non-toxic population cut-off between the islands (Wood et al., 2012b).

Further studies on additional populations of *P. maculata* will help to reveal the exact location of the north-south population differentiation, the existence of other differentiated populations (if any), and various additional factors shaping population structure. Even though I found a correlation between population differentiation and TTX concentrations, my data are unable to say anything conclusive about the origin of TTX in *P. maculata*. It is important to note that the limited sampling might have also

given an impression of the correlation although it does not exist. Studies that aimed to identify TTX-producing microorganisms have not found any evidence supporting a culturable bacterial origin of TTX in *P. maculata* (Chau et al., 2013; Salvitti et al., 2015b). On the other hand, there is strong evidence for a dietary origin of TTX for at least some regions in NZ (Khor et al., 2014; Salvitti et al., 2015c). The ultimate source of TTX in *P. maculata* as well as other TTX-containing organisms is most probably biosynthesis by one or more commensal microorganisms that are not culturable or that cannot produce TTX in-vitro (Chau et al., 2013; Salvitti et al., 2015c). As a further step, the metagenomic approach may reveal *P. maculata*'s associated microflora, and accordingly unculturable microorganism taxa. Additionally, the metagenomic approach may help to identify putative gene clusters that could be involved in biosynthesis of TTX (Gerwick and Moore, 2012).

CHAPTER 5: DIVERSITY AND STRUCTURE OF *P. MACULATA* POPULATIONS AS REVEALED BY MITOCHONDRIAL DATA

5.1 INTRODUCTION

Mitochondrial DNA (mtDNA) is a double-stranded DNA molecule found in eukaryotes. It is generally circular in multi-cellular organisms (Nosek et al., 1998). It has many copies within a cell, and can comprise 25% of the volume of cytoplasm (Ballard and Whitlock, 2004). The sea slug (Gastropoda: Heterobranchia) mtDNA genome, which is approximately 13–15 Mb in size, is compact with a limited number of non-coding sequences. It contains 13 protein-encoding, 22 tRNA-encoding and two ribosomal RNA genes as is characteristically seen in other Metazoans (Medina et al., 2011).

mtDNA markers are the most popular markers in population genetic and phylogenetic analysis of closely related taxa for several reasons, including uni-parental (mostly maternal) inheritance, rare recombination events, simple structure, high copy number and high mutation rates (Funk and Omland, 2003). In addition, sequence data, including whole mtDNA genomes, are available for a wide variety of taxa, which facilitates primer design (or cross-amplification of published primers for non-model organisms). However, analysis based on mtDNA markers alone may not represent the structure and history of the population correctly, as all the genes in the mitochondrial genome evolve as a single linkage unit, which makes them a single locus marker. Therefore, it is generally preferable to use mtDNA markers along with multiple unlinked nuclear DNA (nuDNA) markers to get a better representation of the population structure of the species in question (Zink and Barrowclough, 2008). Thanks to the advances in sequencing and genotyping technologies, the number of studies where diverse sets of markers, including nuclear nuDNA and cytoplasmic DNA markers such as mtDNA and chloroplast DNA, are used has increased (Toews and Brelsford, 2012).

The biogeographical patterns revealed by different marker types are usually congruent with each other. However, several phylogenetic or population genetic studies have observed discordant patterns between mtDNA and nuDNA, and the number of such observations is growing as the number of studies using both marker types increases

(Toews and Brelsford, 2012).

Mitonuclear discordances are possible as the responses of nuDNA and mtDNA to historical and recent evolutionary processes may differ. A common reason for discordance is incomplete linkage sorting (ILS). For two groups that have recently diverged, gene trees are polyphyletic where haplotypes are randomly dispersed between the groups without creating phylogenetic congruence. After sufficient time has passed, lineage sorting is complete and members of isolated groups are represented as monophyletic groups (Funk and Omland, 2003). mtDNA has an approximately four-fold smaller effective population size (N_e) than that of nuclear genes since it is haploid and uni-parentally (mostly maternally) inherited. As the coalescence time when the last common ancestor existed is directly proportional to N_e, these two marker types provide information about demographic and population structure time over different periods. Despite their high mutation rates, nuclear DNA, including microsatellite markers, offer information about a more distant time than that of mtDNA markers (Zink and Barrowclough, 2008). Therefore, ILS can be responsible for cases where mitonuclear discordance is observed. Mitonuclear discordances may also be a result of hybridisation. Populations isolated for a long period may come into secondary contact, or isolated populations might experience past contact. Hybridisation of divergent groups that have accumulated different mutations during isolation will create discordance, as the level of gene flow will be different for the two genomes. Under these circumstances, nuDNA markers may be exchanged at a higher frequency, but mtDNA is more likely to protect its past signature (Rohwer et al., 2001; Toews and Brelsford, 2012).

In the previous chapter, analysis of 12 unlinked nuclear microsatellite data showed that *Pleurobranchaea maculata* populations in New Zealand exhibit a north-south differentiation. In addition to microsatellite data, two mtDNA genes, cytochrome c oxidase I (COI) and cytochrome B (Cytb) were analysed in order to examine the pattern of genetic variation and demographical changes in the *P. maculata* populations in NZ. This chapter presents the mtDNA results, and discusses the results combined with the ones obtained from the microsatellite data in order to understand the population history of the species comprehensively, as both marker types offer unique insights into population processes.

5.2 RESULTS

5.2.1 Identification of polymorphic regions

Four hundred forty-five, 1455 and 1135 bp of 16S rRNA, COI and Cytb genes, respectively, were sequenced for seven individuals. No variation was observed for the 16S rRNA sequences, but 14 and 9 sites were polymorphic across individuals in COI and Cytb, respectively. The 16S rRNA gene was not utilised for further studies, whereas polymorphic portions of the COI and Cytb regions were amplified across all available 156 DNA samples: 5 from Whangarei (WH), 20 from Ti Point (TP), 30 from Auckland (AKL), 3 from Coromandel (CR), 33 from Tauranga (TR), 18 from Wellington (WL), 45 from Nelson, and 1 each from Picton (PC) and Kaikoura (KK). Individuals from WH, CR, PC and KK were excluded from population-wise analysis due to their limited sample sizes.

5.2.2 Genetic diversity

A total of 2213 bp of mtDNA (COI: 1153 bp; Cytb: 1060 bp) was sequenced for all 156 individuals from nine localities. Sequence data for one individual are presented in Supplementary Table 9. One hundred and seventy three variable sites (COI: 105; Cytb: 68), 98 of which are singleton mutations (COI: 59; Cytb: 39) resulting in 130 distinct haplotypes (COI: 103; Cytb: 74; haplotype sequences for variable sites in Supplementary Table 1) were observed. The basic diversity values are presented in Table 5.1. High haplotype diversity (*Hd*), i.e. the average number of nucleotide differences between a pair of sequences divided by the total length of the sequences (Nei, 1987), was observed across all the samples (Total: 0.993, COI: 0.980, Cytb: 0.945). In contrast to high haplotype diversity, the numbers of base changes between the sequences are low, suggesting low nucleotide diversity (π). The overall mean numbers of pairwise differences (k) and π are 7.30 (COI: 3.81, Cytb: 2.75) and 0.330 % (COI: 0.381 %, Cytb: 0.275 %), respectively. The maximum k and π differences observed between the concatenated sequences are 19 bp (COI: 13 bp; Cytb: 10 bp) and 1.127% (COI: 0.9434%; Cytb: 0.8586%), respectively. A similar amount of diversity was observed between the sampling locations. The *Hd* in the sampling locations is also high, ranging from 0.983 to 1.000 for concatenated sequences (0.957–1.000 for COI, 0.922–1.000 for Cytb). Once more, nucleotide diversity is low within the populations; k and π

range from 6.745 to 7.895 (COI: 4.018–4.779; Cytb: 2.687–3.359) and from 0.305 to 0.357% (COI: 0.348–0.512%, Cytb: 0.226–0.317, respectively.

Table 5.1 Summary genetic diversity statistics for the mtDNA data. Statistics are presented only for locations with ≥5 individuals, although the total values were calculated for 156 *P. maculata* individuals sampled from nine localities.

	POP	*N*	*S*	*Sin*	*Hap*	*Hd* ± SD	*k*	*π*± SD %
COI	Whangarei (WH)	5	14	1	5	1 ± 0.126	5.9	0.512 ± 0.077
	Ti Point (TP)	20	27	16	17	0.984 ± 0.020	4.779	0.414 ± 0.048
	Auckland (AKL)	30	39	31	26	0.989 ± 0.013	4.057	0.352 ± 0.033
	Tauranga (TR)	33	39	27	27	0.981 ± 0.015	4.492	0.390 ± 0.045
	Wellington (WL)	18	25	17	16	0.987 ± 0.023	4.19	0.363 ± 0.033
	Nelson (NL)	45	44	33	30	0.957 ± 0.018	4.018	0.348 ± 0.035
	Total	156	105	59	103	0.980 ± 0.0053	4.383	0.381 ± 0.018
Cytb	Whangarei (WH)	5	6	6	5	1 ± 0.016	2.4	0.226 ± 0.050
	Ti Point (TP)	20	17	9	16	0.957 ± 0.021	3.116	0.294 ± 0.041
	Auckland (AKL)	30	23	15	18	0.922 ± 0.034	2.687	0.254 ± 0.043
	Tauranga (TR)	33	29	20	21	0.936 ± 0.032	2.792	0.263 ± 0.041
	Wellington (WL)	18	18	10	13	0.954 ± 0.034	3.359	0.317 ± 0.046
	Nelson (NL)	45	35	23	29	0.941 ± 0.063	2.921	0.276 ± 0.032
	Total	156	68	39	74	0.945 ± 0.011	2.912	0.275 ± 0.018
CONCATENATED	Whangarei (WH)	5	20	20	5	1.000 ± 0.126	8.3	0.375 ± 0.056
	Ti Point (TP)	20	44	25	20	1.000 ± 0.016	7.895	0.357 ± 0.035
	Auckland (AKL)	30	62	46	28	0.993 ± 0.012	6.745	0.305 ± 0.031
	Tauranga (TR)	33	68	47	31	0.996 ± 0.009	7.284	0.329 ± 0.030
	Wellington (WL)	18	43	27	17	0.993 ± 0.021	7.549	0.341 ± 0.028
	Nelson (NL)	45	79	56	39	0.987 ± 0.011	6.939	0.314 ± 0.023
	Total	156	173	98	130	0.993 ± 0.032	7.302	0.330 ± 0.013

N: sample size, *S*: number of segregating (variable) sites, *Sin*: singleton mutations, *Hap*: number of haplotypes detected, *Hd*: haplotype diversity, *k*: number of pairwise nucleotide differences, *π*: nucleotide diversity, SD: standard deviation.

As the sample size varied across the populations and high haplotype diversity was observed, rarefaction analysis was performed for the TP, AKL, TR, WL and NL populations to estimate whether a reasonable amount of diversity had been sampled, or not. For this purpose, the standardised bp distances between individuals were calculated for the Cytb sequences by dividing the total number of bp mismatches between two individuals by the total number of sites that were compared. The total number of operational taxonomic units (OTUs) that is used to cluster individuals was calculated either by clustering individuals with identical sequences, or by clustering individuals that varied by a distance of maximum 0.0030. This cut-off value was chosen to remove the rare sequences whose distance from each other was at most ~1/3 of the maximum distance (0.009434) observed between individuals. Then, using the OTU number as a diversity index, rarefaction analysis was performed for the TP, AKL, TR, WL and NL populations to measure intra-sample diversity in a standardised way so as to compare

the diversity across the populations. The rarefaction curves for the Cytb sequences show that diversity increases for all the populations with the inclusion of additional samples, and that diversity saturation does not occur for any population (Figure 5.1A). Sampling saturation is not achieved even with clustering OTUs with a cut-off at 0.0029 (Figure 5.1B), suggesting that only a portion of the total diversity was sampled.

5.2.3 MegaBlast results

The high amount of haplotype diversity that is observed for the mtDNA data raised the question of whether microbial contamination connected to biota that may inhabit the skin of *P. maculata* interfered with the sequencing data or not because it was not always possible to choose a specific tissue for DNA extraction during DNA extraction. 1060 bp of partial Cytb sequences from 74 unique haplotypes and 1153 bp of partial COI sequences from 103 unique haplotypes were used to interrogate nucleotide collections in the NCBI database using the MegaBlast algorithm. The aim was to test whether the retrieved sequences belonged to a microorganism, or not. The first 100 retrieved sequences that were evaluated for the Cytb and COI sequences are all from respectively mtDNA Cytb or COI sequences from sea slugs, other gastropods, and several insect or arachnid species. During the evaluation, the focus was on the query coverage for each search and pairwise identity score, but also grade score, which is a percentage weighted score for the retrieved sequence comprised of the expected value (*E*), query coverage and the pairwise identity. *E* describes the random number of matches that is expected while searching a database of a particular size. Matches to Cytb sequences with >90% query coverage are from Cytb sequences from five sea slug species (*Aplysia kurodai*, *Platevindex mortoni*, *Onchidium siruma*, *Berthelina* sp. *TLT-2006* and *Placida* sp. (*NY-2013*) with 74–77% pairwise identity and 85–87% grade over 1060bp. The lowest coverage is 26% over 1060bp. The matches to COI sequences with >90% query coverage are mainly from many other sea slug species (e.g., *Pleurobranchaea meckeli, Berthelina sp.TLT-2006, Rostanga pulchra* and *Limax sp*) with 83.6–84.8% pairwise identity and 99.65% grade over 1153 bp, while the lowest query coverage was 54%.

COI and Cytb sequences from a single individual were compared against the nucleotide database in 100 bp intervals to check the presence of any hybrid sequences in case of contamination. The results for the species with the first highest grades are shown in Table 5.2. Pieces from the Cytb gene either do not match any sequences or they match

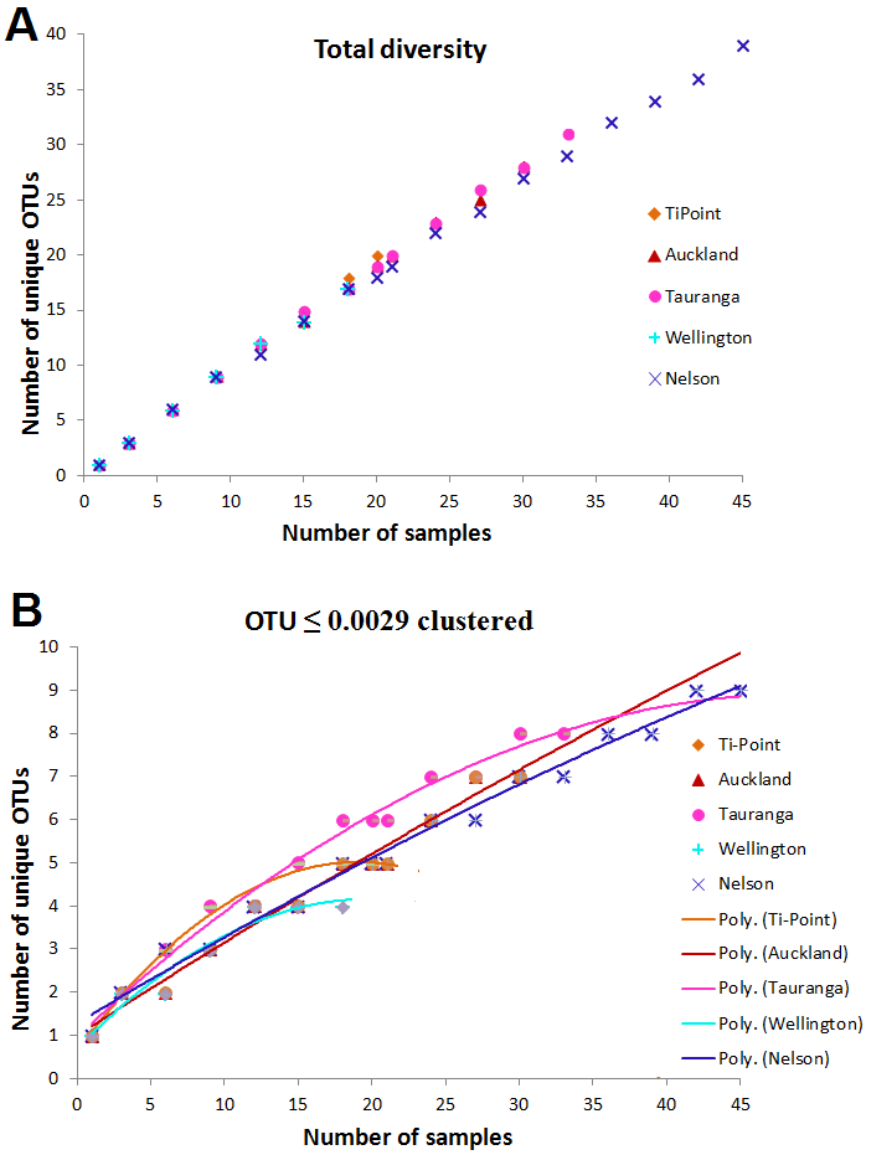


Figure 5.1 Rarefaction curves for the Cytb sequences. **A)** total diversity, **B)** diversity after clustering OTUs that are ≤0.0029 distant from each other. A cut-off point was chosen to remove the rare sequences whose distances vary approximately 1/3 of the maximum distance observed between individuals, i.e. 0.009434. Poly: polynomial trend lines for the data points. Population abbreviations within the text are indicated in parentheses: Ti Point (TP), Auckland (AKL), Tauranga (TR), Wellington (WL) and Nelson (NL).

Cytb sequences from other Metazoan species. Pieces from the COI gene match *P. maculata* COI sequences stored in the database with very high grades (99.5–100%), but also those of some insect species, especially for the 3' half of the sequenced part due to the fact that this region of the gene is not available for the *P. maculata* species in the database. Here, the main concern was microbial contamination, but no matches from microbial genomes are observed. Matching sequences are mainly from other sea slug species with high scores, which also confirms the absence of contamination.

COI and Cytb sequences from a single individual were compared against the nucleotide database in 100 bp intervals to check the presence of any hybrid sequences in case of contamination. The results for the species with the first highest grades are shown in Table 5.2. Pieces from the Cytb gene either do not match any sequences or they match Cytb sequences from other Metazoan species. Pieces from the COI gene match *P. maculata* COI sequences stored in the database with very high grades (99.5–100%), but also those of some insect species, especially for the 3' half of the sequenced part due to the fact that this region of the gene is not available for the *P. maculata* species in the database. Here, the main concern was microbial contamination, but no matches from microbial genomes are observed. Matching sequences are mainly from other sea slug species with high scores, which also confirms the absence of contamination

Table 5.2 MegaBlast results for the COI and Cytb sequences of the *P. maculata* individuals. The sequences were used to interrogate the NCBI nucleotide database. The search was performed by dividing the sequences into 100 bp intervals.

Gene Name	Hit sequence	interval (bp)	Coverage %	Grade %	Pairwise identity %	taxa
Cytb	no significant match	1—100 bp	-	-	-	-
	Gymnothorax undulates Cytb gene	101-200	40	42.5	0.95	moray eel
	no significant match	201—300	-	-	-	-
	Megaselia scalaris Cytb gene	301—400	0.94	91.7—92.7	89.4—91.5	a fly species
	no significant match	401—500	-	-	-	-
	Empoas cavitis or *Spinocalanus abyssalis* Cytb gene	501—600	0.28	39	100	a leaf-hopper and crustacean
	no significant match	601—700	-	-	-	-
	Black burnia sp or *Sepsis fissa* Cytb gene	701—800	88—92	81.8—88.33	88—89.8	insect species
	no significant match	801—900	-	-	-	-
	no significant match	901—1000	-	-	-	-
	no significant match	1001—1060	-	-	-	-
COI	*P. maculata* COI gene	1—100 bp	100	99.5—100	99	study organism
	P. maculata COI gene	101—200	100	99.5—100	99—100	study organism
	P. maculata COI gene	201—300	100	100	100	study organism
	P. maculata COI gene	301—400	100	100	100	study organism
	P. maculata COI gene	401—500	100	100	100	study organism
	P. maculata COI gene	501—600	99—100	99.5—100	100	study organism
	Pleurobranchaea meckeli or *Natalina knysnaensis* COI gene	601—700	0.94	92.7	91.5	sea slug
	Varroa sp COI gene	701—800	100	92.9—94	87—88	honeybee mite
	Sarcophaga albiceps COI *gene*	801—900	91	62.3	83.5	fly species
	Petalifera petalifera COI gene	901—1000	100	68	86.3	sea slug
	Neochrysocharis formosa COI gene	1001—1153	100	95.5—97	91—94.1	fly species

5.2.4 Genetic differentiation

5.2.4.1 Haplotype networks

All organisms share similarities in their genomes as they derive from a common ancestor. However, genetic diversity arises following mutations, duplications, genome reorganisation, and recombination. Phylogenetic analysis aims to reveal the evolutionary relationship between different species, taxa or individuals by comparing various traits; but most popularly, sequence data (DNA or amino acid) of homologous genes (Brinkman and Leipe, 2001). Phylogenetic trees are one of the tools that are used to achieve this task explicitly. A phylogenetic tree is composed of nodes representing points of diversification, internal branches connecting the nodes and terminal branches representing the existing taxa and data (Lipscomb, 1998). Phylogenetic trees can represent the history of taxa if the history is constructed of diversification events and changes that are restricted to branches without interaction between branches. However, trees cannot represent the history of the taxa adequately when there is substantial horizontal gene transfer, hybridisation between taxa, recombination or incomplete lineage sorting. For such events, phylogenetic networks are more suitable to represent the relationship between taxa, allowing visualisation of conflicting and ambiguous signals but without representing the evolutionary history explicitly (Huson and Bryant, 2006; Huson and Scornavacca, 2011).

In this study, the mutational relationship between the haplotypes and their geographic distribution was analysed using a median joining haplotype network (MJN) (Bandelt et al., 1999). The median-joining method is best suited for analysis of closely related sequences that have evolved without recombination, and is widely used in phylogeography and population genetic studies that are based on mtDNA or the Y chromosome. MJN resolves nucleotide distances between haplotypes but also takes individual nucleotide substitutions into account (Bandelt et al., 1999; Huson and Scornavacca, 2011). The Cytb network was utilised (Figure 5.2) to infer the relationship between individuals due to it having a less complicated pattern than that of the COI (Supplementary Figure 3) and concatenated (Supplementary Figure 4) networks. The highly polymorphic structure of the latter two networks could have obscured the evolutionary relationship between individuals. All three haplotype networks resulted in a similar pattern; a dispersed star-like structure in which common haplotypes are shared across the populations. The two most common Cytb haplotypes are shared by 25 (16.0%

of the total dataset) and 22 (14.1% of the total dataset) individuals from all the sampling locations (frequency of haplotypes in Supplementary Table 11).The network is highly complex with many reticulations (cycles) due to homoplasy. Internal nodes can be deduced as ancestral haplotypes; however, the reticulations between many internal nodes create uncertainty about the order of the nucleotide substitutions. In this regard, it is not easy to pinpoint a branch that could be the root of the haplotypes. In addition, an important feature of the network is its star-like topology; many private haplotypes descend from central shared nodes with mostly one to two bp distances (Slatkin and Hudson, 1991), and these private haplotypes are dispersed among all the sampling locations. It is also interesting that the internal nodes from the haplotypes shared by several populations generally belong to individuals from the upper North Island populations (WH, TP, AKL and TR), which may suggest that these populations are the source of the dispersal. Additionally, some of the private haplotypes belonging to the upper North Island populations are several bp away from the rest, and some imaginary nodes are necessary to explain them (please see the right hand side of (Figure 5.2), which may be a sign of the existence of distinctive clusters within the North Island populations.

Errors in the data due to factors such as sequencing artefacts, artificial recombination and alignment problems may result in reticulations and ambiguity in networks (Zimmermann et al., 2014). In order to eliminate sequencing-related quality issues, the amplified fragments were sequenced from both ends. In the case of ambiguous chromatograms, the PCR and sequencing processes were repeated in order to gain better quality data. As the COI and Cytb genes were amplified by two sets of primers, artificial recombination might have occurred if the fragments from the wrong individuals had been assembled. To overcome this problem, several individuals randomly chosen from each population were sequenced for a second time at both loci to ensure that there was no error in the dataset. Additionally, MJN haplotypes for the Cytb sequences were drawn separately for the 5’ (530bp) and the 3’ halves (530bp) amplified by the two sets of primer pairs (Supplementary Figure 5A and B, respectively).The 3’ half included overlapping sequences from the 5’ amplicon. The MJN networks show that the ambiguity and reticulation, as well as the lack of population structure based on the sampling location are still evident for both the 3’ and 5’ fragments. The network for the 5’ fragment is much less complex as this fragment contains less variation. All these

results suggest that the observed complex haplotype networks are not the result of errors that occurred during data collection, but instead represent the real population structure.

The third positions of codons encoding an amino acid are mutated more often than the first and the second positions as the mutations in the third position are generally synonymous, i.e. they do not change the identity of the amino acid due to the redundancy of the genetic code (Lehmann and Libchaber, 2008). A high mutation rate at the third codon may lead to homoplasies, which means that a particular shared mutation is not identical by descent, but instead emerges as a result of parallel evolution

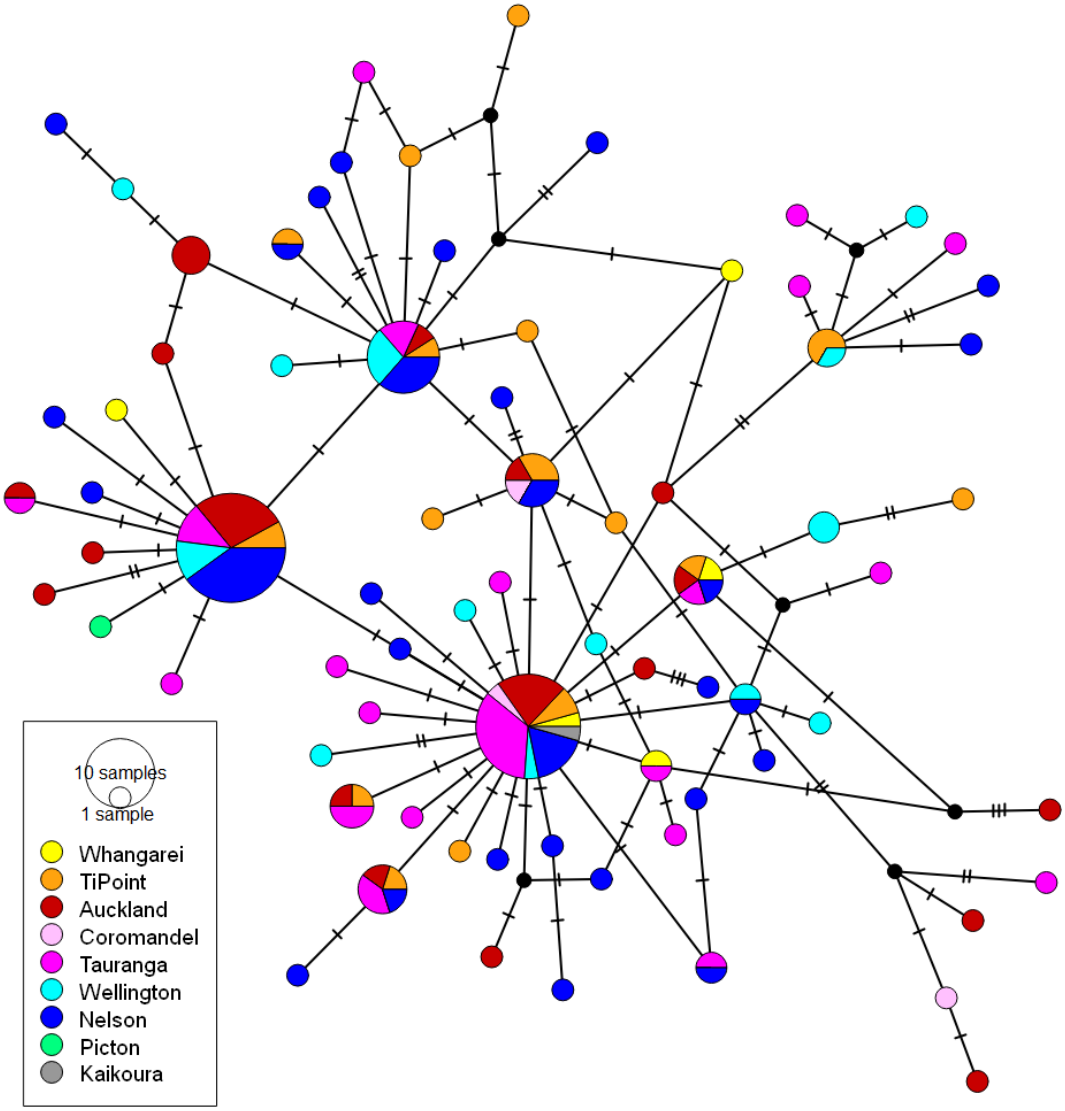


Figure 5.2 Median joining network of the Cytb haplotypes. The network was coloured according to the sampling locations. The diameter of the circles is proportional to the frequency of the haplotypes. The hashes indicate the mutational steps between the haplotypes. The black nodes represent the imaginary haplotypes necessary to create a bridge between the present haplotypes. Abbreviations of the sampling locations are in parentheses: Whangarei (WH), Ti Point (TP), Auckland (AKL), Coromandel (CR), Tauranga (TR), Wellington (WL), Picton (PC), Nelson (NL) and Kaikoura (KK).

(or recombination). In order to test whether the homoplasies obscured any present population structure in the *P. maculata* populations, MJN analysis was performed for the COI gene excluding the third position from the data. This time COI but not Cytb was chosen as the former was considered to have been more severely affected by homoplasies due to its higher levels of polymorphism. The resulting network (Supplementary Figure 5C) is noticeably simple, and the majority of the haplotypes are identical. Again, no structure is evident with respect to the sampling location. The index of substitution saturation (*ISS*) was examined with DAMBE and used to test homoplasy due to multiple substitutions (Xia and Lemey, 2009; Xia et al., 2003). For both symmetrical and asymmetrical tree topology models and for both genes, the observed *ISS* values are significantly larger than the critical *ISS* (*ISS.c*) values (Table 5.3), which indicates that the paired partitions are not saturated and the degree of homoplasy is low. This result may suggest that the possibility of a lack of structure in the MJN network is not a result of homoplasy, but instead due to identical mutations that are identical by descent.

Table 5.3 Results of saturation tests for COI and Cytb sequences.

Gene	# of OTUs	*ISS*	*ISS.cSym*	*t*	*d.f.*	*P*	*ISS.cAsym*	*t*	*d.f.*	*P*
COI	4	0.027	0.825	75.639	242	0.0000	0.793	72.612	242	0.0000
	8	0.028	0.796	73.300	242	0.0000	0.692	63.374	242	0.0000
	16	0.031	0.779	75.979	242	0.0000	0.586	56.391	242	0.0000
	32	0.033	0.758	78.971	242	0.0000	0.46	46.559	242	0.0000
Cytb	4	0.035	0.822	55.039	164	0.0000	0.790	52.802	164	0.0000
	8	0.037	0.791	52.580	164	0.0000	0.686	45.266	164	0.0000
	16	0.037	0.775	55.479	164	0.0000	0.579	40.741	164	0.0000
	32	0.040	0.752	57.984	164	0.0000	0.450	33.365	164	0.0000

ISS: the index of substitution saturation, *ISS.c*: critical *ISS*, Sym: assumes a symmetrical topology, Asym is *ISS.c* assumes an asymmetrical topology. *t*: t-value, *d.f.*: degrees of freedom. A significant difference was observed between *ISS* and *ISS.c* values, and *ISS* < *ISS.c*.

5.2.4.2 Individual distance matrices and the multivariate approach

Genetic differentiation between the populations using both Cytb and COI data was estimated initially based on the standardised nucleotide distance (D_{SEQ}) between individuals that was previously used to calculate the OTUs. In the multivariate approach, samples can be imagined as objects in multi-dimensional space where the distance matrix represents the relationship between each pair of objects (Please refer to the sub-heading “Multivariate analysis” in subsection 3.5.2.1 for information on the

analysis). Both distance matrices were first visualised using a non-metric 2-D Multi-Dimensional Scaling analyses (MDS), which is a multivariate ordination method that

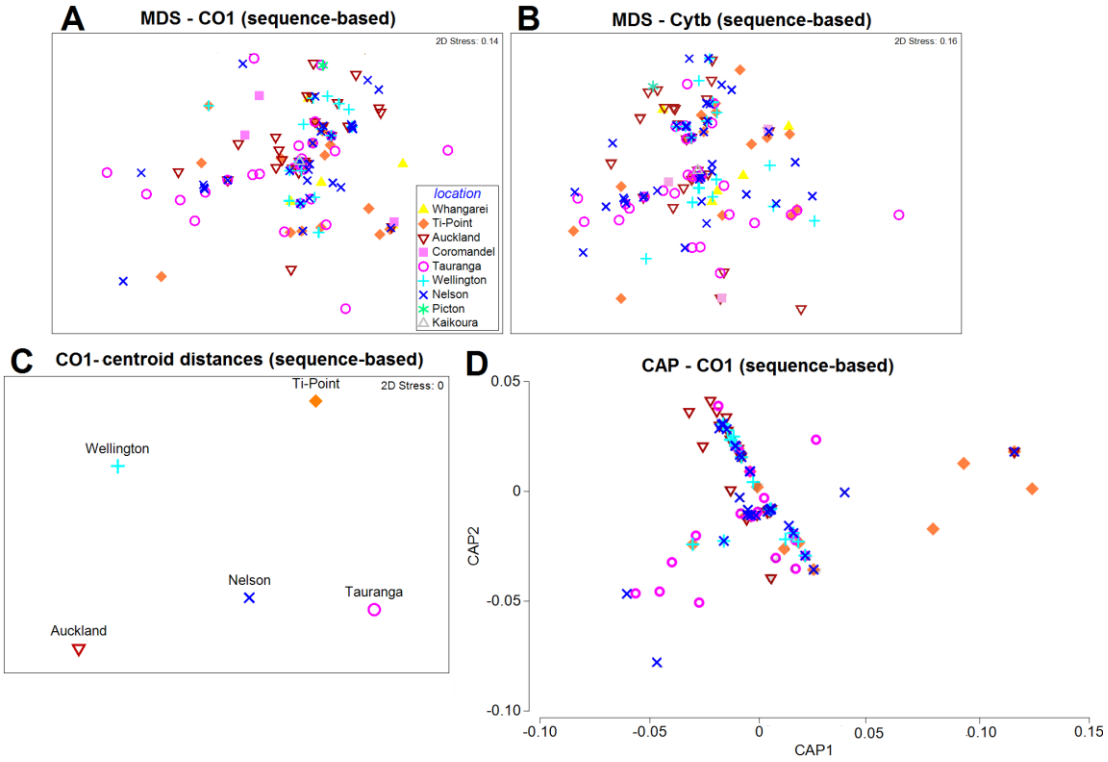


Figure 5.3 Graphical visualisation of the standard nucleotide differences between individuals based on the mtDNA data. Non-metric MDS ordination of distances obtained from A) COI and B) Cytb sequences. C) MDS of distances between centroids of each population for COI (not drawn for Cytb, as PERMANOVA did not support significant differences between populations). D) Canonical correlation ordination (CAP) analysis of COI data at m=4 showing the placement of the points into the axes that maximises the population differentiation. Abbreviations of the sampling locations are in parentheses: Whangarei (WH), Ti Point (TP), Auckland (AKL), Coromandel (CR), Tauranga (TR), Wellington (WL), Picton (PC), Nelson (NL) and Kaikoura (KK).

represents dissimilarities between samples in a low-dimensional space (Kruskal, 1964). Similar to the haplotype network results, MDS analysis of the D_{SEQ} matrices reveals no observable structure at first glance based on sampling locations from either COI or Cytb sequences (Figure 5.3A and B). Yet, for COI data, the TR individuals are clustered slightly separately from the rest. In addition, the AKL samples formed as a distinct cluster in the graph, although these patterns are not very clear. Permutational Multivariate Analysis of Variance (PERMANOVA) (Anderson, 2001; McArdle and Anderson, 2001) was conducted to test the null hypothesis that there is no difference between the main five sampling locations. PERMANOVA suggested that the sampling location had a significant effect on the population structure for the COI sequences (COI: $F_{4, 145}$=1.8791, P=0.0173), but not for the Cytb sequences (Cytb: $F_{4, 145}$=1.3386,

P=0.1806). Pairwise PERMANOVA analysis that was conducted for COI to investigate which sampling locations have diverged from each other shows a significant differentiation between the AKL and TP populations (*P*=0.0276), and the AKL and TR populations (*P*=0.0218) (Table 5.4). This differentiation pattern can be seen in the MDS graph of the centroid distances between the populations, in which AKL clusters with WL and NL, and is the furthest point from TP and TR. On the other hand, individuals from TP, TR, WL and NL are located relatively close to each other (Figure 5.3C).

MDS analysis is an unconstrained method that visualises the data through the dimensions that contain the highest amount of variation. However, the differences between the groups can be hidden at dimensions with smaller variations. Canonical Analysis of Principle Coordinates (CAP), which is a constrained ordination method, projects the observations onto the axis that maximises the variation between the pre-defined groups. The main five populations were introduced as pre-defined groups in this analysis. The dimensions that explain the variation between the populations can be most readily visualised in this way (Anderson and Willis, 2003). CAP analysis was performed for the COI but not for the Cytb data because PERMANOVA did not find significant differences between the groups for the latter. After a diagnostic analysis with CAP, only four principal component axes (*m*) were chosen because there was not a large gain in allocation success rates beyond these values (data not shown). The resulting first four axes explain 74.4% of the total variation in the data, which means that the chosen axes can capture a high proportion of the variation. CAP analysis yielded four canonical axes onto which the samples are projected. The squared canonical correlation coefficients (δ^2), which indicate the strength of the association of the hypothesis and observations, are low for each canonical axis (δ_1^2=0.1155, δ_2^2=0.0698, δ_3^2=0.0296, δ_4^2=0.0037), implying a weak correlation with the hypothesis of population differences and the data space. δ^2 varies between -1 and 1, in which -1 means total negative correlation, 0 means no correlation, and 1 means total positive correlation. The plot produced through CAP (Figure 5.3D) improves the visualisation of group differences somewhat. Similar to the results of the MDS graph, individuals from AKL are clustered separately from some individuals of TP and TR. Cross-validation testing shows high overall misclassification rates of 76.0% (101 out of 146). Considering that a correct classification rate of 20% is expected just by chance (for data constructed of five groups), the overall success rate of 24% that is observed for this data can also be observed just by chance. Therefore, this

data does not support the hypothesis of population differentiation. Despite the low overall success rate, the COI data yield relatively high correct classification for the AKL (42.4 %) and TR (50%) individuals at the population level. Additionally, the TP population has a 30% correct-rate, which may suggest weak differentiation (Table 5.5). Trace statistics (*tr*), which is the sum of the squared canonical correlations (Anderson and Willis, 2003), show a significant differentiation between the centroids of the populations (*tr*=0.21851, *P*=0.0087). As a result, the CAP result is congruent with the structure that is revealed by PERMANOVA. Among the five populations, only AKL is markedly distant from TP and TR for COI data in terms of nucleotide distances.

Table 5.4 Results of the statistical tests performed to compare the distribution of the genetic variation in mtDNA between five *P. maculata* populations. The analysis was performed separately for the COI and Cytb sequences. Standardised nucleotide distances between individuals were used for PERMANOVA analysis. Pseudo-t values (*t*) represent the distance between the populations. Fixation indexes were calculated based on haplotype frequencies (F_{ST}) and nucleotide diversity (Φ_{ST}) where nucleotide distances between the haplotypes were implemented. The Tamura-Nei mutational model was used for Φ_{ST} calculations.

		PERMANOVA		F-statistics			
		Sequence-based		Haplotype-based		Sequence-based	
	Groups	t	*P*	F_{ST}	*P*	Φ_{ST}	*P*
COI	TP-AKL	1.585*	0.0203	0.014**	0.0066	0.024	0.0617
	TP-TR	1.359	0.0916	0.017**	0.0055	0.018	0.1019
	TP-WL	1.233	0.1596	0.014*	0.0239	0.008	0.2861
	TP-NL	1.494	0.0646	0.030**	0.0032	0.029	0.0517
	AKL-TR	1.712*	0.0175	0.015***	0.0007	0.020	0.0556
	AKL-WL	1.202	0.1567	0.012*	0.0183	0.014	0.1276
	AKL-NL	1.283	0.1487	0.028***	0.0002	0.005	0.2341
	TR-WL	1.403	0.0797	0.016*	0.0153	0.017	0.1058
	TR-NL	1.076	0.3111	0.031***	0.0000	0.007	0.1897
	WL-NL	1.326	0.1214	0.029**	0.0055	0.020	0.0946
Cytb[a]	TP-AKL	NA	NA	0.050***	0.0008	0.028	0.0597
	TP-TR	NA	NA	0.043**	0.0021	0.023	0.0951
	TP-WL	NA	NA	0.033***	0.0004	-0.026	0.9083
	TP-NL	NA	NA	0.041**	0.0015	0.000	0.4031
	AKL-TR	NA	NA	0.071***	0.0000	0.019	0.0725
	AKL-WL	NA	NA	0.062***	0.0007	-0.003	0.4722
	AKL-NL	NA	NA	0.068***	0.0000	-0.001	0.4366
	TR-WL	NA	NA	0.056***	0.0010	0.012	0.1999
	TR-NL	NA	NA	0.061***	0.0000	0.026*	0.0226
	WL-NL	NA	NA	0.053***	0.0007	-0.014	0.8416

P values for genetic differentiation: * $P<0.05$, ** $P<0.01$, *** $P<0.001$. NA: calculations were not performed because an overall differentiation was not observed.

PERMANOVA is sensitive to differences in dispersion of samples in the multivariate space between groups. Any significant difference found between populations identified by PERMANOVA can be the result of differences in the location of samples, differences in the dispersion of samples or a mixture of both. The population sizes in this study are variable, which could have led to dispersion differences between the groups. PERMDISP analysis was performed to test whether the significant differences found by PERMANOVA are the result of possible heterogeneity of multivariate dispersions within the groups (Anderson, 2006). The results of the test do not reveal significant differences in dispersion of the variations for either the COI ($F_{4,141}$=0.3793, P=0.847) or Cytb ($F_{4,141}$=0.358, P=0.833) data, meaning that nucleotide distances between individuals are homogeneously distributed among the groups, and that differences in the sample sizes do not affect the output of the statistical test. Non-significant results also show that the significant differences between populations, which are suggested by PERMANOVA, can be confidently attributed to the differences in fragmentation of genetic variation.

Table 5.5 Results of the cross-validation test for the COI data. CAP was used to calculate leave-one-out allocation rates of individuals to populations. The standard nucleotide differences between individuals were used as the distance method.

			Classified					
	Orig. group	**N**	**TP**	**AKL**	**TR**	**WL**	**NL**	**% correct**
COI (m = 4)	Ti-Point	20	6	6	7	1	0	30.0
	Auckland	30	1	15	13	0	1	50.0
	Tauranga	33	5	7	14	3	4	42.4
	Wellington	18	4	8	6	0	0	0
	Nelson	45	6	22	15	2	0	0
	Across populations	146	-	-	-	-	-	24.0 (35)

Abbreviations of the sampling locations are in parentheses: Whangarei (WH), Ti Point (TP), Auckland (AKL), Coromandel (CR), Tauranga (TR), Wellington (WL), Picton (PC), Nelson (NL) and Kaikoura (KK).

5.2.4.3 Differentiation based on F-statistics

Pairwise population differentiation analysis was also performed to calculate the fixation indices. The five main populations were included in the calculations. The fixation index is defined as the variance of allele frequencies between populations relative to the total variation (Holsinger and Weir, 2009). Two different fixation indexes were used: conventional F_{ST} (Weir and Cockerham, 1984), which calculates the differences between haplotype frequencies, and Φ_{ST}, which includes nucleotide differences between the haplotypes during frequency calculations (Excoffier et al., 1992).The Tamura-Nei

mutational model was implemented for Φ_{ST} calculations (Tamura and Nei, 1993). The fixation index varies between 0 and 1.0; 0 means that frequencies are identical between the two populations, while 1 means that the populations do not share any alleles (Wright, 1951; Wright, 1965).

5.2.4.4 Haplotype-based F_{ST}

Pairwise F_{ST} values are low for all the comparisons; they vary between 0.033 and 0.071 for the Cytb data, and between 0.012 and 0.031 for the COI data (Table 5.4). Despite the low values, differentiation between the populations is significant for all the comparisons. The level of significance is higher for the Cytb data (P=0.002–0.0000) than that of the COI data (P=0.0236–0.0000) as the P values are smaller for most of the

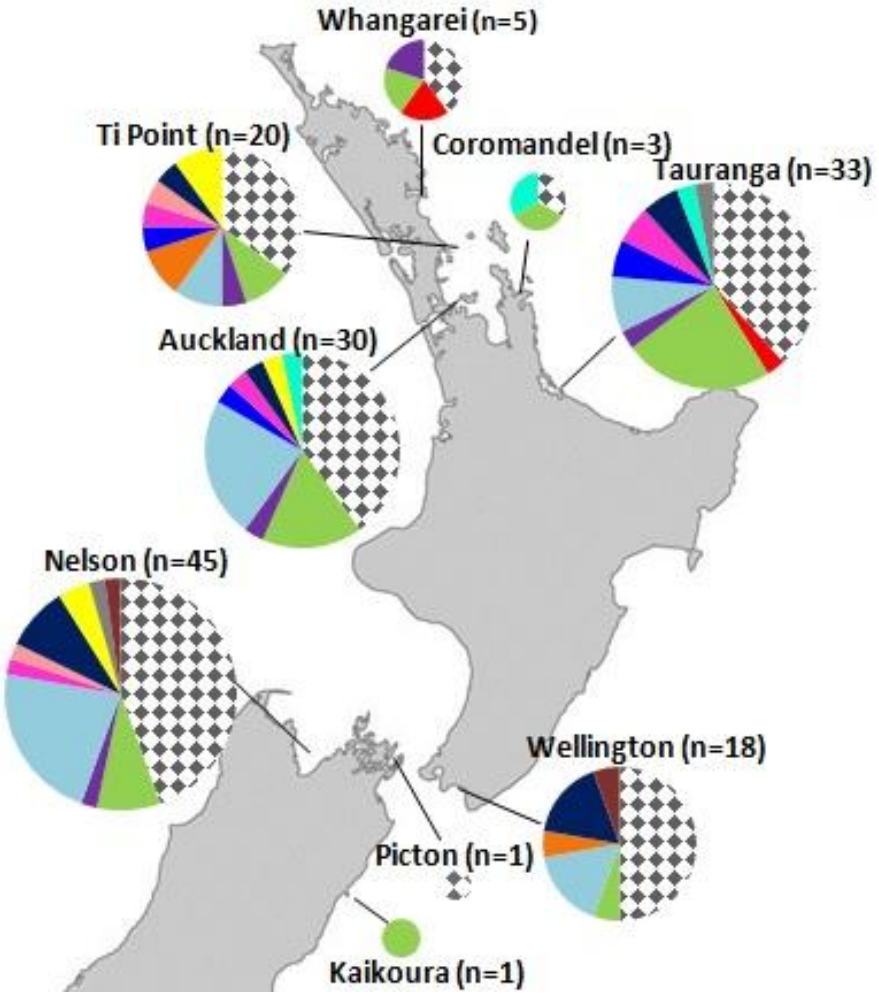


Figure 5.4 The frequencies of Cytb haplotypes at each location. The pie segment represents the relative haplotype frequencies. Each colour corresponds to a different haplotype. The patterned segment represents private haplotypes. The sizes of the circles are proportional to the sample size. Abbreviations of the sampling locations are in parentheses: Whangarei (WH), Ti Point (TP), Auckland (AKL), Coromandel (CR), Tauranga (TR), Wellington (WL), Picton (PC), Nelson (NL) and Kaikoura (KK).

comparisons for the COI data (Table 5.4).The response of F_{ST} to within-population diversity generally underestimates the differentiation, because fixation of alternative alleles is not achievable in the event of high allelic diversity (Bird et al., 2011; Jost, 2008). Additionally, diversity in shared alleles defines the behaviour of F_{ST} more than rare (and private) alleles (Bird et al., 2011). In my case, within-population diversity is high as all the populations contain high numbers of private and several shared haplotypes (distribution of Cytb and COI haplotypes in Figure 5.4 and Supplementary Figure 6, respectively). This may explain the low F_{ST} values despite the high haplotype diversity. The F_{ST} values for COI are even smaller than Cytb as the within-population diversity is higher for COI. The actual differentiation between the populations may be higher than the estimated values.

Significant differentiation between all the population pairs based on the F_{ST} values can be explained by variations in frequencies of the shared haplotypes between the populations. For example, for the Cytb data (Figure 5.4), the light blue haplotype is represented in greater numbers in AKL, WL and NL compared to others. The green haplotype is represented in AKL and TR more frequently than in TP, WL and NL. The dark pink and purple haplotypes are absent in WL, although they are represented in other populations.TP contains almost all the shared haplotypes, whereas the number of shared haplotypes is low for WL (five haplotypes). Based on the haplotype frequencies, I can conclude that there are differences between the populations.

5.2.4.5 Sequence-based Φ_{ST}

According to F_{ST}, all haplotypes are equally distant from each other regardless of the base pair differences between them, which leads to a loss of information. On the other hand, Φ_{ST} incorporates nucleotide distances between haplotypes into the differentiation calculation. Consequently, it includes more information about the genetic composition of the populations (Bird et al., 2011; Excoffier et al., 1992). In this study, similar to the F_{ST} values, the pairwise Φ_{ST} values are low for all the comparisons, ranging from 0.007 to 0.029 for the COI data and from 0.026 to 0.028 for the Cytb data (Table 5.4). None of the populations are significantly differentiated from each other for the COI data, whereas TR and NL are weakly but significantly differentiated for the Cytb sequences (Φ_{ST}=0.026, P=0.0226). These results differ from those of the PERMANOVA and CAP analyses in terms of the identity of significantly differentiated populations for the COI

data. PERMANOVA and CAP also did not find any significant differences between the populations for the Cytb data, but Φ_{ST} did. However, the *P* values obtained from each estimator are very similar to each other (Table 5.4). If one estimator points out a significant differentiation but the others do not, the *P* value of the second is always between 0.05 and 0.1. Different statistics have different properties, and combining them allowed the data to be investigated from different perspectives.

Table 5.6 AMOVA results of concatenated sequences based on F and Θ-statistics.

	Structure	Source of variation	d.f.	Variance components	%var	Fixation indices	*P* value
Θ_{ST}	One group of five populations	Among sites	4	0.046	1.26	Θ_{ST}=0.0126	0.0308
		Within sites	141	3.607	98.74		
		Total	145	3.653	100		
	2 population groups	Among groups	1	-0.022	-0.62	Θ_{CT}=-0.0062	0.8994
		Among sites within groups	3	0.060	1.65	Θ_{SC}=0.0164	0.0262
		Within sites	141	3.607	98.96		
		Total	145	3.645	100	Θ_{ST}=0.0104	0.0316
F_{ST}	One group of five populations	Among sites	4	0.003	0.69	F_{ST} = 0.0069	0.0001
		Within sites	141	0.496	99.31		
		Total	145	0.500	100		
	2 population groups	Among groups	1	0.001	0.16	F_{CT}= 0.0016	0.1003
		Among sites within groups	3	0.003	0.58	F_{SC}=0.0059	0.0062
		Within sites	141	0.496	99.26		
		Total	145	0.500	100	F_{ST}=0.0075	0.0000

5.2.4.6 Molecular variance and the search for the north-south disjunction

The hierarchical distribution of variation was tested with AMOVA based on nucleotide diversity (Θ_{ST}) and haplotype frequencies (F_{ST}) using two models: one in which there was no structure introduced and one in which groups were pre-defined (TP, AKL and TR populations defined a Northern Group, whereas WL and NL formed the Southern Group). Analysis of the microsatellite data identified these two genetically differentiated Northern and Southern clusters (please refer to Chapter 4). The aim of this grouping during analysis of mtDNA sequences was to test whether the north-south differentiation can be observed with the mtDNA data or not. Analysis of both sequences without introducing any population structure showed that differentiation among sampling sites explains between 1.55 % and 5.67 % of the total variation based on Θ_{ST} and F_{ST} statistics, respectively, with high significance (Θ_{ST} = 0.0126, *P*=0.0308; F_{ST} = 0.0096, *P*=0.0001), whereas variation within populations explained 98.74% (Θ_{ST}) or 99.31%

(F_{ST}) of the total variation (Table 5.6). These results were in congruent with the other differentiation analyses results. Grouping of populations into Northern and Southern regions explained a maximum of 0.16% of the total variation, but it was not significant (Θ_{CT}=-0.0062, P=0.8994; F_{CT}=0.0016, P=0.1003). PERMANOVA analysis was performed using the D_{SEQ} matrix by contrasting the TP, AKL and TR populations (north) against WL and NL (south). A significant differentiation was not observed between the Northern and Southern clusters (COI: $F_{1, 145}$=0.819, P=0.5268; Cytb: $F_{1, 145}$=1.2325, P=0.2982). These results indicate that the mtDNA data do not support the north-south differentiation that was revealed by the microsatellite data.

5.2.5 Spatial analysis

The isolation by distance model was tested with the Mantel test (Mantel, 1967), which calculates the correlation between pairwise genetic and geographical distances. The results do not show a significant correlation between pairwise F_{ST} and geographical distance between populations (Figure 5.5, r =-0.0222, P=0.5490), suggesting that the isolation by distance model (IBD) is not supported for the mtDNA data.

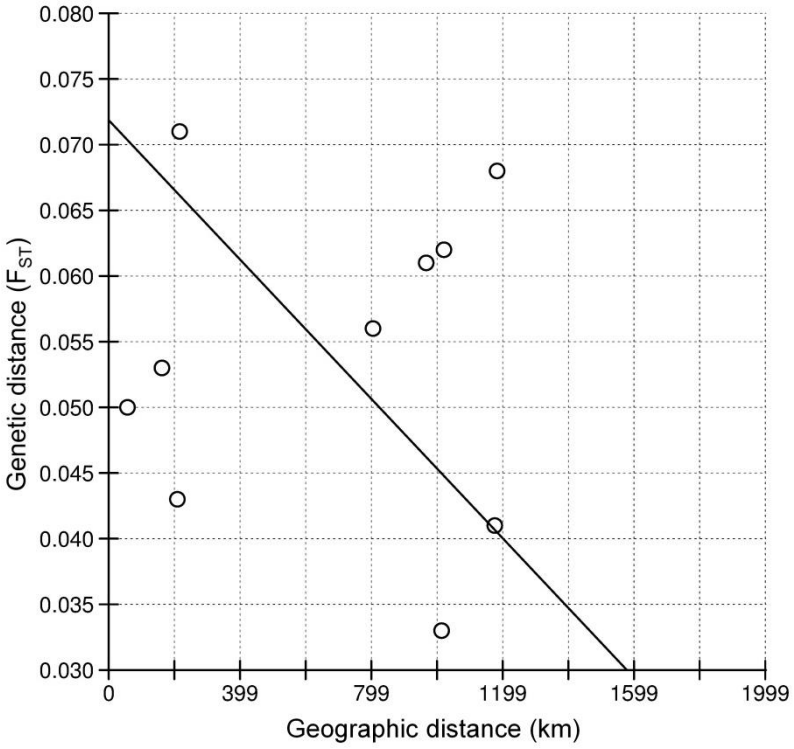


Figure 5.5 Test of isolation by distance, based on the Cytb data. Relationship between pairwise genetic distances (F_{ST}) and geographical distances (km). Isolation by distance model was not supported (r=-0.0222, P=0.5490).

5.2.6 Neutrality tests and demographic changes

According to coalescent theory, which assumes neutral mutation and constant population size, the common ancestor of two randomly chosen sequences lived *2N* generations ago, where *N* is the effective population size. If μ is the neutral mutation rate per nucleotide site per generation for a particular gene, the nucleotide difference between two sequences will be $2N\mu$ for haploid sequences since the last common ancestor. This value of $2N\mu$, which is symbolised by θ, is also named as the mutation rate of population (Watterson, 1975). θ can be calculated from various diversity estimators such as nucleotide diversity (θ_π), the number of segregating sites (θ_S) or haplotype diversity (θ_{Hap}). In a population in mutation-drift equilibrium, θ_π and θ_S are expected to be similar assuming that the compared sequences are closely related enough for all mutations to have occurred at a different site. Although selective or demographic changes affect both values, π is more sensitive to such changes. In a similar way, Tajima's *D* determined whether these two θ values are statistically indistinguishable from each other ($D=\theta_\pi-\theta_S$). Significant deviations from zero are attributable to non-random evolutionary processes, including selection or demographic changes (Tajima, 1989). Similarly, Fu's *Fs* (Fu, 1997) compares θ_{Hap} and θ_S; significant discordance between these two values is accepted as evidence of deviations from neutrality. Negative Tajima's *D* and F_S values may show population expansion or purifying selection, whereas positive values may show a population decline or balancing selection (Holsinger, 2012). Both of the tests were performed for *P. maculata* to investigate deviations from neutrality.

The results of Tajima's *D* and Fu's *Fs* tests for five populations and across populations are shown in Table 5.7. The global values of Tajima's *D* and Fu's *Fs* tests are negative and significant, based on analysis of the COI and Cytb sequences, suggesting either deviation from neutrality (particularly purifying selection) or an overall population expansion for the species around NZ. Regarding demographic changes at the population-level, analysis of the COI sequences show that all the populations, except TP, exhibit significantly negative Tajima's *D* values. Analysis of the Cytb sequences yielded similar results, except that the Tajima's *D* value is not significantly negative for WL in addition to TP. Fu's *Fs* values are significant for all the populations. Tajima's *D* has less power to detect demographic changes compared to alternative methods (Schneider and Excoffier, 1999). This may explain why deviations in the TP and WL

populations were not detected with this method, but are evident with the Fu's *Fs* method.

Table 5.7 Results of the neutrality and demographic tests for the mtDNA data. Time since population expansion (Tau-τ) Tajima's *D*, Fu's *Fs* and the parameters of the mismatch distributions based on the mitochondrial COI, Cytb and concatenated data were calculated for five populations and across the populations.

					Mismatch Distribution	
	Pop	**τ**	**Tajima's *D***	**Fu's *Fs***	***SSD* (P_{SSD})**	***Raggedness* (P_r)**
COI	**Ti Point (TP)**	5.406	-1.45	-10.82***	0.013 (0.225)	0.035 (0.298)
	Auckland (AKL)	4.152	-2.16**	-25.26***	0.002 (0.537)	0.026 (0.417)
	Tauranga (TR)	2.68	-1.93*	-23.87***	0.003 (0.532)	0.015 (0.676)
	Wellington (WL)	4.404	-1.68*	-11.66***	0.012 (0.212)	0.048 (0.215)
	Nelson (NL)	4.127	-2.09**	-24.73***	0.002 (0.761)	0.014 (0.847)
	Total	3.572	-2.34***	-26.43***	0.003 (0.120)	0.032 (0.453)
Cytb	**Ti Point (TP)**	3.008	-1.31	-12.39***	0.007 (0.384)	0.053 (0.289)
	Auckland (AKL)	1.512	-1.90*	-12.68***	0.006 (0.391)	0.047 (0.415)
	Tauranga (TR)	1.371	-2.16**	-17.08***	0.010 (0.307)	0.053 (0.258)
	Wellington (WL)	3.439	-1.38	-6.94***	0.003 (0.718)	0.029 (0.696)
	Nelson (NL)	2.094	-2.17**	-26.33***	0.001 (0.672)	0.029 (0.588)
	Total	2.754	-2.34***	-26.43***	0.003 (0.115)	0.032 (0.454)
Concatenated	**Ti Point (TP)**	8.783	-1.45	-14.57***	0.010 (0.368)	0.033 (0.184)
	Auckland (AKL)	5.254	-2.14**	-23.57***	0.017 (0.028)	0.008 (0.980)
	Tauranga (TR)	7.590	-2.10**	-24.93***	0.001 (0.873)	0.006 (0.968)
	Wellington (WL)	7.691	-1.62*	-9.47***	0.001 (0.944)	0.011 (0.852)
	Nelson (NL)	7.994	-2.20**	-25.08***	0.001 (0.870)	0.006 (0.953)
	Total	7.797	-2.45***	-24.74***	0.001 (0.756)	0.004 (0.929)

Sum of squared deviations (*SSD*) and statistical significance (P_{SSD}) for the validity of the sudden expansion model, Harpending's Raggedness index (*Raggedness*) and its statistical significance (P_r) for the test of goodness-of-fit. *P*-values * <0.05, **<0.01, ***<0.001.

A qualitative test for recent population expansion was performed with mismatch distribution analysis, which analyses the distribution of pairwise base pair differences between DNA sequences. A population that has a stable size for a long time or that passed through a recent bottleneck is expected to have a multimodal and uneven pattern of mismatch distributions, whereas a recent expansion is expected to result in a unimodal and smooth distribution (Rogers, 1995; Rogers and Harpending, 1992; Schneider and Excoffier, 1999). The mismatch analysis of the COI, Cytb and concatenated sequences supports a sudden population expansion: the observed mismatch distributions for pooled populations closely match the expected values under a model of recent sudden population expansion, and mismatch distribution plots show a uni-model pattern for all the populations (Figure 5.6) for the pooled samples, and Supplementary Figure 7 for each population). The expansion model was statistically tested with the sum of the squared deviations (*SSD*) from the expected values

(Schneider and Excoffier, 1999) and Harpending's *raggedness* index (Harpending, 1994). For constant populations with multimodal mismatch distributions, the *raggedness* index provides high values as the sequences are diverse and few individuals share the same haplotypes. Both the *SSD* and *raggedness* values are small and non-significant for all the *P. maculata* populations and pooled samples (Table 5.7), which is another sign of population expansion. All these results suggest that population expansion is not specific to any population, but that all of them underwent a recent expansion. Selection rather than population expansion would yield similar results. However, as the star-like shape of the haplotype networks suggests population expansion (Schneider and Excoffier, 1999), the results of the neutrality tests are attributed to population expansion.

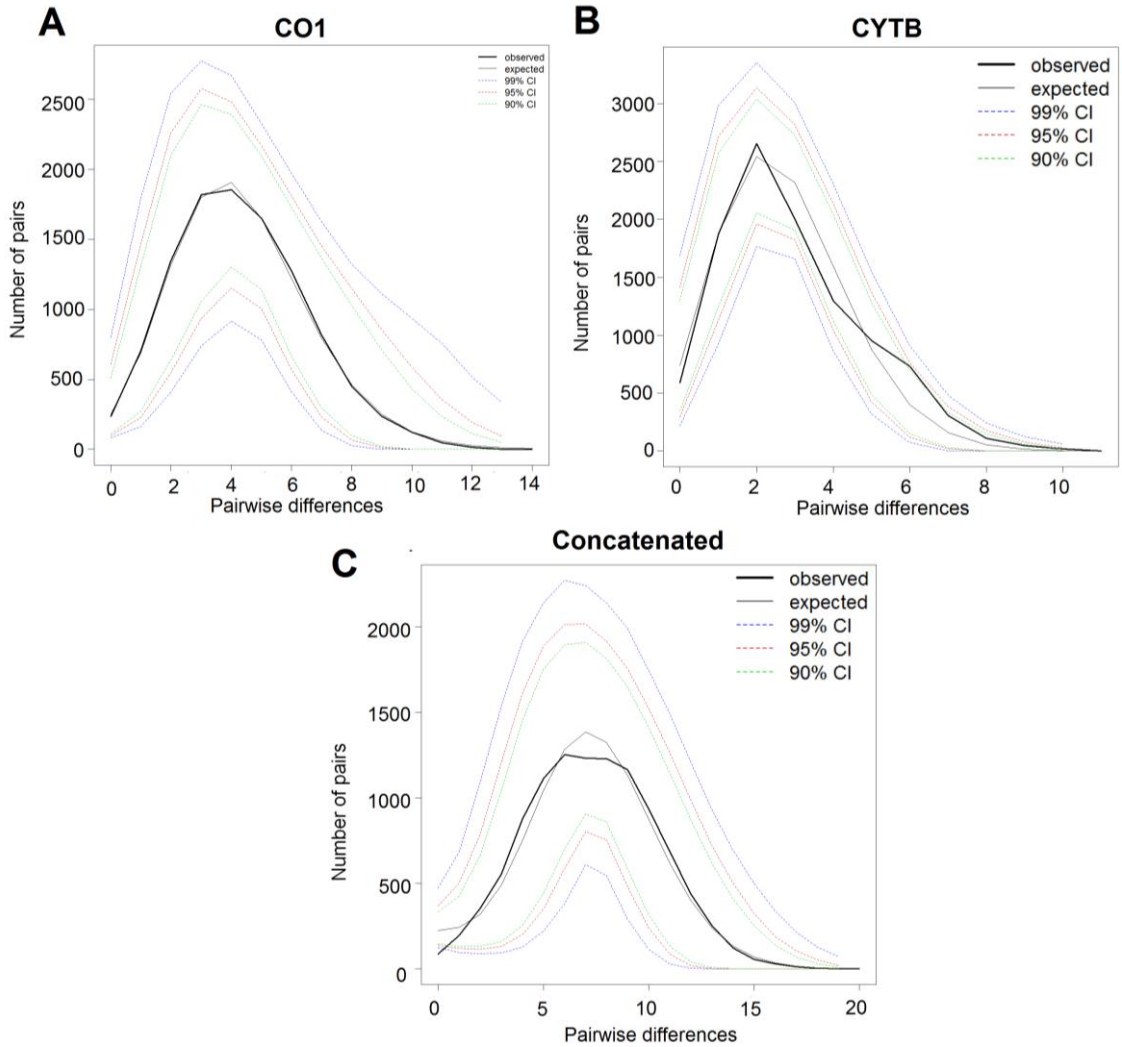


Figure 5.6 Mismatch distributions of pairwise base pair differences between the *P. maculata* mtDNA haplotypes. Observed and expected pairwise differences under the sudden expansion model for A) COI B) Cytb and C) concatenated COI and Cytb sequences. 156 individuals from all locations were pooled for the analysis.

The demographic parameter tau (τ), which represents the mutational time since expansion, was calculated after pooling all five populations as all of them exhibit signs of population expansion. τ was estimated at 2.703 (CI 95% 1.488–3.320) and 2.754 (CI 95%: 1.344–3.615) for the COI and Cytb data, respectively. The time at which the expansion event happened was calculated following the expression $t = \tau/2\ \mu k$, where μ is the mutation rate per site per generation, and k is the sequence length (Liao et al., 2010). The mutation rate was assumed to be 5.3% divergence/mya (average mutation rate for marine invertebrate COI, (Crandall et al., 2012) and one year of generation time was assumed for the calculation. The date of expansion was estimated at 44.2 kya (CI: 34.3–54.3 kya).

5.3 DISCUSSION

5.3.1 Genetic diversity

Analysis of mtDNA for 156 *P. maculata* individuals from various locations along New Zealand's coasts reveals high haplotype diversity for all the populations; the number of haplotypes for each location is close to the sample sizes. This may suggest that only a portion of the total diversity was sampled. This study revealed that *P. maculata* populations in NZ experienced a population expansion, which is evident in the star-like haplotype network that contains many private haplotypes (Schneider and Excoffier, 1999). As expansion leads to a huge number of haplotypes, sampling more individuals might have just led to discovery of other private haplotypes rather than discovery of the total diversity. Therefore, it is possible that haplotype diversity did not provide an accurate estimate of diversity. At the same time, nucleotide diversity is low for *P. maculata*: pairwise nucleotide diversity across individuals is 0.381% for the COI gene, which is less than the 2% threshold accepted to distinguish intra-specific from inter-specific variation (Hebert et al., 2003). One hypothesis was that *P. maculata* individuals from toxic and non-toxic regions are cryptic species. According to the mtDNA dataset, *P. maculata* populations around NZ show no evidence of cryptic species complexes.

5.3.2 Population structure

The existence of several haplotypes shared between all the sampling locations without creating a clear geographical structure, as well as low nucleotide distances between the haplotypes, suggests that all the populations are strongly connected. However,

haplotype-frequency-based F_{ST} analysis suggests that all the populations have diverged significantly from each other. In addition, the high proportion of private alleles within each population implies some degree of genetic distinctiveness in the populations. Nevertheless, only a weak structure was detected when the extent of the differentiation between the populations was estimated based on nucleotide distances between haplotypes. The AKL population is weakly but significantly differentiated from the TP and TR populations. Similarly, the TR and NL populations are weakly but significantly differentiated from each other. However, detection of the differentiation between all these populations was dependant on the method used.

The findings of the F_{ST} and Φ_{ST} analyses appear discordant. This is because the Φ_{ST} calculation includes both the haplotype frequencies and bp distances between the haplotypes (Excoffier et al., 1992) whereas the F_{ST} calculation includes only haplotype frequencies and treats all haplotypes equally distant (Nei, 1987). In this regard, for recently diverged populations, genetic drift for just a few generations may lead to significant F_{ST} values, whereas mutation accumulation is needed for observation of significant Φ_{ST} values (Penant et al., 2013). Therefore, genetic distance between the haplotypes may not provide sufficient resolving power to reveal genetic connectivity of populations when the mutation rate is much smaller than the migration rate (Bird et al., 2011). F_{ST} is therefore more powerful than Φ_{ST} for detecting significant differentiation in recently divergent populations (Penant et al., 2013). Although the mutation rate for mtDNA data for *P. maculata* is unknown, the low nuclear diversity detected here could be interpreted as the signal of a low mutation rate. In this case, the strong population structure suggested by the haplotype-based analysis is more probable than the low structure suggested by the Φ_{ST} analysis for *P. maculata* populations. Therefore, I conclude that the mtDNA differentiation is high between the sampled *P. maculata* populations. It is worth noting that distance based methods would provide meaningful estimates of population structure when hypervariable markers such as microsatellites are utilized (Bird et al., 2011).

The differentiation of the AKL population from the TP and TR populations, which was revealed by distance-based methods, is interesting considering the close proximity between these localities. AKL is the biggest trading city in New Zealand, and Auckland Harbour connects the city to other cities within NZ as well as to other countries where

P. maculata inhabits. Individual sea slugs from other locations may be transferred from and into AKL from other cities more often compared to in neighbouring cities due to the water activities including shipment. This may explain why the AKL population is differentiated from its neighbouring locations.

5.3.3 Incongruence between mtDNA and microsatellite data

As discussed in Chapter 4, the microsatellite data revealed a highly significant population structure. Analysis of mtDNA sequences shows a high level of genetic differentiation among all the sampling localities. This finding is incongruent with that of microsatellite data as the later suggests a strong differentiation between just the Northern and Southern clusters.

In phylogenetic or population genetic analysis where both nuDNA and mtDNA markers are used, the mitochondrial data is mostly congruent with the nuDNA data (Toews and Brelsford, 2012), although discordance between these two marker types has been identified by other researchers (Apple et al., 2010; Kai et al., 2002; Mila et al., 2010; Scheffer and Hawthorne, 2007). There are various possibilities that could explain this discrepancy. Sex-biased asymmetry is one possibility, but this seems an unlikely explanation given that *P. maculata* is an obligate out-crossing simultaneous hermaphrodite that often copulates reciprocally. The non-recombinant nature of mtDNA may bring about a situation where shared haplotypes can be maintained across populations that have recently diverged, whereas the shared history maybe disguised due to recombination events in nuDNA. Consequently, mtDNA may give a glimpse into the evolutionary history of the species before any divergence happened (Yang and Kenagy, 2009), which may be the case for this study. However, a more probable explanation for the mitonuclear discordance is that two marker types provide information about different timeframes of the species' history. mtDNA is considered a "leading indicator", whereas nuclear genes are considered a "lagging indicator" (Barrowclough and Zink, 2009; Funk and Omland, 2003; Toews and Brelsford, 2012; Zink and Barrowclough, 2008). In this context, as observed for *P. maculata* data, it is common to observe more mtDNA structure than nuDNA structure, because lineage sorting would have been completed for mtDNA before nuDNA, which should be reflected as more pronounced structure in mtDNA markers for populations that have diverged recently (Zink and Barrowclough, 2008).

mtDNA genes behave as a linked single locus and thus comprise a single evolutionary unit, whereas the microsatellite markers used in this study are from multiple unlinked nuclear loci. Therefore, it is best to keep in mind that combined information obtained from multiple unlinked loci may provide more robust information about the population structure and the evolutionary processes shaping the species (Edwards and Bensch, 2009) although ILS can explain the mitonuclear discordant observed for *P. maculata*.

One limitation of this study is the gaps in the sampling locations, which has prevented analysis of population structure at a finer scale. This limitation may have also led to mitochondrial discordance with the true population structure remaining hidden. A more detailed sampling regime would be required to eliminate this possibility.

5.3.4 A population expansion dating back to last glacial maxima

One important finding in the haplotype network is its star-like structure with high haplotype diversity with most haplotypes having very low frequency; the haplotypes have been differentiated from the shared haplotypes with only a few base pairs. Such a pattern is evidence of a recent population expansion following a bottleneck (Slatkin and Hudson, 1991). A recent population expansion is supported by several indexes that were used to investigate demographic history including a uni-modal mismatch distribution pattern. The approximate date of overall population expansion for *P. maculata* was estimated at 44.2 kya (*CI*: 24.3–54.3kya) assuming a 5.3% divergence rate for the COI gene (average invertebrate rate for marine invertebrate COI; (Crandall et al., 2012). Although this molecular clock rate may not reflect the intra-species *P. maculata* rate (it is an approximate estimate), the estimated expansion time falls into the late Pleistocene era (~110–15 kya), which is characterised by cyclic climatic oscillations between warm and cold periods (Allcock and Strugnell, 2012; Rosetti et al., 2007). This era is associated with massive reductions in population size, and accordingly a reduction in genetic diversity for many marine species. Shallow-water species that inhabit intertidal or shallow subtidal areas were primarily affected as decreased sea levels destroyed their available habitat (Allcock and Strugnell, 2012; Norris and Hull, 2012). During the last glacial maxima (LGM, 26.5–18 kya), ice sheets were at their maximum size, and the sea level dropped as low as 120 m along New Zealand's coasts (Lewis et al., 1994; Trewick and Bland, 2012)When the LGM ended, a rise in sea levels made the large benthic habitats available again (Allcock and Strugnell, 2012). Such star-like mtDNA haplotype

networks have been also observed for other NZ marine organisms, such as the sea urchin *Evechinus chloroticu*s (Nagel et al., 2015) and abalone *Haliotis iris* (Will et al., 2011). Therefore, population expansion following a bottleneck is common for NZ marine organisms.

Although Willan (1983) states that *P. maculata* can be found throughout New Zealand, the twice-yearly harbour surveys conducted by the National Institute of Water and Atmospheric Research have failed to find this species in Christchurch and Otago regions historically (Chris Woods and Kimberly Seaward, personal communications). There are records of the species in Wanganui and Christchurch dating back to 2009 and 2011, respectively (Te Papa collections), yet the population size of this species is not expected to be high given that the organism has not been regularly identified in these regions. Similarly, Global Dive (dive training, operation and global dive destination travel agent, Auckland) have not found *P. maculata* specimens on Goat Island or in the Poor Knights Islands Marine Reserves (Anna Clague, Andrew Simpson, personal communications). This may suggest that the organism has specific habitat preferences. *P. maculata* is a member of the Pleurobranchidae family, which is distributed worldwide in tropical and temperate waters, and they are rarely found in cold areas (Willan, 1983). Lower sea temperatures at lower latitudes may be the reason why the organism is found rarely in the southern part of the South Island. This habitat restriction may also indicate that there has been a spatial expansion of *P. maculata* populations towards to the southern coasts of New Zealand since the increase in sea temperatures and concomitant rise in sea levels. Alternatively, behavioural patterns of *P. maculata* may have played a role in the scarcity of samples obtained during the course of my study. Another species from the same genus, *Pleurobranchaea californica*, is known to be nocturnally active and to hide under rocks and in old sunken tires during day. Nocturnal activities have been noted also for other *Pleurobranchaea* species (Gillette, 2014). If *P. maculata* also has a nocturnal activity, then divers might have difficulties in finding the individuals as the surveys that in NZ were carried out during day.

Being an organism that inhabits both tropical and temperate waters, *P. maculata* possibly experienced a bottleneck during the glacial cycles due to the changes in temperature and sea levels (Weaver et al., 1998), and community composition (Grosberg and Cunningham, 2001). This might have resulted in a decrease in the range

of available mtDNA haplotypes, which is evident in the closely related mtDNA haplotypes. It is worth noting that the network haplotype for *P. maculata* was also "diffuse" due to presence of a few main haplotypes from which other haplotypes descent, high haplotype diversity and reticulations between different haplotypes, indicating that the bottleneck did not lead to a drastic decrease in haplotypic diversity. Also possibly relevant is the three-week pelagic larval stage (Gardner et al., 2010; Gibson, 2003; Wood et al., 2012b) that confers a high dispersal capacity on the species and might have allowed persistent mtDNA haplotypes to expand to new areas after the LGM (Allcock and Strugnell, 2012). Interestingly, "diffuse" haplotype networks are associated with species whose habitat range extends to the deep sea. For example, a deep-sea shrimp, *Nematocarcinus lanceopes* which lives at depths below 4000 m, was found to have a haplotype network that is very similar to that of *P. maculata* (Raupach et al., 2010). Such species are thought to take refuge in deeper areas away from continental shelves that were affected by drops in the sea level (Allcock and Strugnell, 2012). Although *P. maculata* generally inhabits shallow and sandy subtidal areas, it can also be found in intertidal areas deeper than 300 m (Willan, 1983). Its diverse range of habitats may help *P. maculata* to maintain high haplotypic diversity even though subtidal populations might have undergone a bottleneck.

When mitonuclear discordance in *P. maculata* is evaluated in light of recent population expansion, the more pronounced differentiation revealed by mtDNA markers suggests that isolation and differentiation of populations have become more pronounced in recent times following the demographic expansion. Gene flow between the NZ *P. maculata* populations was reduced after its range extended to new areas. Therefore, mtDNA has most probably caught the more recent events, as being the "leading markers. (Zink, 2010). The hypothesis of population expansion was also tested with microsatellite data (See Chapter 4, section 4.2.5); however, a significant sign of expansion was not detected. Microsatellites may not have captured the sign of the expansion event because, being nuDNA markers, they are less sensitive to bottlenecks due to larger N_e (Zink and Barrowclough, 2008), and also they may not show the evidence of the recently happened expansion event due to ILS.

Although there is strong support for divergence after population expansion for the NZ *P. maculata* populations, the opposite scenario may also work. During the glacial cycles,

the sea slug populations might have survived in multiple isolated shelf refugia where vicariant events happened due to genetic drift and/or adaptation. More specifically, the northern populations could have become isolated from the southern populations during the glacial cycles, and this isolation may be reflected in the microsatellite data. Population bottlenecks in shallow refugia are expected to decrease haplotype diversity (González-Wevar et al., 2011), which could underpin the closely related mtDNA haplotypes. Again, the nuDNA would be affected by bottlenecks, but to a lesser extent. In this case, reunion of the populations might have resulted in discordance between variations in the mtDNA and nuDNA (Chavez et al., 2011; Toews and Brelsford, 2012). One reason for this discordance can be introgressive hybridisation, i.e. where the alleles of one divergent clade spread within the other clade (Toews and Brelsford, 2012). As the mitochondrial genome is more susceptible to introgression than nuclear loci, mtDNA haplotypes from one population might have replaced the haplotypes of the other population and have erased the signals of divergence in the mtDNA, while the signals may have been maintained in the nuDNA (Ballard and Whitlock, 2004). There is some evidence to distinguish introgression from incomplete lineage sorting (ILS), although it is difficult to distinguish between the two. For example, ILS is expected to leave an unpredictable biogeographical pattern (Toews and Brelsford, 2012). On the other hand, introgression leads to biogeographic discordance between the populations, in which the relationship between the populations is identified differently based on marker types (Toews and Brelsford, 2012; Zink and Barrowclough, 2008). Our case is closer to the former, which supports incomplete linage sorting for mtDNA. An alternative explanation is that population expansion may have obscured signs of an ancient population divergence event (Veeramah et al., 2012), and this obscuring event may affect mtDNA to a larger extant compared to nuclear DNA due to its smaller N_e. In this case, relatively slowly evolving microsatellites might have maintained the signal from a past divergence event, whereas fast evolving mtDNA loses it after a population expansion event (Eytan and Hellberg, 2010). Although this possibility of ancient vicariance cannot be ruled out for *P. maculata*, the lower sea temperatures during the LGM make the survival of *P. maculata* individuals in the southern seas unlikely.

It is worth mentioning that there are some difficulties in using microsatellite markers in phylogeographic studies. Firstly, there is not a direct phylogenetic relationship between the numbers of repeats because the coalescence of the repeat numbers cannot be

postulated (Ellegren, 2004; Rubinsztein et al., 1995). Secondly, two individuals may have the same number of repeats due to homoplasy as microsatellites have high evolutionary rates. On the other hand, sequence data may resolve the question of an evolutionary relationship between the individuals. Another drawback of using microsatellites is that the analysis is mostly dependant on allele frequencies, which makes it difficult to compare the data directly with mtDNA sequencing data (Brito and Edwards, 2009; Lee and Edwards, 2008). In contrast, sequence data from nuclear genes can be directly compared to the mtDNA data as the same statistics could be used for both mtDNA and nuDNA sequencing data (Brito and Edwards, 2009; Zink, 2010). Therefore, sequencing multiple unlinked nuclear loci and comparing them with mtDNA using coalescence methods (Brito and Edwards, 2009; Zink, 2010) may provide a more comprehensive understanding of the evolutionary history of *P. maculata*. However, due to their different coalescent times, the interpretation of the data from multiple nuclear coding sequences could be complicated (Brito and Edwards, 2009). An appropriate -and a popular- approach is using non-coding nuclear markers, but especially intronic sequences (e.g. Bensch et al., 2006; Palumbi and Baker, 1994). Designing primers from highly conserved exonic regions that flank introns confers an ease for the process (Brito and Edwards, 2009). By sequencing multiple intronic regions, the date of vicariant event(s) in *P. maculata* can be revealed, leaving little room for doubt.

5.3.5 Human impact on the distribution of *P. maculata*

Invasive or introduced species can assist the population distribution and/or abundance of native species (Rodriguez, 2006), which may be the case for *P. maculata*. Taylor et al. (2014) identified large populations of *P. maculata* on near-shore beds of *Musculista senhousia* in Narrow Neck in 2009 when dog poisoning events happened on Auckland's beaches. *M. senhousia*, an invasive bivalve, has established populations in Auckland since the 1970s, and it forms large beds in sub- and intertidal areas in Hauraki Gulf, Manukau Harbour and Whangarei Harbour (Crooks, 2002). Die-off of *M. senhousia* populations led to a decrease in *P. maculata* numbers. Similarly, establishment of off-shore mussel farms in Tasman Bay (Nelson, New Zealand), where *Perna canaliculus* (Greenshell™ mussel) and *Mytilus edulis* (blue mussel) are cultured, created a new habitat for *P. maculata*: high density populations were found beneath these mussel farms (Taylor et al., 2014). The gut content of *P. maculata* confirms that they feed on these bivalve species, which are available at that site. Taylor et al. (2014) also observed

that adults lay eggs on 3D mussel and shell structures beneath the mussel farms, and juvenile individuals are recruited exclusively to the farms. These findings show that, despite the restricted geographical distribution of *P. maculata* in NZ, the distribution and density of the *P. maculata* populations are affected by spatial and temporal alterations in their habitat and food supplies. Additionally, an outbreak of an *P. maculata* was identified in Argentinean waters in 2009, and the number of the individuals has increased since then (Farias et al., 2015). This invasion confirms that human activities result in the transportation of sea slug individuals to the different geographical areas that act as new habitats for this species.

To conclude, this chapter shows that the all sampled *P. maculata* populations havediverged significantly from each other, and all have undergone a population expansion dating back to the late Pleistocene era. This finding is seemingly not congruent with the microsatellite data, which show two significantly differentiated populations indicative of a north-south disjunction between the *P. maculata* populations. One explanation for this seemingly paradoxical situation is the different evolutionary rates of the two markers types. Microsatellites, being lagging markers, may not have had sufficient time to accumulate diversity sufficient to detect vicariant events. Alternatively, the geographical gaps in my sampling may have led to the impression of discordance, when in fact there is no. The latter, however, seems unlikely. It is entirely possible that populations have become geographically isolated more recently, perhaps due to migration events facilitated by a rise in sea levels, or improvement in the environmental conditions that created new habitats for the species. Further, the role of humans assisted migration cannot be overlooked. The recent isolation shows evidence of a decrease in gene flow and differentiation between the populations.

CHAPTER 6: CONCLUSION AND PERSPECTIVES

INTRODUCTION

Pleurobranchaea maculata, which is native to the western and south Pacific, is known to contain high concentrations of tetrodotoxin (TTX). Significant individual and seasonal variations in TTX concentrations have been observed for populations from various regions of the North Island, whereas no or trace amounts of TTX were found in individuals from South Island populations (McNabb et al., 2010; Wood et al., 2012b). The origin of TTX in *P. maculata* and the reasons underpinning variation are still not fully understood. My PhD project aimed to study the diversity and genetic structure of *P. maculata* individuals from five localities representing four regions in the North Island (Ti-Point-TP, Auckland-AKL, Tauranga-TR and Wellington-WL) and one region from the South Island (Nelson-NL) using twelve nuclear microsatellite markers, and sequences from two mitochondrial genes (COI and Cytb). I analysed the total diversity, partitioning of variation between the populations, the degree of gene flow between the populations and signs of population demographic events using data obtained from both microsatellite and mtDNA markers. The nature of the two marker types helped to reveal the population processes that affected *P. maculata*'s distribution from unique perspectives.

MAIN FINDINGS

The microsatellite data, which reject panmixia between *P. maculata* populations, reveal two significantly differentiated genetic pools that cluster TP, AKL and TR into one group (the Northern cluster), and WL and NL into another group (the Southern cluster). Interestingly, this finding correlates with regional variations in TTX concentration. The North Island populations containing highly toxic individuals are significantly different from the WL and NL populations, which harbour either slightly toxic or non-toxic individuals, respectively. The divergence correlates with the biogeographical barrier on the east coast of the North Island between 37–39°S, which stems most probably from East Cape eddies (Figure 1.2A) (Gardner et al., 2010; Ross et al., 2009). Alternatively, the structure pattern can be explained by the historical connectivity of the North and South Islands (Lewis et al., 1994) that has been used to explain an east-west disjunction

and connectivity of the southern/western North Island and the northern South Island populations for some NZ marine organisms (Keeney et al., 2013; Stevens and Hogg, 2004). Weak differentiation is observed between the WL and NL populations, which can perhaps be explained by the body of water that is Cook Strait (Ross et al., 2009). On the other hand, only very weak conclusions could be drawn due to the limited sampling, because an IBD model can also explain the population structure pattern. mtDNA data revealed high haplotype diversity. Although several haplotypes are shared between the populations, each population pair is significantly differentiated from each other with regard to the distribution of haplotype frequencies and high proportions of private haplotypes in each population. Despite the high haplotype diversity, the haplotypes are very closely related to each other, and differentiation estimators based on the base pair distances between haplotypes reveal the lack of a clear regional structure. Here, only the AKL population is weakly divergent from the TP and TR populations. mtDNA data also show that *P. maculata* experienced a recent population expansion following a bottleneck (Slatkin and Hudson, 1991). The approximate date of the expansion corresponds to the last glacial period.

The ostensibly incongruous results from the mtDNA and microsatellite markers are explained by the fact that each marker type reveals information from different time frames of the species' history (Funk and Omland, 2003; Toews and Brelsford, 2012; Zink, 2010; Zink and Barrowclough, 2008). The mtDNA data, which provide information about the species' more recent past, show that differentiation between the *P. maculata* populations in NZ became more pronounced during times that are more recent. A recent bottleneck is thought to have decreased the haplotype diversity moderately during the Pleistocene cycles. The rise in sea levels and consequently the formation of new habitats, and increased sea temperatures allowed a demographic expansion for the *P. maculata* populations that were possibly geographically restricted. Microsatellite markers, which are expected to provide species' more distant past, are likely to show a less population structure due to incomplete lineage sorting. However, it is worth noting that the limited sampling might have led to the mitonuclear discordance as it prevented a detailed fine-scale structure analysis.

Although beyond the scope of this thesis, this study shows a correlation with geographical variations in TTX concentrations and the population structure that is

revealed by the microsatellite data, but not by mtDNA data. The fact that individuals from non-toxic regions can accumulate high TTX-concentrations when they are fed on a TTX-containing diet, and the identification of TTX-containing food sources in some regions support a dietary origin of TTX in *P. maculata*. Yet, a common toxic dietary source that is associated with individuals containing high TTX concentrations has not been identified in other regions (Khor et al., 2014; McNabb et al., 2010). Ultimately, a bacterial origin of TTX via symbiosis may explain the presence of TTX in *P. maculata*, as well as in other toxic organisms. Oceanographic barriers that result in the north-south disjunction in the *P. maculata* populations (or in other NZ organisms) and/or environmental differences may also determine the distribution of such symbiont microorganisms or their capacity to produce TTX. This would indirectly result in a correlation with the toxicity and population genetic structure in *P. maculata*. However, the limited sampling might have also given an impression of the correlation although it does not exist.

FUTURE DIRECTIONS

Microsatellite markers suggested a north-south disjunction, which could be attributed to biogeographical barriers in NZ or simply to isolation by descent. In this study, the origin of population structure could not be found, possibly due to the limited sample size. A more detailed sampling regime that includes samples from various locations including several points on the east and west coasts of both islands can help to identify the exact location and the origin of the disjunction. This will also help to find if the mitonuclear incongruence is real or not.

This research may suggest a correlation between population differentiation and TTX concentrations when microsatellite data are taken into consideration. However, the mtDNA data, which is expected to give information about a more recent history of the species, does not show such a correlation. Therefore, the phenotype-genotype correlation seems to be solely a reflection of environmental factors that are most probably behind the differential distribution of TTX-producing microorganisms. Despite several attempts, no evidence supporting the bacterial origin of TTX has been found for *P. maculata* (Chau et al., 2013; Salvitti et al., 2015b). These studies focused on culturable bacterial species. However, the ultimate TTX-producing microorganisms may not be culturable, or they may need host organisms to produce TTX (Chau et al., 2013;

Salvitti et al., 2015b). A metagenomic approach targeting the microflora of *P. maculata* individuals but also environmental DNA that is sampled from toxic and non-toxic regions will help to identify the differences in bacterial community structure between distinct geographical regions (Gerwick and Moore, 2012). Temporal variations in TTX concentrations have been observed for TTX-associated *P. maculata* populations and *Styochplona* sp. A metagenomic analysis that is enlarged to a temporal scale may clarify how the structure of the bacterial communities changes throughout the year. Metagenomic approaches can also help to identify biosynthetic genes or pathways that are suggested for biosynthesis of TTX (Chau et al., 2011; Chau et al., 2013).

CONCLUSION

My research provides significant initial insights into the population structure of this Heterobranch gastropod, which was recorded as the first species to contain TTX within this subclass (McNabb et al., 2010). The evolutionary history of TTX-containing taxa is important to understand the biosynthetic pathways of this toxin. The analysed microsatellite data show a north-south discontinuity that is concordant with geographical variations in TTX concentrations, and with the biogeographic barriers on NZ's coasts or a stepping stone model (Gardner et al., 2010; Ross et al., 2009). The mtDNA data are incongruent with the microsatellite data as the later reveal a more structured population where all the sampling locations are significantly differentiated from each other. If I interpret the seemingly contradictory data, the NZ *P. maculata* populations were once more connected until a recent population expansion that happened in the late Pleistocene age. The populations have become more geographically isolated in more recent times perhaps due to migration events facilitated by a rise in sea levels that created new habitats for the species as well as the arrival of humans. This recent isolation is evident in the mtDNA data, which show a higher level of differentiation between populations. When viewed from this perspective, this study provides a good example of mitonuclear discordance where microsatellite data reveal a more distant population process than mtDNA as being "lagging markers" (Zink and Barrowclough, 2008).

REFERENCES

Abdelkrim, J., B. Robertson, J.A. Stanton, and N. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques*. 46:185-192.

Allcock, A.L., and J.M. Strugnell. 2012. Southern Ocean diversity: new paradigms from molecular ecology. *Trends in Ecology & Evolution*. 27:520-528.

Anderson, M., R.N. Gorley, and R.K. Clarke. 2008. Permanova+ for Primer: Guide to Software and Statisticl Methods.

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 26:32-46.

Anderson, M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*. 62:245-253.

Anderson, M.J., and T.J. Willis. 2003. Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. *Ecology*. 84:511-525.

Apple, J.L., T. Grace, A. Joern, P.S. Amand, and S.M. Wisely. 2010. Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. *Molecular ecology*. 19:4012-4028.

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

Asakawa, M., T. Toyoshima, K. Ito, K. Bessho, C. Yamaguchi, S. Tsunetsugu, Y. Shida, H. Kajihara, S.F. Mawatari, T. Noguchi, and K. Miyazawa. 2003. Paralytic toxicity in the ribbon worm Cephalothrix species (Nemertea) in Hiroshima Bay, Hiroshima Prefecture, Japan and the isolation of tetrodotoxin as a main component of its toxins. *Toxicon*. 41:747-753.

Ayers, K., and J. Waters. 2005. Marine biogeographic disjunction in central New Zealand. *Marine Biology*. 147:1045-1052.

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

Ballard, J.W., and M.C. Whitlock. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*. 13:729-744.

REFERENCES

Balloux, F., and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology*. 11:155-165.

Bandelt, H.J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*. 16:37-48.

Barrett, S. 2014. Evolution of mating systems: outcrossing versus selfing. *In: Losos J, ed. The Princeton guide to evolution. Princeton, NJ, USA: Princeton University Press*:356–362.

Barrowclough, G.F., and R.M. Zink. 2009. Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. *Molecular Ecology*. 18:2934-2936.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*. 312:570-572.

Bensch, S., D.E. Irwin, J.H. Irwin, L. Kvist, and S. Åkesson. 2006. Conflicting patterns of mitochondrial and nuclear DNA diversity in Phylloscopus warblers. *Molecular Ecology*. 15:161-171.

Bernal-Ramírez, J., G. Adcock, L. Hauser, G. Carvalho, and P. Smith. 2003. Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents. *Marine Biology.* 142:567-574.

Bhargava, A., and F.F. Fuentes. 2010. Mutational dynamics of microsatellites. *Molecular Biotechnology*. 44:250-266.

Bilgin, R. 2007. Kgtests: a simple Excel Macro program to detect signatures of population expansion using microsatellites. *Molecular Ecology Notes*. 7:416-417.

Bird, C.E., S.A. Karl, P.E. Smouse, and R.J. Toonen. 2011. Detecting and measuring genetic differentiation. *Phylogeography and population genetics in Crustacea*. 19:31-55.

Bonferroni, C. 1936. Statistical theory of classification and calculation of the probability. *Publication of the R Institute Superiore of the Science of Economy and Commerce of Florence*. 8:3-62.

Bradbury, I.R., B. Laurel, P.V. Snelgrove, P. Bentzen, and S.E. Campana. 2008. Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society of London B: Biological Sciences*. 275:1803-1809.

Brinkman, F.S., and D.D. Leipe. 2001. Phylogenetic analysis. *Bioinformatics: a practical guide to the analysis of genes and proteins*. 2:349.

REFERENCES

Brito, P.H., and S.V. Edwards. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*. 135:439-455.

Brodie, J. 1960. Coastal surface currents around New Zealand. *New Zealand Journal of Geology and Geophysics*. 3:235-252.

Brodie Jr, E.D., B. Ridenhour, E. Brodie III, and J. Wiens. 2002. The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution*. 56:2067-2082.

Burley, N. 1983. The meaning of assortative mating. *Ethology and Sociobiology*. 4:191-203.

Chambers, G.K., and E.S. MacAvoy. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 126:455-476.

Chapuis, M.P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*. 24:621-631.

Charlesworth, D., and S.I. Wright. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics & Development*. 11:685-690.

Chau, R., J.A. Kalaitzis, and B.A. Neilan. 2011. On the origins and biosynthesis of tetrodotoxin. *Aquatic Toxicology*. 104:61-72.

Chau, R., J.A. Kalaitzis, S.A. Wood, and B.A. Neilan. 2013. Diversity and biosynthetic potential of culturable microbes associated with toxic marine animals. *Marine Drugs*. 11:2695-2712.

Chavez, A.S., C.J. Saltzberg, and G.J. Kenagy. 2011. Genetic and phenotypic variation across a hybrid zone between ecologically divergent tree squirrels (Tamiasciurus). *Molecular Ecology*. 20:3350-3366.

Cheng, C.A., D.F. Hwang, Y.H. Tsai, H.C. Chen, S.S. Jeng, T. Noguchi, K. Ohwada, and K. Hasimoto. 1995. Microflora and tetrodotoxin-producing bacteria in a gastropod, Niotha clathrata. *Food and Chemical Toxicology*. 33:929-934.

Clarke, K., and R. Gorley. 2006. User manual/tutorial. *PRIMER-E Ltd., Plymouth*:93.

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

Crandall, E.D., E.J. Sbrocco, T.S. DeBoer, P.H. Barber, and K.E. Carpenter. 2012. Expansion dating: calibrating molecular clocks in marine species from

expansions onto the Sunda Shelf following the last glacial maximum. *Molecular Biology and Evolution*. 29:707-719.

Crooks, J.A. 2002. Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. *Oikos*. 97:153-166.

Csencsics, D., S. Brodbeck, and R. Holderegger. 2010. Cost-effective, species-specific microsatellite development for the endangered Dwarf Bulrush (Typha minima) using next-generation sequencing technology. *Journal of Heredity*. 101:789-793.

Czekanowski, J. 1909. Zur differentialdiagnose der Neandertalgruppe. Friedr. Vieweg & Sohn.

Dao, H.V., Y. Takata, S. Sato, Y. Fukuyo, and M. Kodama. 2009. Frequent occurrence of the tetrodotoxin-bearing horseshoe crab *Carcinoscorpius rotundicauda* in Vietnam. *Fisheries Science*. 75:435-438.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*:1-38.

Donald, K.M., M. Kennedy, and H.G. Spencer. 2005. Cladogenesis as the result of long-distance rafting events in South Pacific topshells (Gastropoda, Trochidae). *Evolution*. 59:1701-1711.

Drenth, A. 1998. Practical Guide to Population Genetics. *The University of Queensland, Australia*.

Earl, D.A., and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 4:359-361.

Edwards, S., and S. Bensch. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*. 18:2930-2933; discussion 2934-2936.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*. 5:435-445.

Endler, J.A. 1977. Geographic variation, speciation, and clines. *Princeton University Press.*

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

REFERENCES

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

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

Eyre-Walker, A., and P.D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nature Review Genetics*. 8:610-618.

Eytan, R.I., and M.E. Hellberg. 2010. Nuclear and mitochondrial sequence data reveal and conceal different demographic histories and population genetic processes in caribbean reef fishes. *Evolution*. 64:3380-3397.

Farias, N.E., S. Obenat, and A.B. Goya. 2015. Outbreak of a neurotoxic side-gilled sea slug (Pleurobranchaea sp.) in Argentinian coasts. *New Zealand Journal of Zoology*. 42:51-56.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*. 3:294-299.

Frankham, R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology*. 10:1500-1508.

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

Fuhrman, F.A. 1986. Tetrodotoxin, tarichatoxin, and chiriquitoxin: historical perspectives. *Annals of the New York Academy of Sciences*. 479:1-14.

Funk, D.J., and K.E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*. 34:397-423.

Galarza, J.A., J. Carreras-Carbonell, E. Macpherson, M. Pascual, S. Roques, G.F. Turner, and C. Rico. 2009. The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences*. 106:1473-1478.

Gardner, J., J. Bell, H. Constable, D. Hannan, P. Ritchie, and G. Zuccarello. 2010. Multi-species coastal marine connectivity: a literature review with recommendations for further research. Ministry of Fisheries.

REFERENCES

Geffeney, S.L., and C. Ruben. 2006. The structural basis and functional consequences of interactions between tetrodotoxin and voltage-gated sodium channels. *Marine Drugs*. 4:143-156.

Gerwick, W.H., and B.S. Moore. 2012. Lessons from the Past and Charting the Future of Marine Natural Products Drug Discovery and Chemical Biology. *Chemistry & Biology*. 19:85-98.

Ghiselin, M.T. 1969. The evolution of hermaphroditism among animals. *Quarterly Review of Biology*:189-208.

Gibson, G.D. 2003. Larval development and metamorphosis in *Pleurobranchaea maculata*, with a review of development in the notaspidea (Opisthobranchia). *The Biological Bulletin.* 205:121-132.

Gillette, R. 2014. Pleurobranchaea. *Scholarpedia*. 9:3942.

González-Wevar, C.A., B. David, and E. Poulin. 2011. Phylogeography and demographic inference in Nacella (Patinigera) concinna (Strebel, 1908) in the western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*. 58:220-229.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3).

Goudet, J., N. Perrin, and P. Waser. 2002. Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Molecular Ecology*. 11:1103-1114.

Grosberg, R., and C. Cunningham. 2001. Genetic structure in the sea. *Marine Community Ecology*. 61-84.

Grove, S. 2015. A Guide to the Seashells and other Marine Molluscs of Tasmania. *Available at: www. mollusc softasmania. net*.

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin, and R.J. Petit. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources*. 11:591-611.

Hamrick, J.L., M.W. Godt, A.H. Brown, M.T. Clegg, A. Kahler, and B. Weir. 1990. Allozyme diversity in plant species. *Plant Population Genetics, Breeding, And Genetic Resources.* 43-63.

Hanifin, C.T. 2010. The chemical and evolutionary ecology of tetrodotoxin (TTX) toxicity in terrestrial vertebrates. *Marine Drugs*. 8:577-593.

Hanifin, C.T., E.D. Brodie, Jr., and E.D. Brodie, III. 2008. Phenotypic mismatches reveal escape from arms-race coevolution. *Plos Biology*. 6:471-482.

REFERENCES

Hanifin, C.T., M. Yotsu-Yamashita, T. Yasumoto, E.D. Brodie, and E.D. Brodie. 1999. Toxicity of dangerous prey: Variation of tetrodotoxin levels within and among populations of the newt *Taricha granulosa*. *Journal of Chemical Ecology*. 25:2161-2175.

Harpending, H.C. 1994. Signature of ancient population-growth in a low-resolution mitochondrial-dna mismatch distribution. *Human Biology*. 66:591-600.

Hartl, D.L., and A.G. Clark. 2007. *Principles of Population Genetics. Sinauer Associates Sunderland.*

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

Heath, R. 1982. What drives the mean circulation on the New Zealand west coast continental shelf? *New Zealand Journal of Marine aAnd Freshwater Research*. 16:215-226.

Hebert, P.D.N., S. Ratnasingham, and J.R. deWaard. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B-Biological Sciences*. 270:S96-S99.

Hedgecock, D., P.H. Barber, and S. Edmands. 2007. Genetic approaches to measuring connectivity. *Oceanography (Washington DC)*. 70-79.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. *Evolution*. 59:1633-1638.

Hedrick, P.W. 2011. Genetics of populations. *Jones & Bartlett Learning, Sudbury, MA.*

Hickey, A.J., S.D. Lavery, D.A. Hannan, C.S. Baker, and K.D. Clements. 2009. New Zealand triplefin fishes (family Tripterygiidae): contrasting population structure and mtDNA diversity within a marine species flock. *Molecular Ecology*. 18:680-696.

Holsinger, K. 2012. Tajima's D, Fu's Fs, Fay and Wu's H, and Zeng E et al.'s. *In: Holsinger K (ed) Lecture notes in population genetics*.:239–244.

Holsinger, K.E., and B.S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*. 10:639-650.

Huson, D.H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. 23:254-267.

REFERENCES

Huson, D.H., and C. Scornavacca. 2011. A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution.* 3:23-35.

Hwang, D., S. Lu, and S. Jeng. 1991. Occurrence of tetrodotoxin in the gastropodsRapana rapiformis and *R. venosa venosa*. *Marine Biology.* 111:65-69.

Hwang, D.F., O. Arakawa, T. Saito, T. Noguchi, U. Simidu, K. Tsukamoto, Y. Shida, and K. Hashimoto. 1989. Tetrodotoxin-producing bacteria from the blue-ringed octopus Octopus maculosus. *Marine Biology.* 100:327-332.

Hwang, D.F., and T. Noguchi. 2007. Tetrodotoxin poisoning. *Advances in Food and Nutrition Research.* 52:141-236.

Jacobs, L.L., C. Strganac, and C. Scotese. 2011. Plate motions, Gondwana dinosaurs, Noah's arks, beached Viking funeral ships, ghost ships, and landspans. *Anais da Academia Brasileira de Ciências*. 83:3-22.

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

Jarne, P., and P.J. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*. 11:424-429.

Jensen, J.L., A.J. Bohonak, and S.T. Kelley. 2005. Isolation by distance, web service. *BMC Genetics*. 6:13.

Jones, T.C., C.E. Gemmill, and C.A. Pilditch. 2008. Genetic variability of New Zealand seagrass (*Zostera muelleri*) assessed at multiple spatial scales. *Aquatic Botany*. 88:39-46.

Jost, L. 2008. GST and its relatives do not measure differentiation. *Molecular Ecology*. 17:4015-4026.

Jost, M.C., D.M. Hillis, Y. Lu, J.W. Kyle, H.A. Fozzard, and H.H. Zakon. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular Biology and Evolution*. 25:1016-1024.

Kai, Y., K. Nakayama, and T. Nakabo. 2002. Genetic differences among three colour morphotypes of the black rockfish, *Sebastes inermis,* inferred from mtDNA and AFLP analyses. *Molecular Ecology*. 11:2591-2598.

Kalinowski, S.T. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5:187-189.

REFERENCES

Kalinowski, S.T., and M.L. Taper. 2006. Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*. 7:991-995.

Keeney, D.B., A.D. Szymaniak, and R. Poulin. 2013. Complex genetic patterns and a phylogeographic disjunction among New Zealand mud snails *Zeacumantus subcarinatus* and *Z. lutulentus*. *Marine Biology*. 160:1477-1488.

Khor, S., S.A. Wood, L. Salvitti, D.I. Taylor, J. Adamson, P. McNabb, and S.C. Cary. 2014. Investigating diet as the source of tetrodotoxin in *Pleurobranchaea maculata*. *Marine Drugs*. 12:1-16.

Kimura, M. 1953. Stepping-stone model of population. *Annual Report of the National Institute of Genetics. Japan*. 3:62-63.

Kimura, M., and G.H. Weiss. 1964. Stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*. 49:561-&.

Kruskal, J.B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*. 29:115-129.

Kuchta, S.R., and A.M. Tan. 2005. Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. *Molecular Ecology*. 14:225-244.

Lee, J.Y., and S.V. Edwards. 2008. Divergence across Australia's carpentarian barrier: statistical phylogeography of the red-backed fairy wren (Malurus melanocephalus). *Evolution*. 62:3117-3134.

Lee, M.-J., D.-Y. Jeong, W.-S. Kim, H.-D. Kim, C.-H. Kim, W.-W. Park, Y.-H. Park, K.-S. Kim, H.-M. Kim, and D.-S. Kim. 2000. A Tetrodotoxin-Producing Vibrio Strain, LM-1, from the Puffer Fish *Fugu vermicularis radiatus*. *Applied and Environmental Microbiology*. 66:1698-1701.

Lehmann, J., and A. Libchaber. 2008. Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *RNA*. 14:1264-1269.

Lewis, K.B., L. Carter, and F.J. Davey. 1994. The opening of Cook Strait: interglacial tidal scour and aligning basins at a subduction to transform plate edge. *Marine Geology*. 116:293-312.

Liao, P.-C., D.-C. Kuo, C.-C. Lin, K.-C. Ho, T.-P. Lin, and S.-Y. Hwang. 2010. Historical spatial range expansion and a very recent bottleneck of *Cinnamomum kanehirae* Hay. (Lauraceae) in Taiwan inferred from nuclear genes. *BMC Evolutionary Biology*. 10:124-124.

REFERENCES

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25:1451-1452.

Lipscomb, D. 1998. Basics of Cladistic Analysis. *George Washington University, Washington DC*.

Lischer, H.E., and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 28:298-299.

Lowe, A., S. Harris, and P. Ashton. 2009. Ecological genetics: design, analysis, and application. John Wiley & Sons.

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

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27:209-220.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437:376-380.

Masel, J. 2011. Genetic drift. *Current Biology*. 21:R837-R838.

Matsui, T., Y. Ohtsuka, and K. Sakai. 2000. Recent advances in tetrodotoxin research. *Yakugaku Zasshi*. 120:825-837.

Matsumura, K. 1995. Tetrodotoxin as a pheromone. *Nature*. 378:563-564.

Matsumura, K. 2001. No ability to produce tetrodotoxin in bacteria. *Applied and Environmental Microbiology*. 67:2393-2393.

McArdle, B.H., and M.J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*. 82:290-297.

REFERENCES

McEdward, L.R. 1997. Reproductive strategies of marine benthic invertebrates revisited: facultative feeding by planktotrophic larvae. *Am Nat*. 150:48-72.

McNabb, P., A.I. Selwood, R. Munday, S.A. Wood, D.I. Taylor, L.A. Mackenzie, R. van Ginkel, L.L. Rhodes, C. Cornelisen, K. Heasman, P.T. Holland, and C. King. 2010. Detection of tetrodotoxin from the grey side-gilled sea slug - *Pleurobranchaea maculata*, and associated dog neurotoxicosis on beaches adjacent to the Hauraki Gulf, Auckland, New Zealand. *Toxicon*. 56:466-473.

McNabb, P.S., D.I. Taylor, S.C. Ogilvie, L. Wilkinson, A. Anderson, D. Hamon, S.A. Wood, and B.M. Peake. 2014. First detection of tetrodotoxin in the bivalve *Paphies australis* by liquid chromatography coupled to triple quadrupole mass spectrometry with and without precolumn reaction. *Journal of AOAC International*. 97:325-333.

Medina, M., S. Lal, Y. Valles, T.L. Takaoka, B.A. Dayrat, J.L. Boore, and T. Gosliner. 2011. Crawling through time: Transition of snails to slugs dating back to the Paleozoic, based on mitochondrial phylogenomics. *Marine Genomics*. 4:51-59.

Meglecz, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, and J.F. Martin. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*. 26:403-404.

Meirmans, P.G., and P.W. Hedrick. 2011. Assessing population structure: F-ST and related measures. *Molecular Ecology Resources*. 11:5-18.

Meirmans, P.G., and P.H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*. 4:792-794.

Merritt, T.J.S., L. Shi, M.C. Chase, M.A. Rex, R.J. Etter, and J.M. Quattro. 1998. Universal cytochrome b primers facilitate intraspecific studies in molluscan taxa. *Molecular Marine Biology and Biotechnology*. 7:7-11.

Mila, B., S. Carranza, O. Guillaume, and J. Clobert. 2010. Marked genetic structuring and extreme dispersal limitation in the Pyrenean brook newt *Calotriton asper* (Amphibia: Salamandridae) revealed by genome-wide AFLP but not mtDNA. *Molecular Ecology*. 19:108-120.

Miyazawa, K., T. Noguchi, J. Maruyama, J. Jeon, M. Otsuka, and K. Hashimoto. 1985. Occurrence of tetrodotoxin in the starfishes *Astropecten polyacanthus* and *A. scoparius* in the Seto Inland Sea. *Marine Biology*. 90:61-64.

REFERENCES

Morley, M.S., and B.W. Hayward. 2015. Intertidal records of 'sea slugs' (nudibranchs and allied opistobranch gastropods) from northern North Island, New Zealand. *Records of the Auckland Museum*. 50:51-93.

Muhlin, J.F., C.R. Engel, R. Stessel, R.A. Weatherbee, and S.H. Brawley. 2008. The influence of coastal topography, circulation patterns, and rafting in structuring populations of an intertidal alga. *Molecular Ecology*. 17:1198-1210.

Nagashima, Y., K. Yamamoto, K. Shimakura, and K. Shiomi. 2002. A tetrodotoxin-binding protein in the hemolymph of shore crab *Hemigrapsus sanguineus*: purification and properties. *Toxicon*. 40:753-760.

Nagel, M.M., M.A. Sewell, and S.D. Lavery. 2015. Differences in population connectivity of a benthic marine invertebrate Evechinus chloroticus (Echinodermata: Echinoidea) across large and small spatial scales. *Conservation Genetics*:1-14.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 89:583-590.

Nei, M. 1987. Molecular evolutionary genetics. Columbia university press.

Noguchi, T., and O. Arakawa. 2008. Tetrodotoxin-distribution and accumulation in aquatic organisms, and cases of human intoxication. *Mar Drugs*. 6:220-242.

Noguchi, T., O. Arakawa, and T. Takatani. 2006a. Toxicity of pufferfish *Takifugu rubripes* cultured in netcages at sea or aquaria on land. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 1:153-157.

Noguchi, T., O. Arakawa, and T. Takatani. 2006b. TTX accumulation in pufferfish. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 1:145-152.

Noguchi, T., and J.S.M. Ebesu. 2001. Puffer poisoning: Epidemiology and treatment. *Journal of Toxicology-Toxin Reviews*. 20:1-10.

Noguchi, T., T. Takatani, and O. Arakawa. 2004. Toxicity of puffer fish cultured in netcages. *Journal of the Food Hygienic Society of Japan*. 45:146-149.

Norris, R.D., and P.M. Hull. 2012. The temporal dimension of marine speciation. *Evolutionary Ecology*. 26:393-415.

Nosek, J., L. Tomaska, H. Fukuhara, Y. Suyama, and L. Kovac. 1998. Linear mitochondrial genomes: 30 years down the line. *Trends in Genetics*. 14:184-188.

REFERENCES

Oliveira, E.J., J.G. Pádua, M.I. Zucchi, R. Vencovsky, and M.L.C. Vieira. 2006. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*. 29:294-307.

Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*. 13:55-65.

Palumbi, S. 1991. Simple fool's guide to PCR.

Palumbi, S.R., and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*. 11:426-435.

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*. 28:2537-2539.

Pelc, R., R. Warner, and S. Gaines. 2009. Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*. 36:1881-1890.

Penant, G., D. Aurelle, J.-P. Feral, and A. Chenuil. 2013. Planktonic larvae do not ensure gene flow in the edible sea urchin Paracentrotus lividus. *Marine Ecology Progress Series*. 480:155.

Piry, S., A. Alapetite, J.M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*. 95:536-539.

Piry, S., G. Luikart, and J.-M. Cornuet. 1999. BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. Montpellier, France.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*. 25:1253-1256.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945-959.

Pritchard, J.K., X. Wen, and D. Falush. 2010. Documentation for structure software: Version 2.3. *University of Chicago, Chicago, IL.* .

Proksch, P., R. Edrada, and R. Ebel. 2002. Drugs from the seas – current status and microbiological implications. *Applied Microbiology and Biotechnology*. 59:125-134.

REFERENCES

Quoy, J.R.C., and J.P. Gaimard. 1832. Mollusques.Zoologie du voyage de l'Astrolabe sous les ordres du Capitane Dumont d'Urville, pendant les années 1826-1829. *Paris: J. Tastu, Editeur-Imprimeur*. Vol 2.

Rannala, B., and J.L. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*. 94:9197-9201.

Raupach, M.J., S. Thatje, J. Dambach, P. Rehm, B. Misof, and F. Leese. 2010. Genetic homogeneity and circum-Antarctic distribution of two benthic shrimp species of the Southern Ocean, *Chorismus antarcticus* and *Nematocarcinus lanceopes*. *Marine Biology.* 157:1783-1797.

Raymond, M., and F. Rousset. 1995. An exact test for population differentiation. *Evolution*. 49:1280-1283.

Reich, D.E., M.W. Feldman, and D.B. Goldstein. 1999. Statistical properties of two tests that use multilocus data sets to detect population expansions. *Molecular Biology and Evolution*. 16:453-466.

Reich, D.E., and D.B. Goldstein. 1998. Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences USA*. 95:8119-8123.

Ridenhour, B.J., E.D. Brodie Jr, and E.D. Brodie III. 2007. Patterns of genetic differentiation in *Thamnophis* and *Taricha* from the Pacific Northwest. *Journal of Biogeography*. 34:724-735.

Ridgway, N. 1980. Hydrological conditions and circulation off the west coast of the North Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research*. 14:155-167.

Riginos, C., K.E. Douglas, Y. Jin, D.F. Shanahan, and E.A. Treml. 2011. Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*. 34:566-575.

Roberts, P.E., and L. Paul. 1978. Seasonal hydrological changes in continental shelf waters off the west coast, North Island, New Zealand, and comments on fish distributions. *New Zealand Journal of Marine and Freshwater Research.* 12: 323–339.

Rodriguez, L.F. 2006. Can invasive species facilitate native species? Evidence of how, when, and why these impacts occur. *Biological Invasions*. 8:927-939.

REFERENCES

Rodriguez, P., A. Alfonso, C. Vale, C. Alfonso, P. Vale, A. Tellez, and L.M. Botana. 2008. First toxicity report of tetrodotoxin and 5,6,11-trideoxyTTX in the trumpet shell *Charonia lampas lampas* in Europe. *Analytical Chemistry*. 80:5622-5629.

Rogers, A.R. 1995. Genetic-evidence for a pleistocene population explosion. *Evolution*. 49:608-615.

Rogers, A.R., and H. Harpending. 1992. Population-growth makes waves in the distribution of pairwise genetic-differences. *Molecular Biology and Evolution*. 9:552-569.

Rohwer, S., E. Bermingham, and C. Wood. 2001. Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution*. 55:405-422.

Rosenberg, N.A. 2004. distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*. 4:137-138.

Rosetti, N., J.C. Vilardi, and M.I. Remis. 2007. Effects of B chromosomes and supernumerary segments on morphometric traits and adult fitness components in the grasshopper, *Dichroplus elongatus* (Acrididae). *Journal of Evolutionary Biology*. 20:249-259.

Ross, P.M., I.D. Hogg, C.A. Pilditch, and C.J. Lundquist. 2009. Phylogeography of New Zealand's coastal benthos. *New Zealand Journal of Marine and Freshwater Research*. 43:1009-1027.

Ross, P.M., I.D. Hogg, C.A. Pilditch, C.J. Lundquist, and R.J. Wilkins. 2012. Population Genetic Structure of the New Zealand Estuarine Clam *Austrovenus stutchburyi* (Bivalvia: Veneridae) Reveals Population Subdivision and Partial Congruence with Biogeographic Boundaries. *Estuaries and Coasts*. 35:143-154.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 145:1219-1228.

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*. 8:103-106.

Rubinsztein, D.C., W. Amos, J. Leggo, S. Goodburn, S. Jain, S.-H. Li, R.L. Margolis, C.A. Ross, and M.A. Ferguson-Smith. 1995. Microsatellite evolution--evidence for directionality and variation in rate between species. *Nature Genetics*. 10:337-343.

Rudman, W. 1999. Pleurobranchaea maculata (Quoy & Gaimard, 1832). Sea Slug Forum. Australian Museum, Sydney. http://www.seaslugforum.net/showall/pleumacu.

REFERENCES

Ryman, N., and S. Palm. 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*. 6:600-602.

Sabrah, M., A. El-Ganainy, and M. Zaky. 2006. Biology and toxicity of the pufferfish Lagocephalus sceleratus (Gmelin, 1789) from the Gulf of Suez. *Egyptian Journal of Aquatic Research*. 32:283-297.

Salvitti, L., S.A. Wood, D.I. Taylor, P. McNabb, and S.C. Cary. 2015a. First identification of tetrodotoxin (TTX) in the flatworm *Stylochoplana* sp.; a source of TTX for the sea slug *Pleurobranchaea maculata*. *Toxicon*. 95:23-29.

Salvitti, L.R., S.A. Wood, P. McNabb, and S.C. Cary. 2015b. No Evidence for a Culturable Bacterial Tetrodotoxin Producer in *Pleurobranchaea maculata* (Gastropoda: Pleurobranchidae) and *Stylochoplana* sp. (Platyhelminthes: Polycladida). *Toxins*. 7:255-273.

Salvitti, L.R., S.A. Wood, L. Winsor, and S.C. Cary. 2015c. Intracellular Immunohistochemical Detection of Tetrodotoxin in Pleurobranchaea maculata (Gastropoda) and Stylochoplana sp. (Turbellaria). *Marine Drugs*. 13:756-769.

Scheffer, S.J., and D.J. Hawthorne. 2007. Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera : Agromyzidae): apparent discordance among marker systems. *Molecular Ecology*. 16:2627-2637.

Schenekar, T., and S. Weiss. 2011. High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. *Heredity-Basingstoke*. 107:511.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, and C.F. Weber. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. 75:7537-7541.

Schnabel, K.E., I.D. Hogg, and M.A. Chapman. 2000. Population genetic structures of two New Zealand corophiid amphipods and the presence of morphologically cryptic species: implications for the conservation of diversity. *New Zealand Journal of Marine and Freshwater Research*. 34:637-644.

REFERENCES

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*. 152:1079-1089.

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

Selkoe, K.A., and R.J. Toonen. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*. 436:291-305.

Shanks, A., J. Largier, and J. Brubaker. 2003a. The nearshore distribution of larval invertebrates during an upwelling event. *Journal of Plankton Research.* 25:645-667.

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

Shanks, A.L., and L. Brink. 2005. Upwelling, downwelling, and cross-shelf transport of bivalve larvae: test of a hypothesis. *Marine Ecology Progress Series*. 302:1-12.

Shanks, A.L., B.A. Grantham, and M.H. Carr. 2003b. Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*. 13:159-169.

Shears, N.T., F. Smith, R.C. Babcock, C.A. Duffy, and E. Villouta. 2008. Evaluation of biogeographic classification schemes for conservation planning: Application to New Zealand's coastal marine environment. *Conservation Biology*. 22:467-481.

Silva, E., and C. Russo. 2000. Techniques and statistical data analysis in molecular population genetics. *Hydrobiologia*. 420:119-135.

Silva, M., J. Azevedo, P. Rodriguez, A. Alfonso, L.M. Botana, and V. Vasconcelos. 2012. New gastropod vectors and tetrodotoxin potential expansion in temperate waters of the Atlantic Ocean. *Marine Drugs*. 10:712-726.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science*. 236:787-792.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*. 264-279.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 139:457-462.

Slatkin, M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*. 9:477-485.

REFERENCES

Slatkin, M., and R.R. Hudson. 1991. Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics*. 129:555-562.

Smith, P., G. MacArthur, and K. Michael. 1989. Regional variation in electromorph frequencies in the tuatua, *Paphies subtriangulata,* around New Zealand. *New Zealand Journal of Marine and Freshwater Research*. 23:27-33.

Sponer, R., M.S. Roy, and F. Bonhomme. 2002. Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*. 56:1954-1967.

Stanton, B. 1973. Hydrological investigations around northern New Zealand. *New Zealand Journal of Marine And Freshwater Research*. 7:85-110.

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

Stevens, M.I., and I.D. Hogg. 2004. Population genetic structure of New Zealand's endemic corophiid amphipods: evidence for allopatric speciation. *Biological Journal of the Linnean Society*. 81:119-133.

Stokes, A.N., P.K. Ducey, L. Neuman-Lee, C.T. Hanifin, S.S. French, M.E. Pfrender, E.D. Brodie, 3rd, and E.D. Brodie, Jr. 2014. Confirmation and distribution of tetrodotoxin for the first time in terrestrial invertebrates: two terrestrial flatworm species (*Bipalium adventitium* and *Bipalium kewense*). *PLoS One*. 9:e100718.

Sugita, H., J. Iwata, C. Miyajima, T. Kubo, T. Noguchi, K. Hashimoto, and Y. Deguchi. 1989. Changes in microflora of a puffer fish *Fugu niphobles*, with different water temperatures. *Marine Biology.* 101:299-304.

Sunnucks, P. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution*. 15:199-203.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 105:437-460.

Tajima, F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:585-595.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution*. 10:512-526.

REFERENCES

Taylor, D.I., S.A. Wood, P. McNabb, S. Ogilvie, J. Walker, S. Khor, S.C. Cary, and C. Cornelisen. 2015. Facilitation effects of invasive and farmed bivalves on native populations of *Pleurobranchaea maculata*: population explosions, diet analysis and dog deaths. *Inter Research Marine Ecology Progress Series.* 537:39-48

Team, R.C. 2013. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Toews, D.P., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*. 21:3907-3930.

Trewick, S.A., and K.J. Bland. 2012. Fire and slice: palaeogeography for biogeography at New Zealand's North Island/South Island juncture. *Journal of the Royal Society of New Zealand*. 42:153-183.

Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. 4:535-538.

Veale, A. 2007. Phylogeography of two intertidal benthic marine invertebrates around New Zealand : the waratah anemone (*Actinia tenebrosa*) & the snakeskin chiton (*Sypharochiton pelliserpentis*). *Unpublished MSc thesis, University of Auckland, Auckland, New Zealand.*

Veale, A.J., and S.D. Lavery. 2011. Phylogeography of the snakeskin chiton Sypharochiton pelliserpentis (Mollusca: Polyplacophora) around New Zealand: are seasonal near-shore upwelling events a dynamic barrier to gene flow? *Biological Journal of the Linnean Society*. 104:552-563.

Veeramah, K.R., D. Wegmann, A. Woerner, F.L. Mendez, J.C. Watkins, G. Destro-Bisol, H. Soodyall, L. Louie, and M.F. Hammer. 2012. An early divergence of Khoesan ancestors from those of other modern humans ıs supported by an abc-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*. 29:617-630.

Vergeer, P., and W.E. Kunin. 2013. Adaptation at range margins: common garden trials and the performance of *Arabidopsis lyrata* across its northwestern European range. *New Phytologist*. 197:989-1001.

Wallis, G.P., and S.A. Trewick. 2009. New Zealand phylogeography: evolution on a small continent. *Molecular Ecology*. 18:3548-3580.

REFERENCES

Wang, X.J., R.C. Yu, X. Luo, M.J. Zhou, and X.T. Lin. 2008. Toxin-screening and identification of bacteria isolated from highly toxic marine gastropod Nassarius semiplicatus. *Toxicon*. 52:55-61.

Waters, J., T. King, P. O'loughlin, and H. Spencer. 2005. Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology*. 14:2789-2802.

Watterson, G.A. 1975. Number of segregating sites in genetic models without recombination. *Theoretical Population Biology*. 7:256-276.

Weaver, P.P., L. Carter, and H.L. Neil. 1998. Response of surface water masses and circulation to late Quaternary climate change east of New Zealand. *Paleoceanography*. 13:70-83.

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

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

Wei, K., A.R. Wood, and J.P.A. Gardner. 2013b. Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. *Marine Ecology Progress Series*. 477:107-121.

Weinberg, W. 1908. Über vererbungsgesetze beim menschen. *Molecular and General Genetics.* 1:440-460.

Weir, B.S. 1996. Genetic Data Analysis II. *Sinauer Associates, Sunderland, MA.*

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population-structure. *Evolution*. 38:1358-1370.

West, C.J., and A.M. Thompson. 2013. Small, dynamic and recently settled: responding to the impacts of plant invasions in the New Zealand (Aotearoa) archipelago. *In* Plant Invasions in Protected Areas. *Springer*. 285-311.

Will, M., M.L. Hale, D.R. Schiel, and N.J. Gemmell. 2011. Low to moderate levels of genetic differentiation detected across the distribution of the New Zealand abalone, *Haliotis iris*. *Marine Biology.* 158:1417-1429.

Will, M.C., and N.J. Gemmel. 2008. Genetic population structure of blackfoot paua. *University of Otago report (GEN2007–01) to Ministry of Fisheries: (Unpublished report held by Ministry of Fisheries, Wellington).*

REFERENCES

Willan, R.C. 1983. New-Zealand side-gilled sea slugs (Opisthobranchia,Nnotaspidea, Pleurobranchidae). *Malacologia*. 23:221-270.

Williams, B.L. 2010. Behavioral and chemical ecology of marine organisms with respect to tetrodotoxin. *Marine Drugs*. 8:381-398.

Winnepenninckx, B., T. Backeljau, and R. De Wachter. 1993. Extraction of high molecular weight DNA from molluscs. *Trends in Genetics*. 9:407.

Wood, S.A., M. Casas, D.I. Taylor, P. McNabb, L. Salvitti, S. Ogilvie, and S.C. Cary. 2012a. Depuration of Tetrodotoxin and Changes in Bacterial Communities in *Pleurobranchea maculata* Adults and Egg Masses Maintained in Captivity. *Journal of Chemical Ecology*. 38:1342-1350.

Wood, S.A., D.I. Taylor, P. McNabb, J. Walker, J. Adamson, and S.C. Cary. 2012b. Tetrodotoxin concentrations in *Pleurobranchaea maculata*: temporal, spatial and individual variability from New Zealand populations. *Marine Drugs*. 10:163-176.

Woodward, R.B. 1964. The structure of tetrodotoxin. *Pure and Applied Chemistry.* 9:49.

Wright, S. 1951. The genetical structure of populations. *Ann Eugen*. 15:323-354.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*. 395-420.

Wright, S. 1978. Vol. 4: Variability within and among natural populations. Chicago [etc.]: University of Chicago Press.

Xia, X. 2000. Data analysis in molecular biology and evolution. *Springer Science & Business Media.*

Xia, X., and P. Lemey. 2009. Assessing substitution saturation with DAMBE. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*. 2:615-630.

Xia, X., and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*. 92:371-373.

Xia, X., Z. Xie, M. Salemi, L. Chen, and Y. Wang. 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*. 26:1-7.

Yang, D.-S., and G.J. Kenagy. 2009. Nuclear and mitochondrial DNA reveal contrasting evolutionary processes in populations of deer mice (*Peromyscus maniculatus*). *Molecular Ecology*. 18:5115-5125.

REFERENCES

Yang, G., J. Xu, S. Liang, D. Ren, X. Yan, and B. Bao. 2010. A novel TTX-producing Aeromonas isolated from the ovary of *Takifugu obscurus*. *Toxicon*. 56:324-329.

Yildirim, Y., S. Patel, C.D. Millar, and P.B. Rainey. 2014. Microsatellite development for a tetrodotoxin-containing sea slug (*Pleurobranchaea maculata*). *Biochemical Systematics and Ecology*. 55:342-345.

Yotsu-Yamashita, M., J. Gilhen, R.W. Russell, K.L. Krysko, C. Melaun, A. Kurz, S. Kauferstein, D. Kordis, and D. Mebs. 2012. Variability of tetrodotoxin and of its analogues in the red-spotted newt, *Notophthalmus viridescens* (Amphibia: Urodela: Salamandridae). *Toxicon*. 59:257-264.

Yotsu-Yamashita, M., A. Sugimoto, T. Terakawa, Y. Shoji, T. Miyazawa, and T. Yasumoto. 2001. Purification, characterization, and cDNA cloning of a novel soluble saxitoxin and tetrodotoxin binding protein from plasma of the puffer fish, *Fugu pardalis*. *European Journal of Biochemistry*. 268:5937-5946.

Yu, V.C.-H., P.H.-F. Yu, K.-C. Ho, and F.W.-F. Lee. 2011. Isolation and identification of a new tetrodotoxin-producing bacterial species, *Raoultella terrigena*, from Hong Kong marine puffer fish *Takifugu niphobles*. *Marine Drugs*. 9:2384-2396.

Zhan, A., J. Hu, X. Hu, Z. Zhou, M. Hui, S. Wang, W. Peng, M. Wang, and Z. Bao. 2009. Fine-Scale Population Genetic Structure of Zhikong Scallop (*Chlamys farreri*): Do Local Marine Currents Drive Geographical Differentiation? *Marine Biotechnology*. 11:223-235.

Zimmer, R.K., and R.P. Ferrer. 2007. Neuroecology, chemical defense, and the keystone species concept. *The Biological Bulletin*. 213:208-225.

Zimmermann, B., A.W. Rock, A. Dur, and W. Parson. 2014. Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software. *Croatian Medical Journal*. 55:115-120.

Zink, R.M. 2010. Drawbacks with the use of microsatellites in phylogeography: the song sparrow *Melospiza melodia* as a case study. *Journal of Avian Biology*. 41:1-7.

Zink, R.M., and G.F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*. 17:2107-2121.

APPENDIX ONE

Biochemical Systematics and Ecology 55 (2014) 342–345

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Microsatellite development for a tetrodotoxin-containing sea slug *(Pleurobranchaea maculata)*

Yeşerin Yıldırım [a,*], Selina Patel [b,c], Craig D. Millar [b,c], Paul B. Rainey [a,d]

[a] New Zealand Institute for Advanced Study and Allan Wilson Centre for Molecular Ecology and Evolution, Massey University Albany, North Shore City 0745, New Zealand
[b] The Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand
[c] School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand
[d] Max Planck Institute for Evolutionary Biology, Plön, Germany

ARTICLE INFO

Article history:
Received 15 March 2014
Accepted 12 April 2014
Available online

Keywords:
Grey side-gilled sea slug
Pleurobranchaea maculata
Tetrodotoxin
Microsatellite markers
454 pyrosequencing

ABSTRACT

Using 454 pyrosequencing data, 24 polymorphic microsatellite markers were identified for the grey side-gilled sea slug, *Pleurobranchaea maculata*. The grey side-gilled sea slug is found throughout the western and south Pacific and is known to contain high concentrations of tetrodotoxin. Polymorphism was assessed in 20 individuals obtained from geographically distinct locations within New Zealand. Between 2 and 15 alleles were identified at each locus. The observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0.10 to 1 and 0.10–0.94, respectively. No significant linkage disequilibrium between pairs of loci or deviations from the Hardy–Weinberg proportions were observed. The markers are central to understanding the population biology and genetic structure of *P. maculata*.

1. Introduction

The grey side-gilled sea slug, *Pleurobranchaea maculata*, is native to coastal waters of New Zealand, south-eastern Australia, China, Sri Lanka and Japan (Wilan, 1983). Little is known of its biology, however it is thought to be an opportunistic carnivore. In late 2009, the grey side-gilled sea slug attracted attention in New Zealand after it was implicated in dog deaths on Auckland beaches. Forensic analyses revealed deaths to be a consequence of tetrodotoxin (TTX) poisoning associated with ingestion of *P. maculata* (McNabb et al., 2010). Additional studies have shown differences in concentrations of TTX associated *P. maculata* populations from different regions of New Zealand (Wood et al., 2012), however, the origin of TTX in *P. maculata* is unknown.

Understanding the genetic structure of *P. maculata* populations from different regions of New Zealand and the Pacific provides one route to understanding the origins and evolution of TTX-producing sea slugs. To this end we have developed microsatellite markers and here we report their identity and associated properties.

* Corresponding author. Massey University, New Zealand Institute for Advanced Study, Albany Highway, Gate 4, Building 16, Albany, Auckland 0632, New Zealand. Tel.: +64 9 4140800x41496, +64 2 102670238 (mobile).
E-mail addresses: Y.Yildirim@massey.ac.nz, yeserin@gmail.com (Y. Yıldırım).

http://dx.doi.org/10.1016/j.bse.2014.04.001
0305-1978/

2. Materials and methods

2.1. Sampling, DNA extraction and genome sequencing

The genome of *P. maculata* was partially sequenced using a single sea slug collected from Narrow Neck, Auckland, New Zealand in June 2010. The slug was placed at −80 °C upon collection. DNA extraction included the following steps: grinding of tissue with liquid nitrogen grinding, lysis with a cetyltrimethylammonium bromide buffer followed by phenol/chloroform extraction (Winnepenninckx et al., 1993) The resulting aqueous phase was put through a Qiagen DNeasy Blood & Tissue Kit. Preparation of the DNA library, emulsion-based clonal amplification of the library and sequencing were performed on Roche 454 GS Junior instrumentation using Titanium technology (Roche, NJ, USA) according to the manufacturer's protocols.

The utility of microsatellite markers was determined by assessing levels of polymorphism in a sample of twenty slugs collected from Ti-Point Wharf (Leigh, Auckland) and Tasman Bay (Nelson) in August and November 2012, respectively. Slug DNA was extracted using OMEGA E.Z.N.A.™ Mollusc DNA Kit after grinding the freeze-dried samples with a mortar and pestle.

2.2. Isolation and development of the primers

The bioinformatics software QDD (Meglecz et al., 2010) was used to analyze the 454 sequences, to search for microsatellite motifs, and to design primers. Sequences with significant similarity were first discarded (default settings) in order to avoid redundant sequences, and consensus sequences were generated. Sequences ≥80 bp were screened for perfect and compound microsatellite with minimum of four repeats of di-hexanucleotide motifs. Primer 3 software, built into QDD, was used to design primers using default settings. 90-350 bp of amplicon size was selected.

The 5′ ends of the forward primers of candidate loci were tagged with a 5′ -fluorescent dye (6-FAM, VIC, NED and PET) or with a universal M13(-21) sequence (18bp) (Schuelke, 2000). For the latter loci, a third primer with the M13(-21)-sequence and a 5′ -fluorescent dye (6-FAM, HEX, NED and PET) was used to screen primer pairs. PCRs for each locus were carried out in a total volume of 12.5 μL including 1× PCR buffer (Invitrogen Kit), 2 mM $MgCl_2$, 0.2 μM M13 primer and reverse primers, 0.05 μM forward primer, 20 mM dNTP, 0.25 U/μL Taq polymerase and 10–25 ng genomic DNA. Either standard or touch-down PCR regimes were applied (Table 1). Standard conditions were as following: 94 °C (3 min), then 30 cycles at 94 °C (30 s)/T_m (30 s)/72 °C (30–60 s). Touch-down conditions were as follows: 94 °C (3 min), then 20 cycles at 94 °C (30 s)/63 °C (decreasing by 0.5 °C/cycle, 30 s)/72 °C (20–60 s) followed by 20 cycles at 94 °C (30 s)/53 °C (30 s)/72 °C (20–60 s). Both regimes were followed by 8 cycles at 94 °C (30 s)/53 °C (30 s)/72 °C (20–60 s) if M13(-21) sequence is used for fluorescent labeling. A final extension at 72 °C for 30 min was applied for all reactions. PCR regime, annealing temperature and elongation time were estimated according to primer pair properties, and they are shown in Table 1. The multiplexed amplicons were electrophoresed on a commercial ABI 3730 Genetic Analyzer along with a size standard GeneScan™-500 LIZ™ (Applied Biosystems). The resulting genotyping data was analyzed with Microsatellite Plug-in on Geneious Pro 5.5.6.

2.3. Data analysis

Genotyping data was tested for genotyping errors, null alleles, large allele dropout and departure from Hardy–Weinberg equilibrium (HWE) using MICRO-CHECKER v.2.2.3 (van Oosterhout et al., 2004). GenoDive v2.0b24 was used to calculate allelic richness (Meirmans and van Tienderen, 2004). Observed heterozygosity (H_o), expected heterozygosity (H_e) and linkage disequilibrium (LD) for each pair of loci were estimated using ARLEQUIN v3.5 (Excoffier et al., 2005). The number of permutations and initial conditions for the EM algorithm were set to 20,000 and 5, respectively. R package (R Development Core Team, 2013) was used to adjust LD results for multiple simultaneous comparisons based on Bonferroni correction.

3. Results and discussion

The 454 sequencing run yielded 115,579 reads with an average length of 241 bp giving a total of 27.456 Mbp of sequence data. (NCBI SRA accession number SRR1168412). From these sequences 30,130 (23%) candidate microsatellite loci with ≥4 di- to hexa repeats were obtained. Size of the loci ranged from 90 to 866 bp in length. Among these sequences, Primer 3 created putative primer sets for 639 and 201 loci, having perfect and interrupted repeats, respectively. Generally, several primer pairs

Table 1
PCR annealing temperature and elongation time for each primer pair.

Locus ID	Annealing temperature (°C)	Elongation time (sec)
[b]Pm01, [b]Pm07, [b]Pm09, [b]Pm10, Pm12, Pm18, Pm19, [b]Pm20, Pm21	57	45
[b]Pm02, Pm04, Pm06, [b]Pm13, Pm14, [b]Pm17, Pm15, Pm16, [b]Pm23, [b]Pm24	63–53[a]	45
[b]Pm08, [b]Pm22,	63–53[a]	30
[b]Pm11	49	60
Pm03, Pm05	55	45

[a] Touchdown regime.
[b] Forward primers directly tagged with a fluorescent dye.

were suggested for each locus. Among them, primer pairs with the "best" score, which correspond to pairs with the lowest penalty, were considered as candidate pairs. Among the designed primers, 65 loci (8 di-, 34 tri-, 16 tetra-, 1 penta- and 6 hexanucleotide repeats) were selected for polymorphism analysis. Thirty of the 65 loci yielded successful amplicons of the expected size. These loci were further examined for consistency of amplification and polymorphism with a panel of 20 slugs from two different geographic regions (Leigh and Nelson, New Zealand). Six loci did not amplify consistently. Genotyping of the remaining 24 loci showed that two loci had limited polymorphism (two alleles), but 22 loci yielded polymorphism in the range of 3–15 alleles (Table 2). The average number of alleles per locus was 5.13. The H_o ranged from 0.10 to 1, and H_e ranged from 0.10 to 0.94 (Table 2).

MICROCHECKER analysis showed no evidence of large allele drop-out or scoring error due to stuttering, however, manual analysis of band patterns revealed evidence of stuttering at loci Pm23 and Pm25. Deviations from HWE were observed at loci Pm06, Pm23 and Pm24 for the Leigh population and at Pm06 for the Nelson population. These are most likely a consequence of chance effects due to small sample size. No significant linkage disequilibrium was found after Bonferroni correction ($P < 0.05$).

Table 2
Characteristics of 24 polymorphic microsatellite loci in *Pleurobranchaea maculata* isolated using 454 high-throughput sequencing.

Locus name & GenBank no	Primer sequence (5′ to 3′)	Dye	Repeat motif	[a]Na	Allele size (bp)	Leigh (n:10)		Nelson (n:10)	
						Ho	He	Ho	He
Pm01	F: TGGTGATCACGACACAAATG	NED	$(ATTC)_{15}$	15	108–208	0.70	0.77	0.70	0.93
KJ510176	R: CGCAAAATAAAATAGCAAGCC								
Pm02	F: TGTAGGAACAAATAGGCAGTGAGA	PET	$(CAGA)_{12}$	7	105–133	0.70	0.79	0.60	0.74
KJ510177	R: TTCGAAGAGGAATGGAATGG								
Pm03	F: ACCACTTTGCAGGTCAAACC	PET	$(GCGTGT)_7$	5	140–170	0.30	0.52	0.50	0.67
KJ510178	R: GTTCACACTGCATCATTGCC								
Pm04	F: AACGATCTAATCGGTGGCTG	PET	$(CTGC)_7$	3	246–254	0.40	0.57	0.40	0.54
KJ510179	R: GAAATTTGTACCTCGCGCTC								
Pm05	F: AAAGCAATGATAACAGCGCC	FAM	$CAGG)_6$	3	119–127	0.20	0.19	0.10	0.10
KJ510180	R: CGGGTGAAAGTTGACAGAC								
Pm06	F: TTATCTCTGTTTCCCCACCG	NED	$(GCAG)_6$	5	144–160	0.80	0.66	0.70	0.56
KJ510182	R: TGAGATAGTGTGCCATGTCCTT								
Pm07	F: ACCGACATGCAAACAGACAG	NED	$(CAGG)_7$	5	132–148	0.50	0.73	0.80	0.72
KJ510183	R: TTGTCTCCAAGTGGCCAAG								
Pm08	F: GGGACAGACAAATGGCTTGA	VIC	$(CTTT)_{11}$	5	140–164	0.60	0.70	0.80	0.67
KJ510184	R: GAAAAGAAAGAATGAAAGAAAGAATGA								
Pm09	F: GGCATAGGATGAAAAGGCAA	FAM	$(ATG)_{11}$	9	103–136	0.90	0.78	0.80	0.77
KJ510185	R: CAGGCTCCAGATCGGTAAAG								
Pm10	F: GGAATTGGGCGCTATACAAT	PET	$(TTA)_{12}$	5	110–122	0.70	0.70	0.60	0.66
KJ510186	R: AGCCGCTTGGTACTCAAAGA								
Pm11	F: CAAAGACCCGATATTCACGG	VIC	$(GTT)_9$	8	160–181	1.00	0.89	0.60	0.82
KJ510187	R: GAAACAATAGCAACACAACGG								
Pm12	F: TCAAGGTTAGTTTCACCATCACC	PET	$(CAT)_9$	3	122–128	0.70	0.62	0.60	0.66
KJ510188	R: TAGTGATAGTGGCAGTGGCG								
Pm13	F: TCAACTGGGAACAAACATCG	FAM	$(GAT)_{10}$	4	121–133	0.50	0.54	0.50	0.49
KJ510189	R: ATGTCATTGGTTGACGTTGG								
Pm14	F: TGCTACCACTGCTACTATGCTATTG	HEX	$(CTGCTA)_{10}$	4	164–194	0.50	0.54	0.60	0.64
KJ510190	R: CATCCACGATCCACAGTCAA								
Pm15	F: AGGTCGAACCCTTAACCTCG	PET	$(CAC)_8$	4	241–250	0.40	0.53	0.90	0.63
KJ510191	R: AATGGTTTTAGCCTTGGCCT								
Pm16	F: GCAGCCAACAAGGAAGGATA	FAM	$(AGGTGA)_5$	3	160–172	NA	NA	0.20	0.19
KJ510192	R: GGTAACCATGGATTTGACGC								
Pm17	F: ACTGAATGGGGTCGAGTACG	VIC	$(GTTT)_5$	4	167–179	0.50	0.57	0.30	0.27
KJ510193	R: TTTGTGCTAAAACCTTCCCG								
Pm18	F: GCTCACTCATCGCATTCATC	NED	$(CTA)_8$	2	175–178	NA	NA	0.10	0.10
KJ510194	R: ATGTGCCTTTGTGCTGTCAC								
Pm19	F: TGTTCTTGATTCGTTATGGGG	NED	$(GTT)_8$	5	169–181	0.20	0.19	0.80	0.59
KJ510195	R: TGTCCGATGGTGAGCAGTAA								
Pm20	F: AGTCGATATTATGGCGTCGG	FAM	$(CTT)_9(CTC)_6$	2	120–126	0.60	0.87	0.10	0.10
KJ510196	R: TTGTTTAAACTTGTTTGAGAAAAGG								
Pm21	F: TTTAACTGGCTGTAGGCCAAA	FAM	$(CTT)_9(CTC)_6$	4	198–207	0.60	0.63	0.60	0.50
KJ510197	R: CGAAGAGAAAGCCGGTAGTG								
Pm22	F: CTGTCTTGGCCATAGGCTCT	VIC	$(ATC)_{18}$	8	101–122	0.40	0.76	0.60	0.67
KJ510198	R: GGCATAGGATGAAAAGGCAA								
Pm23	F: CAATAACAAAGGCCTGGGAA	NED	$(AG)_{14}$	7	156–170	0.70	0.77	0.60	0.63
KJ510199	R: GGCATGTCGATTGCACTAAA								
Pm24	F: TGGAAGGTTGTGTATCAAGCA	PET	$(AC)_{14}$	6	179–191	0.70	0.79	0.60	0.74
KJ510200	R: ATCCATCAAATGGTGGCAAT								

Na: total number of alleles, Ta: annealing temperature (63–53 means touchdown regime), n: total number of samples, NA: not available due to only one allele.
[a] data from the individual used for 454 sequencing was included to the calculations.

The results demonstrate identification of a robust set of microsatellite markers suitable for bio-geographic and population structure analysis of *P. maculata*.

Acknowledgments

This work was supported by grants from the Allan Wilson Center for Ecology and Evolution, Massey University, the University of Auckland and Auckland Regional Council. We thank the Centre for Genomics, Proteomics and Metabolomics at the University of Auckland and Allan Wilson Center Genome Service. We also thank David Taylor and Susanna Wood from Cawthron Institute, Richard Huges and Paul Caiger from the Leigh Marine Laboratory at the University of Auckland, and Severine Hannam and Wilma Blom from Auckland War Memorial Museum for provision of samples. We are grateful to David Taylor and Susanna Wood for sharing their experience and knowledge about *P. maculata*.

References

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

McNabb, P., Selwood, A.I., Munday, R., Wood, S.A., Taylor, D.I., Mackenzie, L.A., van Ginkel, R., Rhodes, L.L., Cornelisen, C., Heasman, K., Holland, P.T., King, C. 2010. Detection of tetrodotoxin from the grey side-gilled sea slug – Pleurobranchaea maculata, and associated dog neurotoxicosis on beaches adjacent to the Hauraki Gulf, Auckland, New Zealand. Toxicon 56, 466–473.

Meglecz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N., Martin, J.F., 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26, 403–404.

Meirmans, P.G., van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol Notes 4, 792–794.

R Development Core Team, 2013. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18, 233–234.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538.

Wilan, R.C., 1983. New-Zealand side-gilled sea slugs (Opisthobranchia, Notaspidea, Pleurobranchidae). Malacologia 23, 221–270.

Winnepenninckx, B., Backeljau, T., De Wachter, R., 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet. 9, 407.

Wood, S.A., Taylor, D.I., McNabb, P., Walker, J., Adamson, J., Cary, S.C., 2012. Tetrodotoxin concentrations in *Pleurobranchaea maculata*: temporal, spatial and individual variability from New Zealand populations. Mar. Drugs 10, 163–176.

APPENDIX TWO: SUPPLEMENTARY FIGURES AND TABLES

Supplementary Table 1 Primer pairs for candidate microsatellite loci that were selected for amplification and polymorphism analysis.

Primer code	Sequence code	PCR product size (bp)	Primer sequence (5' to 3')	Repeat motif
1	BSDD8	124	F: TGGTGATCACGACACAAATG; R: CGCAAAATAAAATAGCAAGCC	(ATTC)15
2	BGGSJ	133	F: TGTAGGAACAAATAGGCAGTGAGA; R: TTCGAAGAGGAATGGAATGG	(CAGA)12
3	BSGB0	164	F: ACCGGCATAACGGGTAGTAA; R: GGACAAGAATCGATACCGGA	(AGTA)11
4*	AFQZJ	148	F: ACCACTTTGCAGGTCAAACC; R: GTTCACACTGCATCATTGCC	(GTGTGC)7
5	A3ZG1	126	F: CCATTCCATTACCGTTACCA; R: GGCACAGTTCTCCGGTAGTC	(CATTAC)7
7	BYGLD	228	F: AACGATCTAATCGGTGGCTG; R: GAAATTTGTACCTCGCGCTC	(CTGC)7
8	AOFL9	141	F: ACCTTGTAACACTGGAGCCG; R: AAAGTAGCAGCCTAACGCCA	(ACT)23
9	AEBZY	120	F: CACGATATCGGGGCTGAATA; R: AGGTCATTTGCCTGTTTTGAG	(ACA)19
10	ADQBS	182	F: ATCGAGATTTGGGTCGTGAG; R: AAAGAAAACTACTTCTACCGACTGA	(TAGG)12
11*	A0DQG	110	F: AAAGCAATGATAACAGCGCC; R: CGGGTGAAAGTTGACAGACA	(GCAG)6
12	A525S	149	F: ATGCATCTGGTCAACATCAA; R: TGCATCAATTGACGGTAATG	(TATC)6
13	A88ZO	98	F: CCCTATCTGACCCGTTAGCA; R: TGGTTTCAATGAAGAAGGTGC	(AGGGTT)6
14*	AIS5W	171	F: AGCCGACTTGTCAACATCGT; R: GGCATGTGCATAATGACCAG	(TATC)6
15	AKEWQ	147	F: TTATCTCTGTTTCCCCACCG; R: TGAGATAGTGTGCCATGTCCTT	(GCAG)6
16	AOMLR	194	F: CCAAGCCCTGTGTTGCTATT; R: AAGGACCCGGAACGAACTAT	(TGTC)6
17	BPN7J	110	F: CTTCGTGCGTAGTGTGGGAT; R: TCCAACACTCGTGCTACACC	(TTAT)6
18	A31P8	107	F: TAATGCACCTTGAGCACGTC; R: ATGTCGATCATGGTTAAGCG	(TAAAG)7
19	A7RE7	304	F: TGCTCTCCGAATCATTACCC; R: CAGATCACCACCAATGAACG	(TGTT)7
20	AWK5B	140	F: ACCGACATGCAAACAGACAG; R: TTGTCTCCAAGTGGCCAAG	(CAGG)7
21	AB5G4	201	F: ATGAATTTTACGTCAGCGGC; R: TGACAGCTTCTGGCTAACTGC	(GTA)10
22	ACFKD	248	F: GCTCCCGTTTTGTTCATTTC; R: ATCCGTTTCATTTTCGTTGC	(TTG)10
23	AL5S7	130	F: ATCAAGTCGCAAGAATGCCT; R: GGGGCAAAAGGTGTAGACAA	(GCT)10

24	AQMBQ	107	F: GGTCCAATTTTCCTTTTCCC; R: GACTTTTGCCAGTTCAGGCT	(CTA)10
25	ABPJ5	239	F: TGACTACCTGGGGATTTTGG; R: TCTGTCTCTGTCCGCGTTTA	(TAA)11
26	AGL33	152	F: GGGACAGACAAATGGCTTGA; R: GAAAAGAAAGAATGAAAGAAAGAATGA	(CTTT)11
27	AWLS4	107	F: GGCATAGGATGAAAAGGCAA; R: CAGGCTCCAGATCGGTAAAG	(ATG)11
28	BW1AV	223	F: CCTTCCCACCTCCTTTCCTA; R: CCTCTCCACAGAAAAGCCAG	(TTC)11
29	AGTDL	195	F: GAGGAGGCATATGAAGCCAT; R: CACCAACACTAATTGCACCG	(GTA)12
30	AI6YC	95	F: TCATCGTCTTGTCTCGTCTCA; R: TGAGTGAGTGTAGGAGAGTTGAGTG	(TCA)12
31	BJC5A	103	F: ACATGGACTTCCCCATCTTG; R: TGGGTAGAAGCCGTAGATGA	(TCA)12
32	BXW3K	116	F: GGAATTGGGCGCTATACAAT; R: AGCCGCTTGGTACTCAAAGA	(TTA)12
33	AFXR2	162	F: TTCGTGTGGAACCAAATTGA; R: GAAAAGGAACGACAACAACGA	(GTT)14
34	BUA8X	96	F: TTGCTGCTACAAAGGAAGAGAA; R: AGCACTACCACTAGCACCACTAC	(AGT)14
35	AJ8JC	160	F: GGTTTGTCGTTGTTGATAGTTCA; R: TGAACTGGCCAAGGTTAGGT	(GTT)16
36	BCRU5	145	F: AAGTGCACAGATGTTTCCCC; R: ACTCACCATCACCTTCAGGG	(ATC)13
37*	BN5YF	117	F: TCCATTTCACAAGAAACGGTATT; R: CATGTTCTCGCCTGTGGTTA	(CTA)9
38	B26YF	158	F: CAAAGACCCGATATTCACGG; R: GAAACAATAGCAACACAACGG	(GTT)9
39	ABFLV	121	F: TTATAACGCCGCAATCATCA; R: GTGAGTGGATAGGGGTGGTG	(CAT)9
40	ANPY6	137	F: ACGCACATTGCAACAGTGAT; R: TAGAGTCAGAAGCAAGGCCG	(TAG)9
41	BV58Z	114	F: TCAAGGTTAGTTTCACCATCACC; R: TAGTGATAGTGGCAGTGGCG	(CAT)9
42*	APQ1T	129	F: TCAACTGGGAACAAACATCG; R: ATGTCATTGGTTGACGTTGG	(ATG)10
43*	BNDY8	167	F: TGCTACCACTGCTACTATGCTATTG; R: CATCCACGATCCACAGTCAA	(CTGCTA)9
44*	A0Z98	175	F: CAGACAGACAAACAGACAGACAAA; R: TGTATCTATCTACGTCTAGCCTTCTG	(ACAG)5(ATAG)7(GTA)6
45	BDSXV	211	F: TGAGTACCGTAGGCAGAGCC; R: GGATCGCATTTCCTGAAGAG	(TAC)8
46	B1OR6	229	F: AGGTCGAACCCTTAACCTCG; R: AATGGTTTTAGCCTTGGCCT	(CAC)8
47	BSV4N	190	F: CGCGTTTAGAGCCTGTATTTG; R: TGGGGAGGAGTCAAAGAATG	(ATT)8
48*	AARB4	96	F: TTGAAGTCCACGGTTTAGCA; R: TGAGGCTACTAGACACAAGTACAAAGA	(TTTG)5
49	BC0V2	155	F: GCAGCCAACAAGGAAGGATA; R: GGTAACCATGGATTTGACGC	(AGGTGA)5
50	AM33L	150	F: ACCCGGGATCCAGAACTTT; R: AACTTGCGAGTCGATCTGGT	(TCTAGA)5

51	B140P	174	F: ACTGAATGGGGTCGAGTACG; R: TTTGTGCTAAAACCTTCCCG	(GTTT)5
52	BD22A	162	F: GCTCACTCATCGCATTCATC; R: ATGTGCCTTTGTGCTGTCAC	(CTA)8
53	BZ6L5	126	F: TTTCTGTTGCACCGTTTCAG; R: TAACGCTTTGAGCTCTTCGG	(TAA)9
54*	AEIT1	157	F: TGTTGGCCTACGAAACAATG; R: CGACCTCTCACTTCCTTTGC	(TAC)8
61*	ACW30	159	F: TGTTCTTGATTCGTTATGGGG; R: TGTCCGATGGTGAGCAGTAA	(TTG)8
65*	AH92H	127	F: AGTCGATATTATGGCGTCGG; R: TTGTTTAAACTTGTTTGAGAAAAGG	(CTT)9(CTC)6
66*	ATRX6	183	F: TTTAACTGGCTGTAGGCCAAA; R: CGAAGAGAAAGCCGGTAGTG	(CTT)9(CTC)6
67*	A6SU2	118	F: CTGTCTTGGCCATAGGCTCT; R: GGCATAGGATGAAAAGGCAA	(TCA)18
87*	A3VRK	142	F: ATTCTGGTTGTCTCGTTGCC; R: ACCGGTAACCGTAACTAACGTA	(TG)25(GT)6
88	BRHL3	162	F: GGAGAAATAGGGAGAACGCA; R: GGTCGTGAGGGTCGAGTCTA	(GA)18
89	BTPKD	162	F: CAATAACAAAGGCCTGGGAA; R: GGCATGTCGATTGCACTAAA	(AG)14
90	BRQ3Q	179	F: TGGAAGGTTGTGTATCAAGCA; R: ATCCATCAAATGGTGGCAAT	(AC)14
91	A325W	234	F: CAAGGTGGGTTGAGTCGTTT; R: GGTGTGGTGCGAAAACATAG	(GT)14
92	A70H4	194	F: CACCTTCATTCAACCAAAGGA; R: GCGTGGAGTTGGGAGTAGAG	(CA)13
93	BHFLC	220	F: TTTGTCTGTAGTTGGCCCGT; R: TCACAAAGCAGACGCTTGTT	(GA)11
94	BT1RC	133	F: TCTTGAGCTTGCATTGTTGG; R: GGGGAGTGAGTGAGAACCAG	(GTT)8

* Interrupted and/or compound repeats; otherwise perfect repeat motifs.

Supplementary Table 2 Summary statistics for the 12 microsatellites and 5 populations of *P. maculata*. *N*: Sample size. *Na*: Number of alleles. A_R and PA_R: Allelic richness and private richness, respectively based on 18 diploid individuals. Ho: observed heterozygosity. H_e: unbiased expected heterozygosity. G_{IS}: Nei's (1987) inbreeding coefficient. *P*-HWE: P values for deviations from HWE, [a] Significant deviation from HWE (P<0.05).

Pop	**Locus**	***N***	***Na***	A_R	PA_R	H_o	H_e	G_{IS}	**P-HWE**
Ti Point									
	1	20	9	9,492	0,028	0.800	0.771	-0.039	0.500
	2	20	8	5,892	0,173	0.600	0.724	0.175	0.108
	20	20	6	6,792	0,036	0.750	0.776	0.034	0.476
	26	20	5	3,900	0,600	0.650	0.695	0.066	0.415
	27	20	9	8,600	0,491	0.800	0.756	-0.059	0.428
	32	20	5	5,892	0,060	0.800	0.703	-0.143	0.211
	38	20	8	6,885	0,164	0.900	0.829	-0.087	0.299
	42	20	5	4,800	0,027	0.600	0.544	-0.107	0.359
	51	20	5	3,800	0,067	0.750	0.600	-0.258	0.091
	65	20	2	2,900	0,000	0.400	0.508	0.216	0.295

	89	20	8	9,399	0,273	0.700	0.705	0.007	0.594
	61	20	3	3,900	0,000	0.400	0.550	0.278	0.082
	Mean	20	6.083	6,021	0.282	0.679	0.680	0.001	0.490
	SE	0	2.314	0.655	0.297	0.157	0.105	-	-
Auckland									
	1	30	12	9,815	0,028	0.833	0.780	-0.070	0.274
	2	30	7	7,449	0,173	0.667	0.714	0.067	0.324
	20	30	7	5,444	0,036	0.667	0.746	0.108	0.185
	26	30	4	4,541	0,600	0.733	0.701	-0.047	0.440
	27	30	9	9,025	0,491	0.700	0.720	0.029	0.467
	32	30	6	5,436	0,060	0.767	0.738	-0.039	0.445
	38	30	8	7,133	0,164	0.900	0.831	-0.084	0.224
	42	30	6	5,140	0,027	0.600	0.597	-0.005	0.597
	51	30	4	4,141	0,067	0.400	0.473	0.156	0.235
	65	30	3	2,000	0,000	0.567	0.493	-0.152	0.268
	89	30	10	7,404	0,273	0.700	0.653	-0.074	0.314
	61	30	4	3,000	0,000	0.633	0.559	-0.136	0.222
	Mean	30	6.667	5,877	0.460	0.681	0.667	-0.021	0.265
	SE	0	2.741	0.679	0.571	0.129	0.114	-	-
Tauranga									
	1	33	13	10,535	0,329	0.879	0.823	-0.069	0.250
	2	33	8	6,597	0,452	0.667	0.730	0.088	0.240
	20	33	6	5,501	0,033	0.848	0.710	-0.199	0.028
	26	33	5	4,527	0,000	0.758	0.669	-0.134	0.166
	27	33	11	9,651	0,070	0.818	0.768	-0.067	0.267
	32	33	5	4,544	0,000	0.727	0.731	0.005	0.551
	38	33	10	8,508	0,131	0.879	0.835	-0.053	0.335
	42	33	5	4,091	0,022	0.636	0.587	-0.085	0.319
	51	33	4	3,343	0,565	0.424	0.510	0.170	0.108
	65	33	2	2,000	0,000	0.455	0.506	0.103	0.407
	89	33	8	6,777	0,764	0.727	0.683	-0.067	0.330
	61	33	5	4,594	0,058	0.545	0.556	0.019	0.512
	Mean	33	6.833	5,889	0,202	0.697	0.676	-0.032	0.155
	SE	0	3.215	0.750	0.076	0.157	0.113	-	-
Wellington									
	1	18	13	10,535	1,621	0.833	0.921	0.097	0.163
	2	18	6	6,597	0,007	0.667	0.752	0.117	0.261
	20	18	5	5,501	0,000	0.556	0.632	0.124	0.308
	26	18	5	4,527	0,455	0.722	0.632	-0.148	0.257
	27	18	7	9,651	1,005	0.611	0.713	0.146	0.182
	32	18	5	4,544	0,000	0.722	0.675	-0.073	0.431
	38	18	9	8,508	0,248	0.778	0.830	0.065	0.364
	42	18	5	4,091	0,077	0.556	0.617	0.103	0.348
	51	18	4	3,343	0,601	0.333	0.346	0.038	0.512
	65	18	2	2,000	0,000	0.389	0.322	-0.214	0.489
	89	18	9	6,777	1,587	0.667	0.730	0.089	0.301

	61	18	2	4,594	0,000	0.444	0.457	0.029	0.655
	Mean	18	6	5,889	0,467	0.606	0.636	0.047	0.139
	SE	0	3.133	0.750	0.178	0.156	0.182	-	-
Nelson									
	1	45	21	15,254	3,843	0.867	0.919	0.057	0.146
	2	45	8	6,552	0,182	0.756	0.791	0.045	0.330
	20	45	10	6,868	1,234	0.867	0.738	-0.177	0.015[a]
	26	45	3	3,000	0,000	0.689	0.601	-0.148	0.123
	27	45	11	7,113	1,276	0.733	0.721	-0.018	0.500
	32	45	5	4,693	0,000	0.667	0.648	-0.029	0.462
	38	45	9	7,450	0,473	0.733	0.737	0.005	0.547
	42	45	5	3,799	0,800	0.467	0.534	0.127	0.155
	51	45	6	4,145	0,403	0.378	0.333	-0.136	0.156
	65	44	4	2,806	0,818	0.227	0.208	-0.093	0.672
	89	43	7	5,820	0,000	0.721	0.691	-0.044	0.476
	61	44	4	3,616	0,208	0.591	0.473	-0.253	0.03[a]
	Mean	44.667	7.750	5,926	0,770	0.641	0.616	-0.041	0.114
	SE	0.651	4.883	0.973	0.309	0.194	0.201	-	-

Supplementary Table 3 Estimating the null allele frequency using the EN algorithm.

Locus	Population	Null allele freq	Locus	Population	Null allele freq
1	Ti Point	0	38	Ti Point	0
1	Auckland	0	38	Auckland	0
1	Tauranga	0	38	Tauranga	0
1	Wellington	0.02735	38	Wellington	0
1	Nelson	0.00504	38	Nelson	0.00001
2	Ti Point	0.0646	42	Ti Point	0.00001
2	Auckland	0.05122	42	Auckland	0
2	Tauranga	0.04664	42	Tauranga	0
2	Wellington	0.03764	42	Wellington	0.00051
2	Nelson	0.02708	42	Nelson	0.03862
20	Ti Point	0.02094	51	Ti Point	0
20	Auckland	0.06045	51	Auckland	0.04422
20	Tauranga	0	51	Tauranga	0.04277
20	Wellington	0.01853	51	Wellington	0
20	Nelson	0	51	Nelson	0
26	Ti Point	0.00001	65	Ti Point	0.06388
26	Auckland	0	65	Auckland	0
26	Tauranga	0	65	Tauranga	0.02919
26	Wellington	0	65	Wellington	0
26	Nelson	0	65	Nelson	0.00001
27	Ti Point	0.00001	89	Ti Point	0
27	Auckland	0.00001	89	Auckland	0
27	Tauranga	0	89	Tauranga	0
27	Wellington	0.04675	89	Wellington	0.00703
27	Nelson	0	89	Nelson	0

32	Ti Point	0	61	Ti Point	0.09584
32	Auckland	0.00006	61	Auckland	0
32	Tauranga	0	61	Tauranga	0.00001
32	Wellington	0	61	Wellington	0.00058
32	Nelson	0	61	Nelson	0

Supplementary Table 4 The estimation of the possible effect of null alleles on population structure analysis. Estimated **A**) global F_{ST} of Weir (1996) for each microsatellite locus and **B**) pairwise F_{ST} for each pair of populations with and without using the ENA correction described in Chapuis and Estoup (2007).

A

All loci

F_{ST} not using ENA	F_{ST} using ENA
0.06649	0.066884

Per locus

Locus	F_{ST} not using ENA	F_{ST} using ENA
1	0.06369	0.06397
2	0.01425	0.01452
20	-0.00360	-0.00290
26	0.12868	0.12867
27	0.04845	0.04779
32	0.07372	0.07372
38	0.03052	0.03052
42	-0.01015	-0.01040
51	0.05831	0.05943
65	0.23490	0.23507
89	0.14677	0.14647
61	0.03309	0.03446

B

Overall loci

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00928			
Tauranga	-0.00364	-0.00352		
Wellington	0.08740	0.09514	0.07380	
Nelson	0.11826	0.12138	0.09684	0.00841

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00854			
Tauranga	-0.00297	-0.00321		
Wellington	0.08842	0.09607	0.07409	

Nelson	0.11876	0.12181	0.09708	0.00740

Loc *1*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00814			
Tauranga	-0.00445	-0.00238		
Wellington	0.03627	0.05243	0.03302	
Nelson	0.12232	0.12038	0.09669	0.03120

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00814			
Tauranga	-0.00445	-0.00238		
Wellington	0.03964	0.05514	0.03496	
Nelson	0.12255	0.12035	0.09662	0.02913

Loc *2*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01377			
Tauranga	-0.00244	0.00623		
Wellington	0.01528	0.01635	0.00533	
Nelson	0.03889	0.03542	0.01737	-0.01069

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01036			
Tauranga	0.00212	0.00546		
Wellington	0.01782	0.01534	0.00741	
Nelson	0.03884	0.03285	0.01715	-0.0101

Loc *20*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01736			
Tauranga	0.00207	-0.00391		
Wellington	0.01770	0.00869	-0.00552	
Nelson	-0.00349	-0.00500	-0.00916	-0.00115

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01558			
Tauranga	0.00245	-0.00061		
Wellington	0.01553	0.01255	-0.00720	
Nelson	-0.00349	-0.00323	-0.00916	-0.00317

Loc *26*

F_{ST} not using ENA

Loc *38*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01019		
Tauranga	-0.00742	-0.01078	
Wellington	0.03127	0.03603	0.03287
Nelson	0.06274	0.06426	0.04942

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01019		
Tauranga	-0.00742	-0.01078	
Wellington	0.03127	0.03603	0.03287
Nelson	0.06273	0.06426	0.04942

Loc *42*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01634		
Tauranga	-0.00983	-0.01196	
Wellington	-0.00944	-0.00898	-0.00524
Nelson	-0.01739	-0.01029	-0.00789

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01634		
Tauranga	-0.00983	-0.01196	
Wellington	-0.00941	-0.00898	-0.00525
Nelson	-0.01583	-0.01135	-0.0088

Loc *51*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.00105		
Tauranga	-0.00930	-0.01481	
Wellington	0.09859	0.05872	0.08405
Nelson	0.13058	0.08571	0.11402

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.00463		
Tauranga	-0.01119	-0.01356	
Wellington	0.09859	0.06197	0.08781
Nelson	0.13058	0.08808	0.11696

Loc *65*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01280			
Tauranga	-0.00816	0.01072		
Wellington	0.20549	0.19794	0.18510	
Nelson	0.20794	0.19738	0.17636	-0.00247

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.01280			
Tauranga	-0.00816	0.01072		
Wellington	0.20549	0.19794	0.18600	
Nelson	0.20794	0.19738	0.17636	-0.00247

Loc *27*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	0.00426			
Tauranga	0.00714	-0.00739		
Wellington	0.00855	0.02062	0.01338	
Nelson	0.08332	0.09188	0.08194	0.06017

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	0.00426			
Tauranga	0.00714	-0.00739		
Wellington	0.00734	0.01915	0.01125	
Nelson	0.08332	0.09188	0.08194	0.05544

Loc *32*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00036			
Tauranga	0.01972	-0.00751		
Wellington	0.13717	0.08211	0.06428	
Nelson	0.17522	0.11636	0.09915	-0.01553

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga	Wellington
Auckland	-0.00036			
Tauranga	0.01972	-0.00751		
Wellington	0.13717	0.08211	0.06428	
Nelson	0.17522	0.11654	0.09915	-0.01553

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.00918		
Tauranga	-0.02274	0.00271	
Wellington	0.21415	0.29334	0.18921
Nelson	0.39282	0.45508	0.34767

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.00614		
Tauranga	-0.02047	0.00348	
Wellington	0.21485	0.29334	0.18954
Nelson	0.39119	0.45508	0.34704

Loc *89*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01209		
Tauranga	-0.00494	-0.00859	
Wellington	0.21691	0.25091	0.22738
Nelson	0.20447	0.22357	0.20159

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01209		
Tauranga	-0.00494	-0.00859	
Wellington	0.21588	0.24973	0.22607
Nelson	0.20447	0.22357	0.20159

Loc *61*

F_{ST} not using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01553		
Tauranga	-0.01601	0.00511	
Wellington	0.05089	0.10207	0.03326
Nelson	0.04613	0.08638	0.01938

F_{ST} using ENA

pop	Ti-Point	Auckland	Tauranga
Auckland	-0.01131		
Tauranga	-0.01271	0.00511	
Wellington	0.05766	0.10196	0.03319
Nelson	0.17522	0.11636	

Supplementary Table 5 Results of the linkage equilibrium test for 12 microsatellite markers. The calculations were performed for each population and across the populations. *P* values were estimated by 6600 permutations. The *P*-value for the 5% nominal level is 0.000152 after Bonferroni correction. Linkage disequilibrium between any pair of loci was not observed after the correction.

Loci	Ti Point	Auckland	Tauranga	Wellington	Nelson	Overall
1 X 2	0.81924	0.98515	0.72652	1.0000	0.31399	0.86106
1 X 20	0.87697	0.72152	0.40970	0.30455	1.00000	0.68409
1 X 26	0.91530	0.21091	0.79197	1.00000	0.67792	0.64470
1 X 27	0.63227	0.83258	0.00712	0.20924	1.0000	0.15894
1 X 32	0.86636	0.98682	0.69621	0.38303	0.69394	0.87106
1 X 38	0.67030	0.53621	0.18682	1.0000	0.24057	0.12500
1 X 42	0.57409	0.38455	0.88758	1.0000	0.89792	0.81909
1 X 51	0.93909	0.20167	0.58818	0.72318	0.01661	0.11561
1 X 65	0.88712	0.59167	0.25742	0.74697	0.64158	0.58985
1 X 89	0.16152	0.60909	0.70652	1.00000	0.46336	0.40985
1 X 61	0.99303	0.76742	0.35864	1.00000	0.34141	0.76121
2 X 20	0.96742	0.84000	0.935	0.56803	0.30578	0.84515
2 X 26	0.97136	0.52379	0.39379	0.0303	0.94727	0.6997
2 X 27	0.07712	0.14515	0.63091	0.75333	0.62534	0.19379
2 X 32	0.95242	0.43364	0.92227	0.70621	0.73302	0.91076
2 X 38	0.80182	0.88712	0.97106	0.37455	0.28805	0.72182
2 X 42	0.73909	0.68227	0.36955	0.92682	0.22655	0.51712
2 X 51	0.45636	0.53545	0.52742	0.57909	0.0888	0.22121
2 X 65	0.38515	0.89288	0.98894	0.97879	0.81632	0.99318
2 X 89	0.54379	0.75621	0.90848	1.0000	0.89751	0.96848
2 X 61	0.74939	0.84788	0.40939	0.62955	0.66934	0.81591
20 X 26	0.60606	0.45515	0.05409	0.98894	0.47701	0.34727
20 X 27	0.84015	0.73182	0.92091	0.11515	0.45496	0.54136
20 X 32	0.82379	0.84076	0.17985	0.66015	0.87339	0.73318
20 X 38	0.07788	0.93955	0.35591	1.00000	0.45026	0.42500
20 X 42	0.27030	0.32727	0.92561	0.50909	0.59180	0.60515
20 X 51	0.79879	0.16045	0.53364	0.36879	0.45765	0.37439
20 X 65	0.80000	0.39197	0.66409	1.00000	0.07875	0.45106
20 X 89	0.92894	0.10667	0.23833	0.72788	0.96035	0.63106
20 X 61	0.20318	0.15273	0.61636	0.59364	0.90050	0.53152
26 X 27	1.00000	0.70152	0.71167	0.85561	0.51109	0.84545
26 X 32	0.54348	0.94273	0.59061	0.94500	0.39336	0.84621
26 X 38	1.00000	0.54333	0.00121	0.25394	0.70118	0.13955
26 X 42	1.00000	0.21485	0.19121	0.06667	0.55496	0.30712
26 X 51	0.77242	0.53833	0.71455	0.27606	0.56631	0.6597
26 X 65	0.00879	0.02636	0.67288	0.69576	0.12984	0.01409
26 X 89	0.73030	0.47788	0.47530	1.0000	0.69422	0.80939
26 X 61	0.89939	0.22409	0.98470	0.13197	0.95692	0.91364
27 X 32	0.39606	0.51591	0.84591	0.47470	0.08904	0.18091
27 X 38	0.60424	0.82742	0.79470	1.00000	0.95592	0.95712

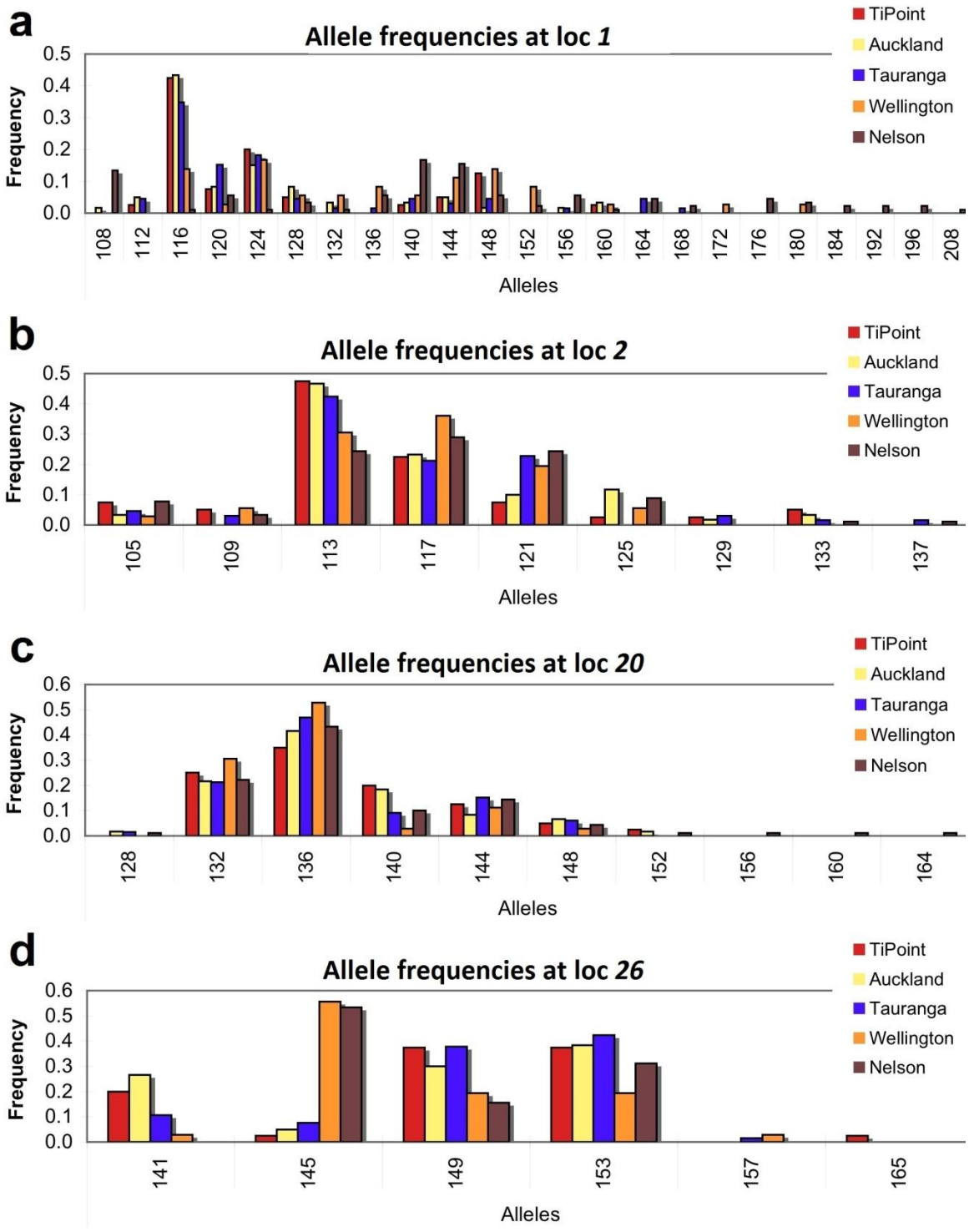
27 X 42	0.84879	0.52227	0.67909	0.35212	0.08052	0.25242
27 X 51	0.99303	0.68667	0.59333	0.07121	0.09781	0.31076
27 X 65	0.75864	0.37939	0.95409	0.24833	0.70624	0.77818
27 X 89	0.09667	0.71636	0.35985	0.19970	0.02113	0.02288
27 X 61	0.84000	0.90455	0.11394	0.35348	0.58146	0.53424
32 X 38	0.82379	0.98500	1.00000	1.00000	0.31481	0.95439
32 X 42	0.77091	0.05773	0.22424	0.72167	0.25988	0.11061
32 X 51	0.07955	0.49167	0.50833	0.80409	0.71723	0.50788
32 X 65	0.84061	0.65333	0.98455	0.64727	0.79675	0.98318
32 X 89	0.61242	0.85379	0.75061	0.00515	0.54974	0.41394
32 X 61	0.03318	0.05667	0.37500	0.88758	0.35618	0.05712
38 X 42	0.57909	0.99152	0.11000	0.72485	0.99279	0.93455
38 X 51	0.73439	0.59318	0.70939	0.08424	0.1661	0.23985
38 X 65	0.33591	0.75470	0.61727	1.00000	0.44158	0.73500
38 X 89	0.76773	0.64000	0.82091	1.00000	0.82685	0.91167
38 X 61	0.16106	0.31879	0.17727	0.71742	0.31657	0.09000
42 X 51	0.08076	0.33773	0.3953	0.51788	0.86564	0.40697
42 X 65	0.29424	0.57515	0.62242	0.77136	0.46031	0.63106
42 X 89	0.17121	0.18606	0.56030	0.78379	0.56401	0.2997
42 X 61	0.29303	0.08439	0.74000	0.76545	0.23881	0.22818
51 X 65	0.47697	0.07864	0.80015	0.38015	0.59897	0.41864
51 X 89	0.84591	0.57818	0.18985	0.80076	0.48462	0.57348
51 X 61	0.09712	0.12939	0.26439	0.83970	0.23129	0.09894
6 5X 89	0.66167	0.75667	0.05985	0.31667	0.70504	0.43273
6 5X 61	0.41848	0.47576	0.17136	1.00000	0.51053	0.45818
89 X 61	0.43697	0.29515	0.10303	0.32955	0.20853	0.06288

Supplementary Table 6 Allele frequencies at 12 microsatellite loci by population.

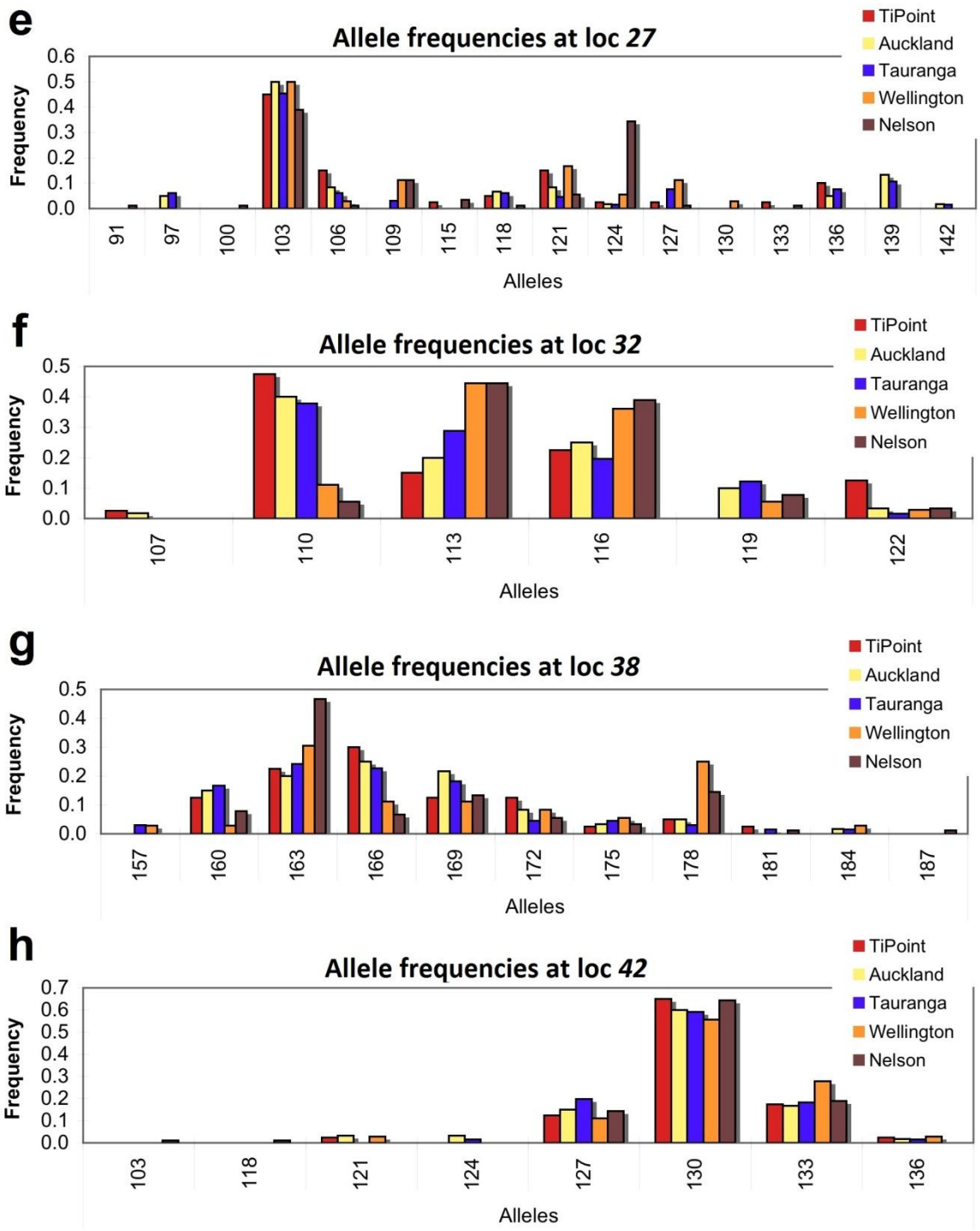
Locus	Allele (bp)	Ti Point	Auckland	Tauranga	Wellington	Nelson
1	**Sample size**	20	30	33	18	45
	108	0.000	0.017	0.000	0.000	0.133
	112	0.025	0.050	0.045	0.000	0.000
	116	0.425	0.433	0.348	0.139	0.011
	120	0.075	0.083	0.152	0.028	0.056
	124	0.200	0.150	0.182	0.167	0.011
	128	0.050	0.083	0.045	0.056	0.033
	132	0.000	0.033	0.015	0.056	0.011
	136	0.000	0.000	0.015	0.083	0.056
	140	0.025	0.033	0.045	0.056	0.167
	144	0.050	0.050	0.030	0.111	0.156
	148	0.125	0.017	0.045	0.139	0.056
	152	0.000	0.000	0.000	0.083	0.022
	156	0.000	0.017	0.015	0.000	0.056
	160	0.025	0.033	0.000	0.028	0.011
	164	0.000	0.000	0.045	0.000	0.044
	168	0.000	0.000	0.015	0.000	0.022
	172	0.000	0.000	0.000	0.028	0.000
	176	0.000	0.000	0.000	0.000	0.044
	180	0.000	0.000	0.000	0.028	0.033
	184	0.000	0.000	0.000	0.000	0.022
	192	0.000	0.000	0.000	0.000	0.022
	196	0.000	0.000	0.000	0.000	0.022
	208	0.000	0.000	0.000	0.000	0.011
2	**N**	20	30	33	18	45
	105	0.075	0.033	0.045	0.028	0.078
	109	0.050	0.000	0.030	0.056	0.033
	113	0.475	0.467	0.424	0.306	0.244
	117	0.225	0.233	0.212	0.361	0.289
	121	0.075	0.100	0.227	0.194	0.244
	125	0.025	0.117	0.000	0.056	0.089
	129	0.025	0.017	0.030	0.000	0.000
	133	0.050	0.033	0.015	0.000	0.011
	137	0.000	0.000	0.015	0.000	0.011
20	**N**	20	30	33	18	45
	128	0.000	0.017	0.015	0.000	0.011
	132	0.250	0.217	0.212	0.306	0.222
	136	0.350	0.417	0.470	0.528	0.433
	140	0.200	0.183	0.091	0.028	0.100
	144	0.125	0.083	0.152	0.111	0.144
	148	0.050	0.067	0.061	0.028	0.044
	152	0.025	0.017	0.000	0.000	0.011
	156	0.000	0.000	0.000	0.000	0.011
	160	0.000	0.000	0.000	0.000	0.011

	164	0.000	0.000	0.000	0.000	0.011
26	**N**	20	30	33	18	45
	141	0.200	0.267	0.106	0.028	0.000
	145	0.025	0.050	0.076	0.556	0.533
	149	0.375	0.300	0.379	0.194	0.156
	153	0.375	0.383	0.424	0.194	0.311
	157	0.000	0.000	0.015	0.028	0.000
	165	0.025	0.000	0.000	0.000	0.000
27	**N**	20	30	33	18	45
	91	0.000	0.000	0.000	0.000	0.011
	97	0.000	0.050	0.061	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.011
	103	0.450	0.500	0.455	0.500	0.389
	106	0.150	0.083	0.061	0.028	0.011
	109	0.000	0.000	0.030	0.111	0.111
	115	0.025	0.000	0.000	0.000	0.033
	118	0.050	0.067	0.061	0.000	0.011
	121	0.150	0.083	0.045	0.167	0.056
	124	0.025	0.017	0.015	0.056	0.344
	127	0.025	0.000	0.076	0.111	0.011
	130	0.000	0.000	0.000	0.028	0.000
	133	0.025	0.000	0.000	0.000	0.011
	136	0.100	0.050	0.076	0.000	0.000
	139	0.000	0.133	0.106	0.000	0.000
	142	0.000	0.017	0.015	0.000	0.000
32	**N**	20	30	33	18	45
	107	0.025	0.017	0.000	0.000	0.000
	110	0.475	0.400	0.379	0.111	0.056
	113	0.150	0.200	0.288	0.444	0.444
	116	0.225	0.250	0.197	0.361	0.389
	119	0.000	0.100	0.121	0.056	0.078
	122	0.125	0.033	0.015	0.028	0.033
38	**N**	20	30	33	18	45
	157	0.000	0.000	0.030	0.028	0.000
	160	0.125	0.150	0.167	0.028	0.078
	163	0.225	0.200	0.242	0.306	0.467
	166	0.300	0.250	0.227	0.111	0.067
	169	0.125	0.217	0.182	0.111	0.133
	172	0.125	0.083	0.045	0.083	0.056
	175	0.025	0.033	0.045	0.056	0.033
	178	0.050	0.050	0.030	0.250	0.144
	181	0.025	0.000	0.015	0.000	0.011
	184	0.000	0.017	0.015	0.028	0.000
	187	0.000	0.000	0.000	0.000	0.011
42	**N**	20	30	33	18	45
	103	0.000	0.000	0.000	0.000	0.011

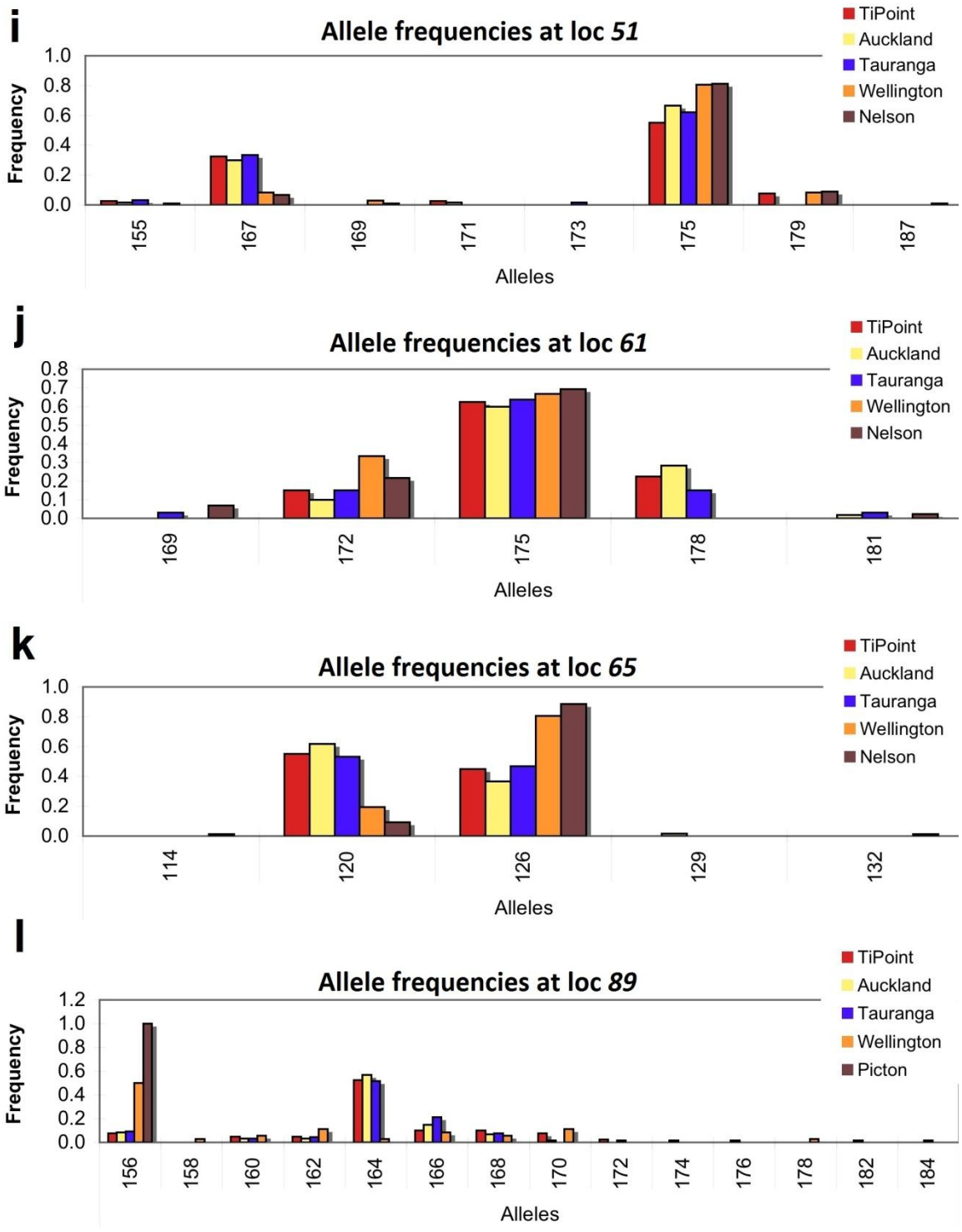
	118	0.000	0.000	0.000	0.000	0.011
	121	0.025	0.033	0.000	0.028	0.000
	124	0.000	0.033	0.015	0.000	0.000
	127	0.125	0.150	0.197	0.111	0.144
	130	0.650	0.600	0.591	0.556	0.644
	133	0.175	0.167	0.182	0.278	0.189
	136	0.025	0.017	0.015	0.028	0.000
51	**N**	20	30	33	18	45
	155	0.025	0.017	0.030	0.000	0.011
	167	0.325	0.300	0.333	0.083	0.067
	169	0.000	0.000	0.000	0.028	0.011
	171	0.025	0.017	0.000	0.000	0.000
	173	0.000	0.000	0.015	0.000	0.000
	175	0.550	0.667	0.621	0.806	0.811
	179	0.075	0.000	0.000	0.083	0.089
	187	0.000	0.000	0.000	0.000	0.011
65	**N**	20	30	33	18	44
	114	0.000	0.000	0.000	0.000	0.011
	120	0.550	0.617	0.530	0.194	0.091
	126	0.450	0.367	0.470	0.806	0.886
	129	0.000	0.017	0.000	0.000	0.000
	132	0.000	0.000	0.000	0.000	0.011
89	**N**	20	30	33	18	43
	156	0.075	0.083	0.091	0.500	0.512
	158	0.000	0.000	0.000	0.028	0.012
	160	0.050	0.033	0.030	0.056	0.140
	162	0.050	0.033	0.045	0.111	0.093
	164	0.525	0.567	0.515	0.028	0.116
	166	0.100	0.150	0.212	0.083	0.116
	168	0.100	0.067	0.076	0.056	0.012
	170	0.075	0.017	0.000	0.111	0.000
	172	0.025	0.000	0.015	0.000	0.000
	174	0.000	0.017	0.000	0.000	0.000
	176	0.000	0.017	0.000	0.000	0.000
	178	0.000	0.000	0.000	0.028	0.000
	182	0.000	0.017	0.000	0.000	0.000
	184	0.000	0.000	0.015	0.000	0.000
61	**N**	20	30	33	18	44
	169	0.000	0.000	0.030	0.000	0.068
	172	0.150	0.100	0.152	0.333	0.216
	175	0.625	0.600	0.636	0.667	0.693
	178	0.225	0.283	0.152	0.000	0.000
	181	0.000	0.017	0.030	0.000	0.023



Supplementary Figure 1 Allele frequencies for each population at each locus.



Supplementary Figure (continues)



Supplementary Figure (continues)

Supplementary Table 7 Pairwise population matrix of the *Dest* values at each locus. *Dest* values below the diagonal. The *P* values calculated with 9999 permutations are shown above the diagonal.

Loc *1*

	TP	AKL	TR	WL	NL
TP	-----	0.757	0.586	0.023	0.000
AKL	-0.028	-----	0.525	0.002	0.000
TR	-0.017	-0.009	-----	0.010	0.000
WL	0.205	0.302	0.222	-----	0.002
NL	0.793	0.787	0.730	0.374	-----

Loc *2*

	TP	AKL	TR	WL	NL
TP	-----	0.843	0.449	0.193	0.021
AKL	-0.036	-----	0.228	0.146	0.010
TR	-0.007	0.016	-----	0.284	0.063
WL	0.044	0.045	0.015	-----	0.743
NL	0.129	0.113	0.057	-0.037	-----

Loc *20*

	TP	AKL	TR	WL	NL
TP	-----	0.936	0.353	0.171	0.568
AKL	-0.054	-----	0.561	0.245	0.678
TR	0.004	-0.011	-----	0.548	0.979
WL	0.044	0.021	-0.013	-----	0.438
NL	-0.013	-0.015	-0.024	-0.004	-----

Loc *26*

	TP	AKL	TR	WL	NL
TP	-----	0.734	0.608	0.000	0.000
AKL	-0.030	-----	0.164	0.000	0.000
TR	-0.019	0.023	-----	0.000	0.000
WL	0.510	0.498	0.426	-----	0.437
NL	0.466	0.449	0.368	-0.004	-----

Loc *27*

	TP	AKL	TR	WL	NL
TP	-----	0.281	0.199	0.236	0.000
AKL	0.012	-----	0.810	0.097	0.000
TR	0.023	-0.021	-----	0.129	0.000
WL	0.024	0.053	0.038	-----	0.007
NL	0.254	0.261	0.258	0.161	-----

Loc *32*

	TP	AKL	TR	WL	NL
TP	-----	0.390	0.102	0.000	0.000
AKL	0.000	-----	0.663	0.002	0.000
TR	0.052	-0.021	-----	0.012	0.000
WL	0.349	0.218	0.165	-----	0.916
NL	0.433	0.293	0.242	-0.029	-----

Loc *38*

	TP	AKL	TR	WL	NL
TP	-----	0.816	0.702	0.052	0.001
AKL	-0.049	-----	0.945	0.015	0.000
TR	-0.036	-0.053	-----	0.021	0.001
WL	0.157	0.181	0.167	-----	0.180
NL	0.235	0.245	0.188	0.031	-----

Loc *42*

	TP	AKL	TR	WL	NL
TP	-----	0.909	0.649	0.566	0.932
AKL	-0.021	-----	0.893	0.568	0.788
TR	-0.012	-0.017	-----	0.488	0.654
WL	-0.013	-0.014	-0.009	-----	0.469
NL	-0.019	-0.013	-0.010	-0.007	-----

Loc *51*

	TP	AKL	TR	WL	NL
TP	-----	0.327	0.506	0.007	0.000
AKL	-0.001	-----	0.765	0.047	0.002
TR	-0.010	-0.014	-----	0.020	0.000
WL	0.099	0.046	0.074	-----	1.000
NL	0.115	0.060	0.090	-0.009	-----

Loc *65*

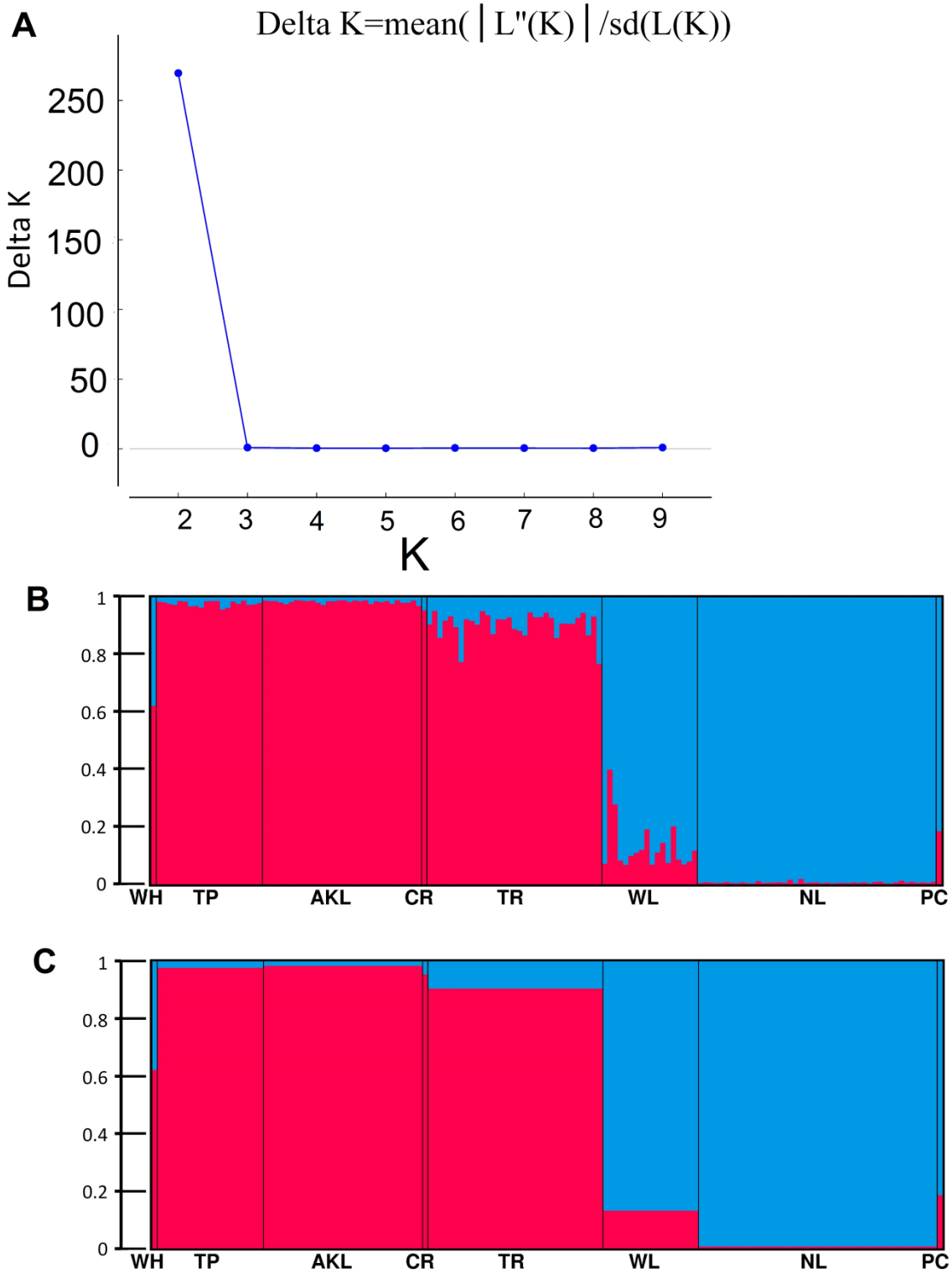
	TP	AKL	TR	WL	NL
TP	-----	0.500	0.928	0.003	0.000
AKL	-0.010	-----	0.282	0.000	0.000
TR	-0.023	0.003	-----	0.002	0.000
WL	0.197	0.300	0.178	-----	0.148
NL	0.301	0.414	0.278	0.006	-----

Loc *89*

	TP	AKL	TR	WL	NL
TP	-----	0.874	0.561	0.000	0.000
AKL	-0.026	-----	0.814	0.000	0.000
TR	-0.012	-0.017	-----	0.000	0.000
WL	0.706	0.738	0.699	-----	0.332
NL	0.590	0.591	0.553	0.005	-----

Loc *61*

	TP	AKL	TR	WL	NL
TP	-----	0.806	0.811	0.073	0.020
AKL	-0.020	-----	0.258	0.003	0.000
TR	-0.020	0.007	-----	0.077	0.058
WL	0.056	0.120	0.037	-----	0.269
NL	0.047	0.098	0.020	0.003	-----



Supplementary Figure 2 Results of Bayesian clustering analysis where the sampling location has not been introduced as a priori for the calculations. **A)** Plot of ΔK versus K indicating that the data is explained best by clusters K=2 (Evanno et al., 2005). **B)** Population structure in 149 individuals of *P. maculata* from 8 different sampling sites at K=2. Each individual is represented by a vertical line divided into two segments indicating proportional membership in the two clusters. **C)** Group assignments, indicating proportional membership in K=2 clusters.

Supplementary Table 8 Detection of first generation migrants. Likelihood ratio: L_home/L_max. The number of individuals with a probability below the threshold value is 4. The potential migrants are labelled in red ($P < 0.01$) and the most likely population in green. Ind: individual.

ID	Assigned	=-LOG (L_H /L_{max})	Probability	Ti-Point -log(L)	Auckland -log(L)	Tauranga -log(L)	Wellington -log(L)	Nelson -log(L)
37A	Ti Point	0.870	0.196	12.699	11.829	12.241	17.776	20.535
37B	Ti Point	0.135	0.329	15.807	16.863	15.671	22.433	24.073
37C	Ti Point	***4.803***	***0.002***	15.646	13.24	***10.843***	15.690	16.353
37D	Ti Point	0.703	0.220	12.836	12.833	12.133	14.555	16.744
37E	Ti Point	1.523	0.112	14.277	13.136	12.754	18.732	21.571
37F	Ti Point	0.000	0.679	12.738	13.805	13.579	19.384	19.033
37G	Ti Point	1.749	0.089	11.302	9.553	9.765	13.665	14.350
37H	Ti Point	2.32	0.055	12.957	10.637	12.437	13.967	15.300
38A	Ti Point	1.916	0.076	14.305	12.388	12.804	15.592	18.043
38B	Ti Point	0.553	0.241	12.928	12.375	12.793	19.507	22.379
38C	Ti Point	0.501	0.256	12.172	11.671	13.855	16.210	20.824
38D	Ti Point	2.207	0.056	13.790	11.583	12.245	18.428	21.726
38E	Ti Point	0.000	0.684	13.212	14.952	16.88	14.879	19.415
38F	Ti Point	0.238	0.307	18.196	17.958	19.186	18.721	23.898
38G	Ti Point	3.490	0.015	15.925	12.435	14.687	18.544	22.028
38H	Ti Point	1.342	0.131	16.388	16.018	15.046	17.940	22.345
38I	Ti Point	0.911	0.191	11.683	10.771	14.289	19.080	23.395
38J	Ti Point	0.000	0.680	15.212	16.03	16.423	19.309	20.151
38K	Ti Point	0.000	0.677	12.004	12.039	12.688	14.503	17.551
38L	Ti Point	0.000	0.682	11.871	12.664	13.147	16.008	19.460
NN1	Auckland	2.773	0.016	13.722	14.511	11.739	17.804	18.213
33A	Auckland	1.155	0.119	12.592	13.746	13.118	18.333	18.170
33B	Auckland	2.543	0.023	13.632	14.123	11.579	21.279	20.972
33C	Auckland	1.134	0.120	13.919	15.054	14.213	18.856	19.322
33D	Auckland	0.746	0.179	12.022	12.769	12.968	17.559	19.944
33E	Auckland	0.000	0.669	19.169	18.772	19.798	25.307	30.278
40A	Auckland	0.000	0.670	9.505	9.371	11.132	17.511	20.389
40B	Auckland	1.499	0.086	14.397	15.896	15.796	19.820	23.360
36A	Auckland	1.724	0.060	12.277	12.533	10.809	17.038	16.803
36B	Auckland	0.715	0.188	12.600	11.419	10.704	16.164	20.338
36C	Auckland	1.700	0.065	10.513	12.214	11.461	15.881	17.717
41A	Auckland	0.000	0.671	15.038	13.644	16.174	15.658	16.334
41B	Auckland	0.000	0.669	10.052	9.786	9.931	14.841	17.407
41C	Auckland	1.424	0.087	13.522	14.946	14.739	15.138	22.624
41D	Auckland	0.194	0.288	12.701	12.49	12.296	20.354	22.458
41E	Auckland	0.000	0.672	15.040	13.201	13.980	22.845	25.478
41F	Auckland	1.369	0.097	13.409	14.504	13.134	15.777	19.398
41G	Auckland	0.000	0.676	14.081	13.335	16.056	20.091	24.280
41H	Auckland	0.714	0.184	13.133	11.582	10.868	15.463	20.372
42A	Auckland	0.503	0.226	15.706	11.801	11.298	19.949	23.311
42B	Auckland	0.577	0.208	15.685	16.261	17.315	17.647	19.131
42C	Auckland	1.389	0.092	12.019	12.922	11.533	21.253	22.097
42D	Auckland	0.774	0.175	11.223	11.997	11.300	15.976	16.925

42E	Auckland	0.000	0.669	16.611	15.208	15.768	20.353	25.569
42F	Auckland	0.902	0.148	11.388	11.091	10.189	12.544	13.995
42G	Auckland	0.000	0.674	12.661	11.391	14.238	19.865	23.603
44A	Auckland	0.461	0.231	12.284	11.344	10.883	13.950	13.799
44B	Auckland	0.917	0.149	14.153	13.296	12.379	14.635	18.269
44C	Auckland	1.309	0.106	12.183	13.492	13.214	22.853	23.127
44D	Auckland	0.729	0.184	15.742	13.131	12.401	14.919	16.195
34A	Tauranga	0.000	0.649	14.483	11.962	11.458	14.618	14.582
34B	Tauranga	**3.300**	**0.005**	**10.412**	11.088	13.712	19.390	22.504
34C	Tauranga	0.000	0.652	17.488	18.383	15.671	15.918	17.399
34D	Tauranga	0.000	0.650	14.750	11.041	10.232	15.715	17.598
34E	Tauranga	0.000	0.652	9.742	10.906	9.148	13.876	16.688
34F	Tauranga	0.000	0.651	20.544	18.353	17.271	22.240	23.618
34G	Tauranga	0.000	0.652	16.792	16.664	14.369	15.469	15.426
34H	Tauranga	0.381	0.214	16.340	14.834	15.215	18.862	23.123
34I	Tauranga	0.178	0.264	11.933	13.369	12.111	17.146	16.695
34J	Tauranga	0.000	0.653	14.294	15.252	12.964	15.933	20.132
34K	Tauranga	0.482	0.199	11.371	11.12	11.602	18.191	20.082
34L	Tauranga	0.000	0.650	13.319	11.276	11.080	16.102	17.925
34M	Tauranga	0.000	0.653	13.039	12.338	10.682	15.041	12.121
34N	Tauranga	1.579	0.070	15.610	12.207	13.786	19.327	18.762
34O	Tauranga	0.000	0.649	14.937	13.157	12.727	21.232	21.812
47A	Tauranga	0.880	0.134	12.675	12.1	12.980	16.402	19.859
47B	Tauranga	1.191	0.097	15.503	15.796	16.694	17.146	18.834
47C	Tauranga	0.681	0.166	17.315	15.608	16.289	17.843	20.028
47D	Tauranga	0.676	0.165	14.708	15.119	14.492	13.816	17.496
47E	Tauranga	1.412	0.079	10.288	11.082	11.700	18.625	18.377
47F	Tauranga	0.711	0.163	12.396	12.346	13.057	18.636	21.027
NIA	Tauranga	0.875	0.137	11.671	9.069	9.944	14.729	17.415
NIB	Tauranga	2.652	0.015	10.988	11.343	13.639	19.154	21.359
NIC	Tauranga	0.114	0.275	10.235	9.618	9.732	14.707	14.496
NID	Tauranga	1.603	0.065	14.696	14.674	15.110	13.507	16.657
NIE	Tauranga	0.907	0.130	13.540	14.139	14.447	19.434	21.087
NIF	Tauranga	0.695	0.163	16.351	12.702	13.397	16.559	18.985
NIG	Tauranga	0.770	0.154	14.349	10.824	11.594	15.096	14.771
32B	Tauranga	0.194	0.258	11.164	10.724	10.918	15.717	18.170
32D	Tauranga	1.170	0.103	16.534	14.554	15.724	23.061	26.487
32E	Tauranga	0.000	0.650	15.335	15.436	14.957	17.283	17.758
32F	Tauranga	0.000	0.653	14.914	13.341	13.159	20.914	20.007
32G	Tauranga	1.023	0.118	14.656	16.303	15.452	14.428	15.170
43A	Wellington	1.413	0.066	15.626	14.881	14.793	10.590	9.177
43B	Wellington	3.166	0.014	12.647	14.072	11.518	14.683	15.318
43C	Wellington	2.833	0.018	14.227	17.344	15.931	17.060	16.586
43D	Wellington	0.000	0.594	15.345	15.151	15.360	11.730	12.407
43E	Wellington	0.000	0.593	21.716	21.868	20.096	14.915	16.082
43F	Wellington	1.205	0.076	18.499	17.495	16.761	15.103	13.898
43G	Wellington	2.765	0.017	17.196	18.855	18.067	17.353	14.588
43H	Wellington	0.000	0.598	17.990	16.889	15.698	12.585	15.265
43I	Wellington	0.000	0.596	16.295	17.964	18.296	13.088	17.514

43J	Wellington	0.408	0.136	18.355	17.872	17.609	13.471	13.063
43K	Wellington	0.000	0.593	18.773	21.42	19.942	13.920	16.637
43L	Wellington	0.000	0.591	16.270	17.327	15.386	13.207	13.436
43M	Wellington	0.764	0.110	17.599	18.428	15.735	12.484	11.720
43N	Wellington	0.000	0.597	17.465	18.816	16.732	16.237	17.241
43O	Wellington	0.000	0.592	15.812	17.072	15.468	11.940	11.952
43P	Wellington	0.000	0.594	19.846	20.828	19.698	13.268	14.783
43R	Wellington	0.25	0.149	15.061	14.631	13.749	11.304	11.054
43S	Wellington	1.096	0.082	15.010	14.333	13.630	13.286	12.190
45A	Nelson	0.000	0.540	22.515	23.241	22.351	20.298	15.822
45B	Nelson	0.000	0.540	17.124	16.146	13.598	12.244	10.778
45C	Nelson	0.000	0.538	17.532	18.036	17.187	14.948	13.692
45D	Nelson	0.000	0.539	16.544	18.178	17.453	11.326	10.495
45E	Nelson	0.533	0.047	25.330	24.198	21.983	16.576	17.108
45F	Nelson	**3.885**	**0.001**	15.273	16.056	14.757	**10.087**	13.971
45G	Nelson	0.000	0.540	20.261	21.235	20.657	17.630	13.505
45H	Nelson	0.000	0.538	16.141	15.637	14.188	9.280	8.821
45I	Nelson	0.000	0.540	19.091	18.092	18.264	16.982	12.195
45J	Nelson	0.000	0.540	20.595	20.629	19.484	13.119	11.247
45K	Nelson	0.000	0.539	20.146	19.211	16.868	12.223	9.291
45L	Nelson	0.282	0.059	19.005	18.625	17.601	15.755	16.037
45M	Nelson	0.169	0.065	14.355	14.799	15.262	10.263	10.432
45N	Nelson	0.965	0.029	16.239	18.043	14.808	11.034	11.999
45O	Nelson	0.000	0.539	17.738	17.887	14.133	13.110	11.345
45P	Nelson	0.000	0.537	16.718	17.139	17.699	15.647	15.001
45R	Nelson	0.000	0.539	23.276	22.46	20.047	16.985	15.539
45S	Nelson	**2.442**	**0.005**	16.205	17.889	17.152	**13.140**	15.582
45T	Nelson	0.000	0.540	21.687	20.648	20.367	15.648	10.760
45U	Nelson	0.000	0.538	18.753	17.728	18.104	17.576	17.357
24A	Nelson	0.000	0.539	17.758	16.14	15.638	12.583	11.689
24AA	Nelson	0.000	0.540	20.680	17.825	16.341	14.246	10.979
24AB	Nelson	0.000	0.538	16.867	15.061	15.581	13.879	10.148
24AD	Nelson	0.000	0.538	24.669	22.754	23.469	16.524	15.078
24B	Nelson	0.000	0.540	19.032	19.083	17.123	13.178	11.579
24G	Nelson	0.000	0.539	18.065	20.288	19.671	11.113	9.935
24I	Nelson	0.000	0.538	14.621	13.939	15.003	11.051	9.244
24K	Nelson	0.000	0.537	19.394	20.667	19.325	17.218	15.868
24L	Nelson	0.000	0.538	16.225	16.543	17.045	14.154	9.935
24M	Nelson	0.000	0.539	19.303	21.019	18.908	15.594	14.174
24N	Nelson	0.000	0.538	17.136	15.593	17.809	12.194	12.053
24Q	Nelson	0.000	0.538	20.472	20.009	18.545	17.838	15.046
24U	Nelson	0.000	0.539	20.197	18.491	15.551	10.499	8.852
24V	Nelson	0.000	0.538	16.369	16.373	16.701	14.473	11.764
24W	Nelson	0.000	0.537	22.765	19.492	21.007	21.384	16.103
N1	Nelson	0.000	0.539	20.853	20.363	20.200	16.107	10.520
N2	Nelson	0.496	0.049	17.864	19.488	17.922	10.062	10.558
N3	Nelson	0.000	0.539	20.579	19.252	21.109	18.420	14.605
N4	Nelson	0.561	0.047	20.518	21.346	15.858	18.365	16.418
N5	Nelson	0.000	0.540	13.674	13.027	12.289	9.774	8.267

N6	Nelson	1.235	0.023	19.566	20.127	16.233	14.805	16.039
N7	Nelson	0.000	0.540	18.665	18.628	17.734	12.881	11.640
N8	Nelson	0.000	0.539	17.496	17.648	15.166	14.537	12.681
N9	Nelson	0.000	0.539	17.800	18.077	19.603	14.164	12.160
N10	Nelson	1.047	0.029	18.076	20.589	18.794	13.344	14.391

Supplementary Table 9 Partial mtDNA sequence data for one individual, 17B. First 1153 bp is from the COI sequences, whereas the remaining 1060 bp is from the Cytb genes.

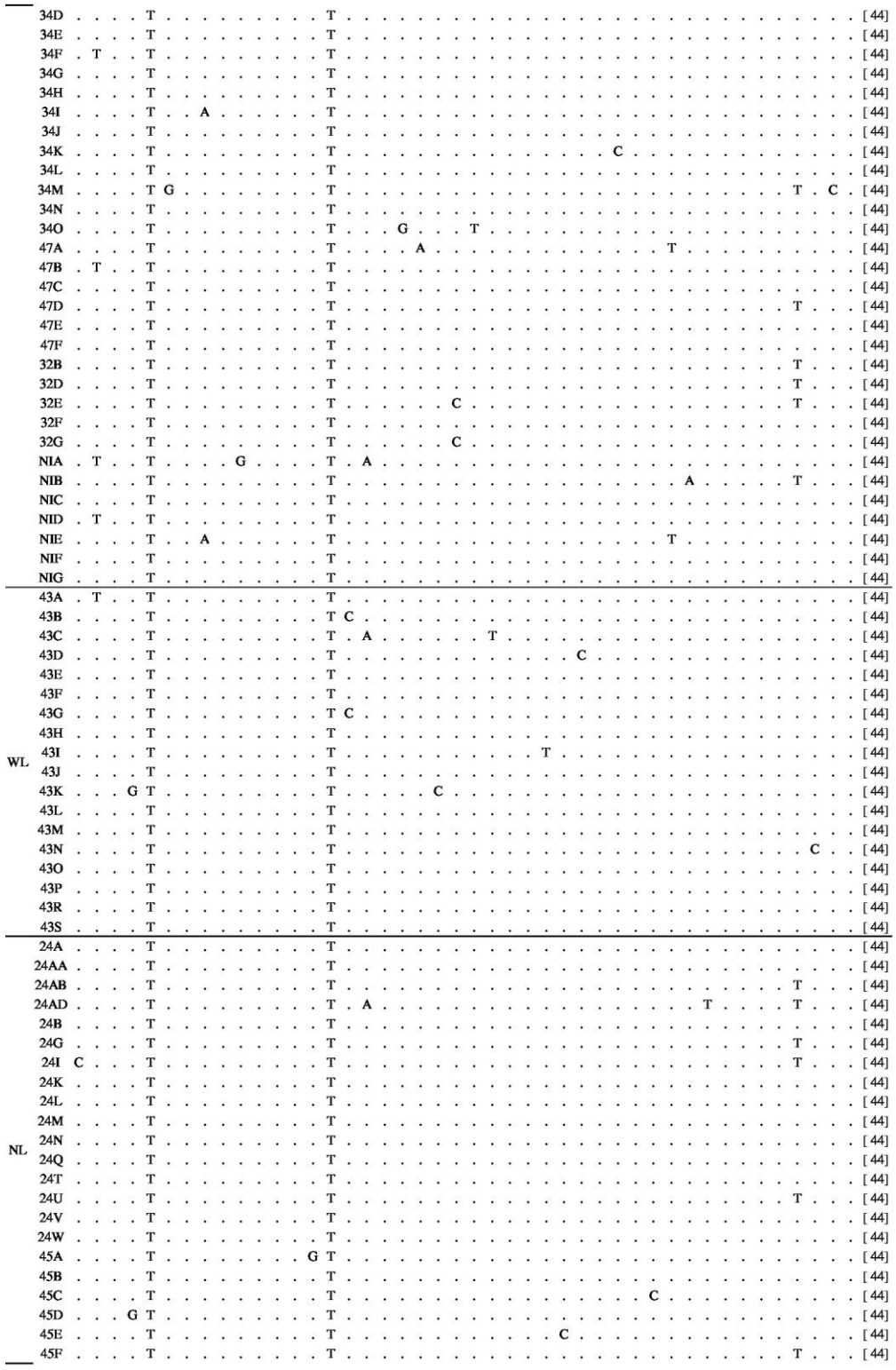
T A G G G G G C T C A G T C T T C T A A T T C G T T T C G A G T T G G G G A C T [40]
T C A G G A G C C T T C T T A G G T G A C G A T C A T T T T T A T A A T G T A A [80]
T C G T A A C T G C T C A T G C T T T C G T T A T G A T T T T T T T T T A T G G T [120]
G A T A C C A C T A A T A A T T G G T G G A T T T G G A A A C T G A A T G G T G [160]
C C T T T A T T A A T T G G T G C A C C A G A C A T G A G T T T T C C C C G A A [200]
T A A A T A A T A T G A G G T T T T G A C T T T T T A C C T C C T T C T T T T A T [240]
T T T G C T T T T T A T G T T C A A C A A T G A T A G A A G G C G G A G C C G G A [280]
A C A G G T T G A A C G G T G T A T C C A C C A T T G T C T G G T C C T A T T G [320]
G T C A T G G G G G T A C C T C G G T T G A T T T A G T G A T T T T T T C T T T [360]
A C A T T T A G C G G G G G C G T C T T C T C T A C T A G G T G C T G T A A A T [400]
T T C A T T A C A A C A A T T T T T A A T A T A C G A T C T C C T G C A A T A A [440]
C T A T G G A G C G T G T T A G T T T A T T T G T T T G A T C T G T T T T T A G T [480]
G A C T G C A T T T T T T A T T G C T T T T T A T C C C T T C C A G T A T T G G C G [520]
G G G G C T A T T A C T A T G T T A C T A A C A G A T C G G A A C T T T A A C A [560]
C T A G T T T C T T T G A T C C G G C A G G G G G A G G T G A T C C T A T T C T [600]
T T A C C A A C A T C T G T T T T G G T T T T T T T G G G C A T C C T G A G G T A [640]
T A T A T T T T A A T T T T T A C C A G G A T T T G G T A T T A T T T C T C A T A [680]
T T T T G A G A A A T T T T T C A T C A A A G C C A G C T T T T G G A A C T T T [720]
A G G A A T G A T T T A T G C T A T G A T T T C A A T T G G G A T T T T A G G T [760]
T T T A T T G T A T G A G C T C A T C A T A T G T T C A C T G T A G G A A T A G [800]
A T G T T G A T A C T C G T G C T T A T T T T A C G G C G G C T A C T A T A G T [840]
A A T T G C C G T T C C A A C A G G T A T T A A G A T T T T T A G A T G A T T A [880]
A T A A C A C T T T A C G G A A G T C G A G G A C C G T T T T C A G C T G C A A [920]
T G T A C T G A G T C T T A G G A T T T A T T T T C T T A T T C A C T T T A G G [960]
A G G G T T A A C A G G G A T T G T T C T T T C A A A T T C T T C T T T A G A T [1000]
A T T G T G T T A C A T G A T A C T T A T T A T G T G G T T G C C C A T T T T C [1040]
A T T A T G T A T T G T C A A T A G G G G C T G T A T T T G C T A T T T T T G G [1080]
G G G T T T C G T T T A T T G A T T C C C T A T A A T A A C A G G T C T T A C A [1120]
T T A C A C G A G C G G T G A G C A A A G G C T C A T T T C T T A T C T C T A T [1160]
T T G A T G A A A T G G A G G A T C A A T C T T A G G A T T A C T T C T A G T T [1200]
C T T C A A A T T C T G A C C G G T T T A T T T T T T A T C T A T A C A C T A C A [1240]
C A G C A G A C A T G G T A A A T T C A T T T G C T T C T G T A A T T C A T A T [1280]
T G T G C G A G A T G C A C C A G G A G G G T G A C T G T T T C G G A A C C T T [1320]
C A T G C T A A T G G T G C T T C T T T G T T C T T T G T T T T T T T G T A T A [1360]
T A C A T A T T G G T C G A G G G T T G T A C T A T C A A A G T T A T A T T A A [1400]
C C A T C C T C A T A C C T G A A T G G T G G G G G T T A C A A T T T T C T T T [1440]
G T G A G T A T A G G A A C T G C G T T T C T T G G G T A T G T T C T T C C T T [1480]
G A G G A C A A A T A T C T T T T T G G G G G G C T A C T G T A A T T A C A A A [1520]
C C T A G T A T C A G C T A T T C C T T A T T G A G G A C C T A G A A T A G T A [1560]
G A G T G A G T T T G A G G A G G A T T T T C T G T T G G G C A A C C A A C T T [1600]
T A A A T C G T T T T T T C T C T T T A C A T T T C A T T T T A C C T T T C T T [1640]
G A T T G G A G G A T T A A G A C T C C T T C A T T T A G T T T T T T T T A C A T [1680]
G A A A A A G G G T C A A G T A A T C C T T T A G G T G A T T T A A A T C A T T [1720]
T A A G T A A A A T T C C G T T T C A T C C T T A T T T T T C T T G A A A A G A [1760]

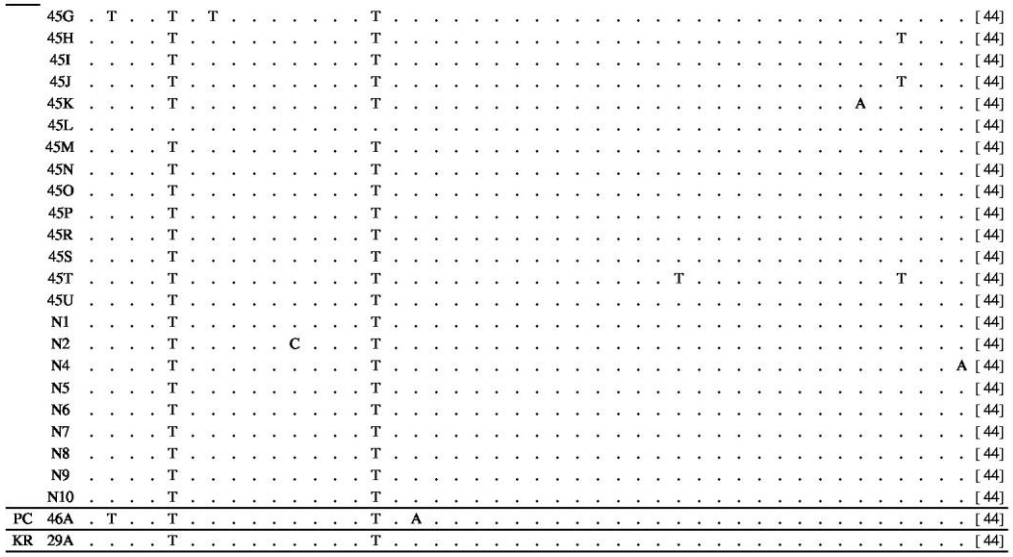
C A T T G T T G G A T T T C T T T T T G T T T T T A T G A T T T T A G C C C T T [1800]
T T A G G A T T T T T T C T T T C C G A C A T T A C T T G G A G A T C C T G A A A [1840]
A C T A T A A T C A T G C T A G A T C T A T A G T A A C A C C T G T A C A C A T [1880]
T C A A C C A G A A T G G T A C T T C T T A T T T G C T T A T G C A A T T C T T [1920]
C G T T C T A T T C C T A G G A A G T T G G G G G G A G T T A T T G C T C T T G [1960]
T A A T A T C A A T T G C A A T C C T T T A T C T T C T T C C T T T T A G G G G [2000]
A G T T G G A A A A A T C A T T C C G T C T T C G T T T G T A C C T C T T T A C [2040]
C A G G T G G T A T T T T G G T C T T T A G T T G T A A C G T T T A T T A T T T [2080]
T A A C T T G G T T A G G G G C T T G T C C T A T T G A A G A G C C C T A T G C [2120]
T A C T T T A G C A G T T C C T A T T T C A A T T T T G T A T T T T C T A C T T [2160]
T A C G T A G T T T T A A T T A G C C T A C C A A G T A T T T G G G G G T T T A [2200]
T T T T A G A A G G T G A [2213]

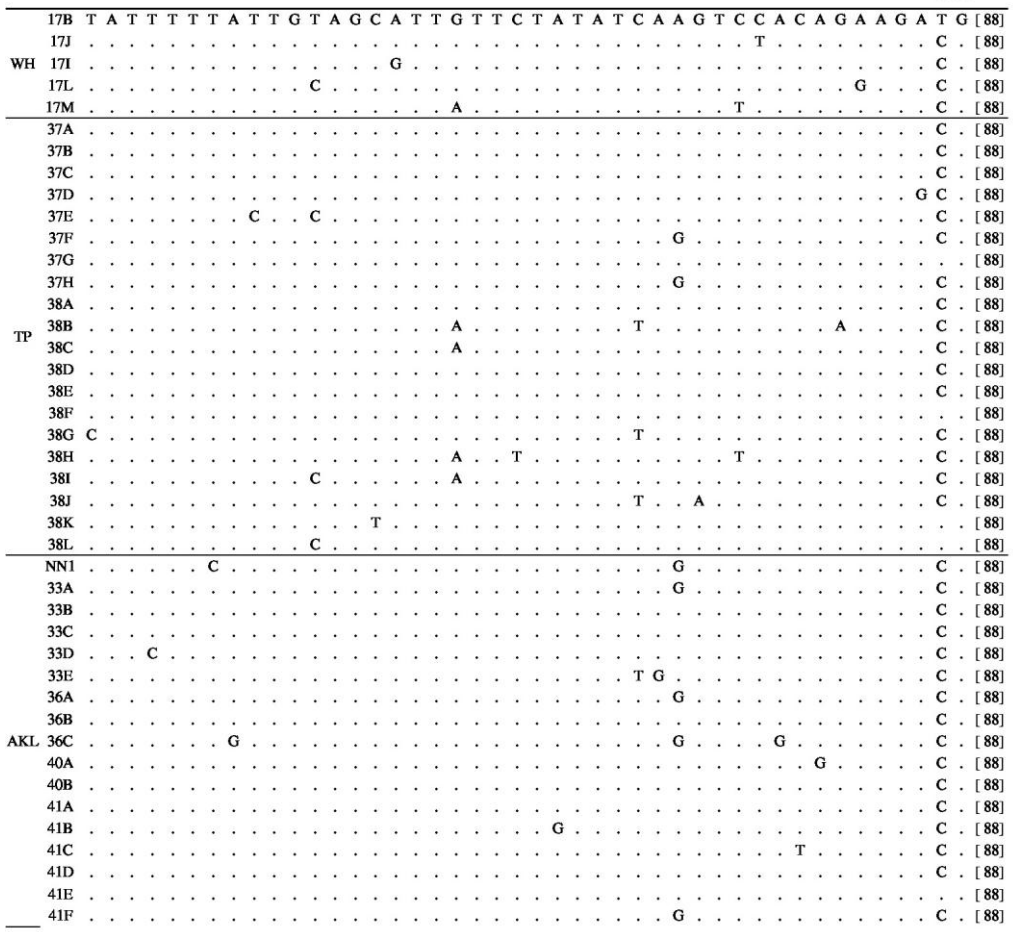
Appendix Two: Supplementary Figures and Tables

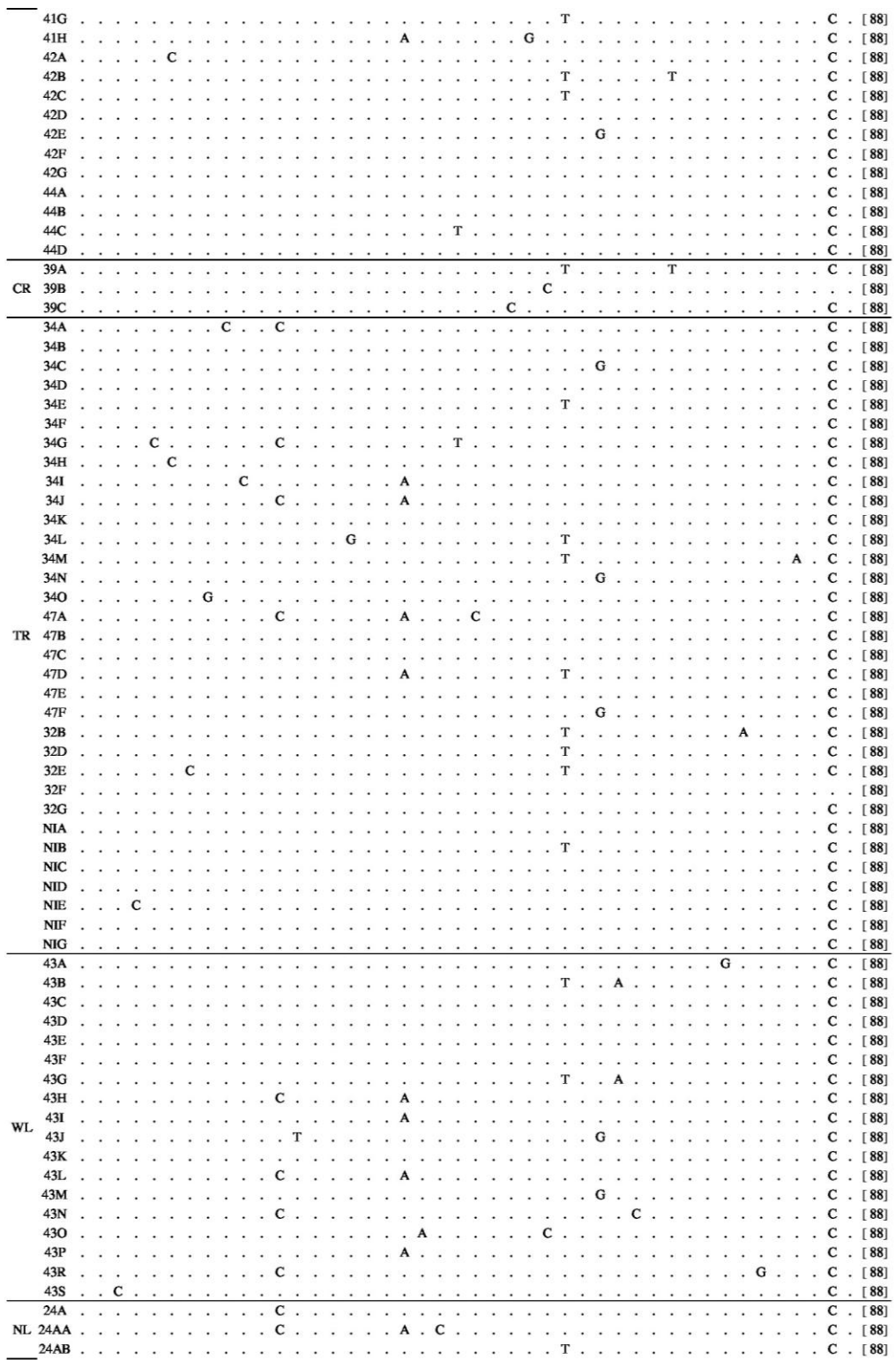
Supplementary Table 10 Partial concatenated mitochondrial COI and Cytb alignment for 170 *P. maculata* haplotypes. The sequences only include variable sites. The first 105 sites belong to the COI gene; the remaining 68 sites belong to the Cytb gene. The dots symbolize identical bases. *The numbers in the first row indicate the location of the mutations in the sequence data in Supplementary Table 9.

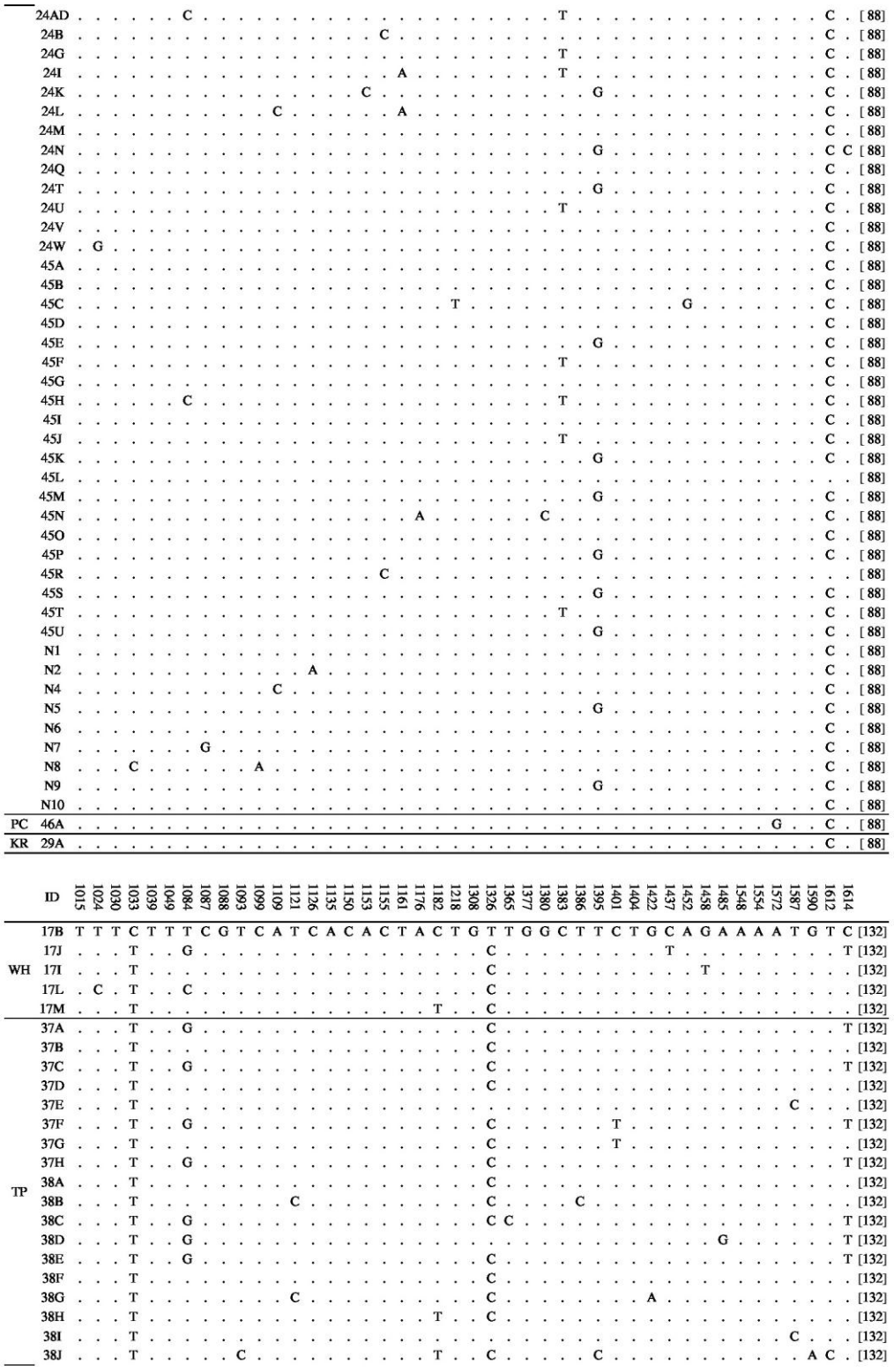
Site	ID	1*	10	17	19	28	46	82	121	127	133	145	184	187	190	196	217	295	298	301	307	310	313	334	337	370	383	386	394	400	403	515	529	536	539	550	553	559	577	583	598	611	616	625	628	
	17B	T	C	C	A	C	A	C	G	A	A	T	C	G	T	C	T	G	T	A	G	T	T	C	G	G	C	C	T	T	C	T	T	T	C	G	C	C	G	G	T	C	T	T	G	[44]
	17J	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
WH	17I	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	17L	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	17M	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37A	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37B	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37C	.	.	.	.	T	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37D	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37E	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37F	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	37H	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
TP	38B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	.	[44]
	38C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38D	.	.	.	.	T	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38E	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38H	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38I	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	[44]
	38J	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	38L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	NN1	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	33A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	33B	.	.	.	.	T	.	.	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	[44]
	33C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	33D	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	33E	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	[44]
	36A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	36B	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	36C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	40A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	40B	.	.	.	G	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41B	.	.	.	G	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41C	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
AKL	41D	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41F	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	41H	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	42A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	42B	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	42C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	[44]
	42D	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	42E	.	.	.	.	T	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	[44]
	42F	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	42G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	44A	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	44B	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	44C	.	.	.	G	T	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	44D	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	39A	.	T	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
CR	39B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	39C	.	.	.	.	T	.	.	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	[44]
	34A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
TR	34B	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]
	34C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	[44]

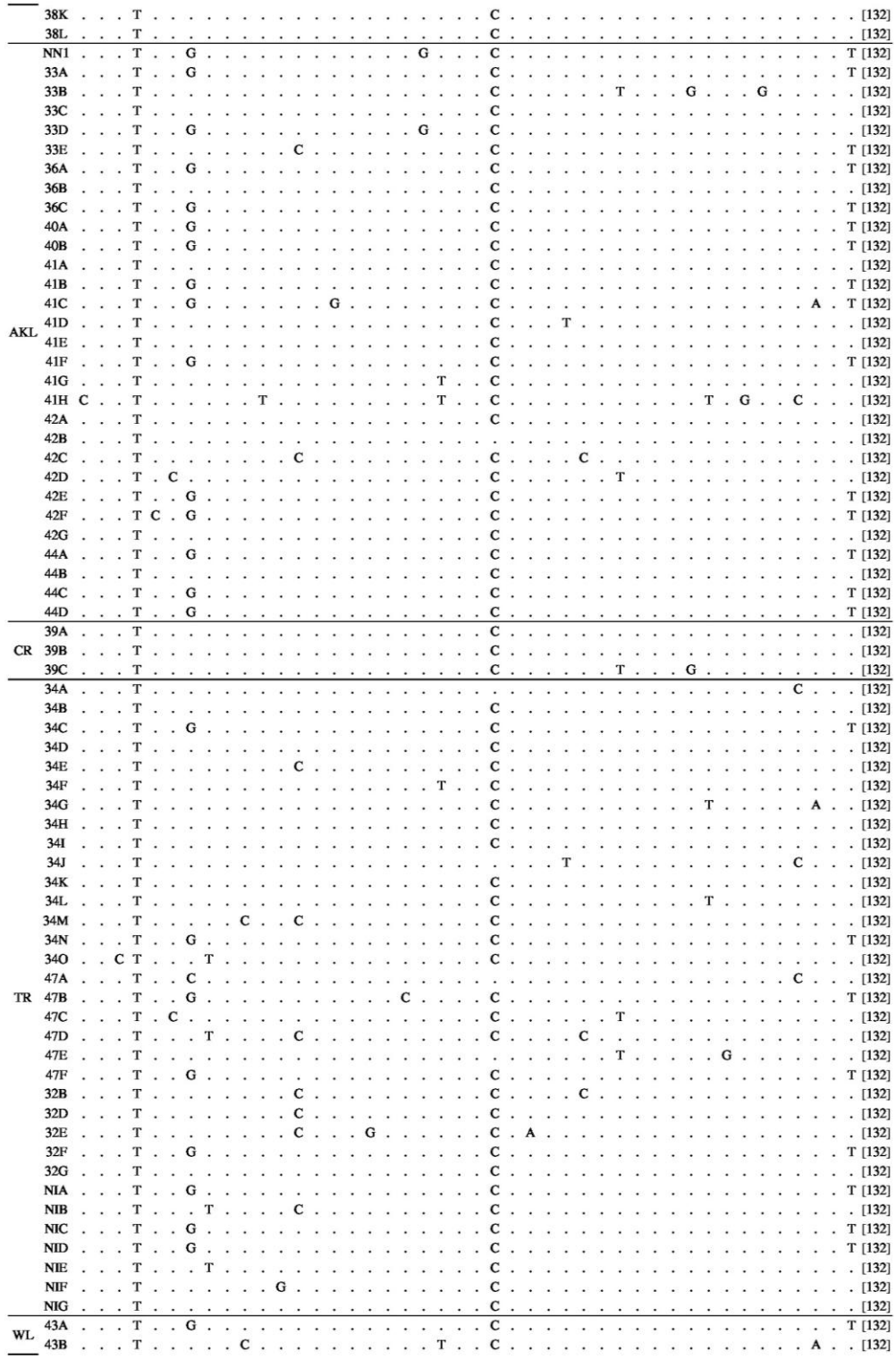

34D [44]
34E [44]
34F [44]
34G [44]
34H [44]
34I [44]
34J [44]
34K [44]
34L [44]
34M [44]
34N [44]
34O [44]
47A [44]
47B [44]
47C [44]
47D [44]
47E [44]
47F [44]
32B [44]
32D [44]
32E [44]
32F [44]
32G [44]
NIA [44]
NIB [44]
NIC [44]
NID [44]
NIE [44]
NIF [44]
NIG [44]
43A [44]
43B [44]
43C [44]
43D [44]
43E [44]
43F [44]
43G [44]
43H [44]
WL
43I [44]
43J [44]
43K [44]
43L [44]
43M [44]
43N [44]
43O [44]
43P [44]
43R [44]
43S [44]
24A [44]
24AA [44]
24AB [44]
24AD [44]
24B [44]
24G [44]
24I [44]
24K [44]
24L [44]
24M [44]
NL
24N [44]
24Q [44]
24T [44]
24U [44]
24V [44]
24W [44]
45A [44]
45B [44]
45C [44]
45D [44]
45E [44]
45F [44]

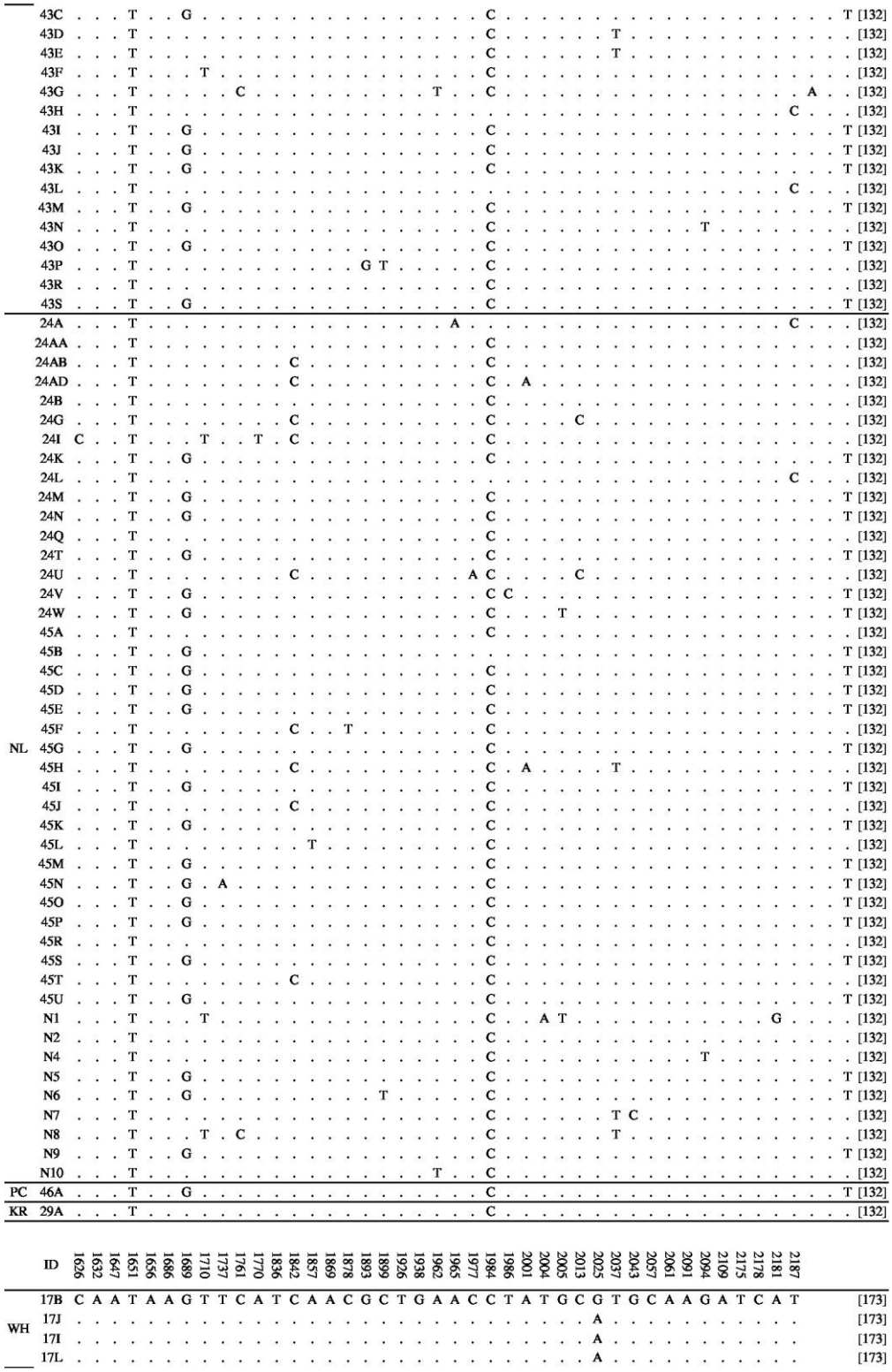

45G . T . . T . T . . . . . T . . . . . [44]
45H . . . . T . . . . . T . . . . . T . . . [44]
45I . . . . T . . . . . T . . . . . [44]
45J . . . . T . . . . . T . . . . . T . . . [44]
45K . . . . T . . . . . T . . . . . A . . . . . [44]
45L . . . . . [44]
45M . . . . T . . . . . T . . . . . [44]
45N . . . . T . . . . . T . . . . . [44]
45O . . . . T . . . . . T . . . . . [44]
45P . . . . T . . . . . T . . . . . [44]
45R . . . . T . . . . . T . . . . . [44]
45S . . . . T . . . . . T . . . . . [44]
45T . . . . T . . . . . T . . . . . T . . . . . T . . . [44]
45U . . . . T . . . . . T . . . . . [44]
N1 . . . . T . . . . . T . . . . . [44]
N2 . . . . T . . . . . C . . . T . . . . . [44]
N4 . . . . T . . . . . T . . . . . A [44]
N5 . . . . T . . . . . T . . . . . [44]
N6 . . . . T . . . . . T . . . . . [44]
N7 . . . . T . . . . . T . . . . . [44]
N8 . . . . T . . . . . T . . . . . [44]
N9 . . . . T . . . . . T . . . . . [44]
N10 . . . . T . . . . . T . . . . . [44]
PC 46A . T . . T . . . . . T . A . . . . . [44]
KR 29A . . . . T . . . . . T . . . . . [44]

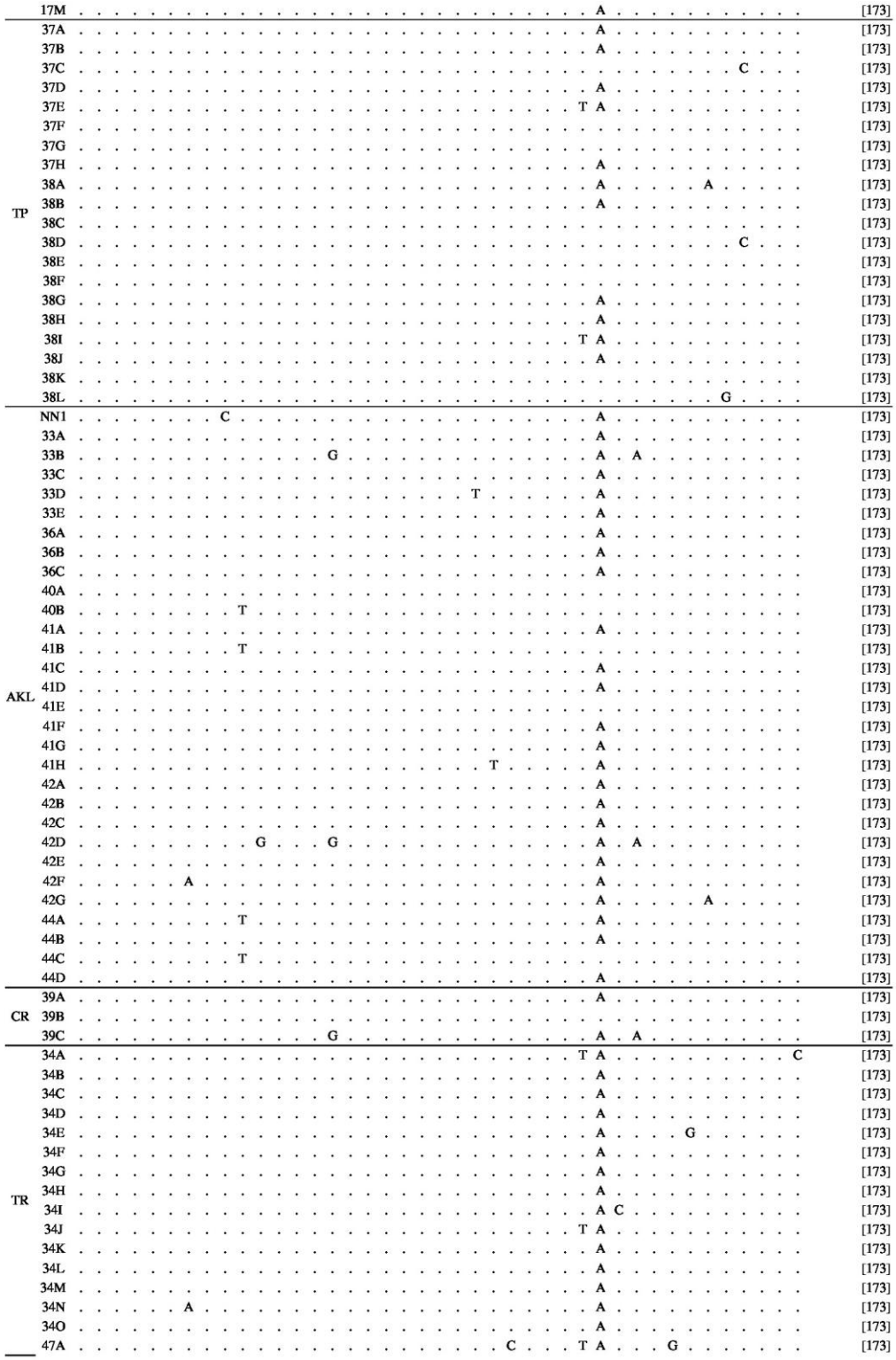

ID 631 640 643 679 691 694 712 715 719 742 751 763 769 784 787 793 805 817 829 832 835 847 850 853 862 874 878 892 901 904 907 916 925 931 937 952 961 964 967 970 973 997 1000 1006
17B T A T T T T T A T T G T A G C A T T T G T T C T A T A T C A A A G T C C A C A G A A G A T G [88]
17J . . . . . T . . . . . C . [88]
WH 17I . . . . . G . . . . . C . [88]
17L . . . . . C . . . . . G . . . C . [88]
17M . . . . . A . . . . . T . . . . . C . [88]
37A . . . . . C . [88]
37B . . . . . C . [88]
37C . . . . . C . [88]
37D . . . . . G C . [88]
37E . . . . . C . . C . . . . . C . [88]
37F . . . . . G . . . . . C . [88]
37G . . . . . [88]
37H . . . . . G . . . . . C . [88]
38A . . . . . C . [88]
TP 38B . . . . . A . . . . . T . . . . . A . . . . C . [88]
38C . . . . . A . . . . . C . [88]
38D . . . . . C . [88]
38E . . . . . C . [88]
38F . . . . . [88]
38G C . . . . . T . . . . . C . [88]
38H . . . . . A . . T . . . . . T . . . . . C . [88]
38I . . . . . C . . . . . A . . . . . C . [88]
38J . . . . . T . . A . . . . . C . [88]
38K . . . . . T . . . . . [88]
38L . . . . . C . . . . . [88]
NN1 . . . . . C . . . . . G . . . . . C . [88]
33A . . . . . G . . . . . C . [88]
33B . . . . . C . [88]
33C . . . . . C . [88]
33D . . . C . . . . . C . [88]
33E . . . . . T G . . . . . C . [88]
36A . . . . . G . . . . . C . [88]
36B . . . . . C . [88]
AKL 36C . . . . . G . . . . . G . . . . G . . . . . C . [88]
40A . . . . . G . . . . . C . [88]
40B . . . . . C . [88]
41A . . . . . C . [88]
41B . . . . . G . . . . . C . [88]
41C . . . . . T . . . . . C . [88]
41D . . . . . C . [88]
41E . . . . . [88]
41F . . . . . G . . . . . C . [88]

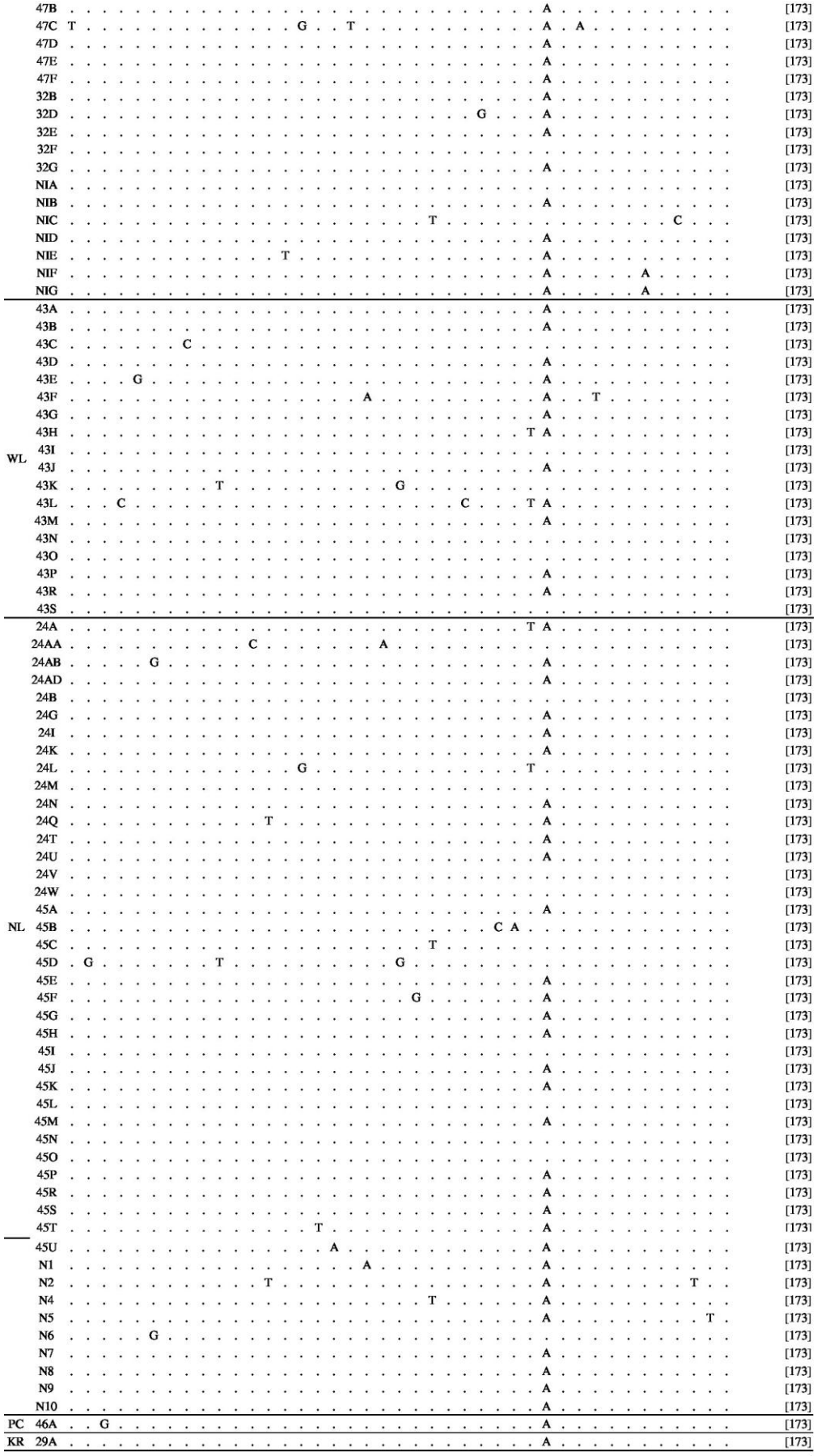

41G
41H
42A
42B
42C
42D
42E
42F
42G
44A
44B
44C
44D
39A
CR 39B
39C
34A
34B
34C
34D
34E
34F
34G
34H
34I
34J
34K
34L
34M
34N
34O
47A
TR 47B
47C
47D
47E
47F
32B
32D
32E
32F
32G
NIA
NIB
NIC
NID
NIE
NIF
NIG
43A
43B
43C
43D
43E
43F
43G
43H
WL 43I
43J
43K
43L
43M
43N
43O
43P
43R
43S
24A
NL 24AA
24AB


ID 1015 1024 1030 1033 1039 1049 1084 1087 1088 1093 1099 1109 1121 1126 1135 1150 1153 1155 1161 1176 1182 1218 1308 1326 1365 1377 1380 1383 1386 1395 1401 1404 1422 1437 1452 1458 1485 1548 1554 1572 1587 1590 1612 1614
24AD [88]
24B [88]
24G [88]
24I [88]
24K [88]
24L [88]
24M [88]
24N [88]
24Q [88]
24T [88]
24U [88]
24V [88]
24W [88]
45A [88]
45B [88]
45C [88]
45D [88]
45E [88]
45F [88]
45G [88]
45H [88]
45I [88]
45J [88]
45K [88]
45L [88]
45M [88]
45N [88]
45O [88]
45P [88]
45R [88]
45S [88]
45T [88]
45U [88]
N1 [88]
N2 [88]
N4 [88]
N5 [88]
N6 [88]
N7 [88]
N8 [88]
N9 [88]
N10 [88]
PC 46A [88]
KR 29A [88]
17B T T T C T T T C G T C A T C A C A C T A C T G T T G G C T T C T G C A G A A A A T G T C [132]
17J [132]
WH 17I [132]
17L [132]
17M [132]
37A [132]
37B [132]
37C [132]
37D [132]
37E [132]
37F [132]
37G [132]
37H [132]
TP 38A [132]
38B [132]
38C [132]
38D [132]
38E [132]
38F [132]
38G [132]
38H [132]
38I [132]
38J [132]


38K
38L
NN1
33A
33B
33C
33D
33E
36A
36B
36C
40A
40B
41A
41B
41C
41D
41E
41F
41G
41H
42A
42B
42C
42D
42E
42F
42G
44A
44B
44C
44D
39A
39B
39C
34A
34B
34C
34D
34E
34F
34G
34H
34I
34J
34K
34L
34M
34N
34O
47A
47B
47C
47D
47E
47F
32B
32D
32E
32F
32G
NIA
NIB
NIC
NID
NIE
NIF
NIG
43A
43B
AKL
CR
TR
WL
[132]


ID 1626 1632 1647 1651 1656 1686 1689 1710 1737 1761 1770 1836 1842 1857 1869 1878 1893 1899 1926 1938 1962 1965 1977 1984 1986 2001 2004 2005 2013 2025 2037 2043 2057 2061 2091 2094 2109 2175 2178 2181 2187
17B C A A T A A G T T C A T C A A C G C T G A A C C C T A T G C G T G C A A G A T C A T [173]
17J . . . . . A . . . . . A . . . . . [173]
17I . . . . . A . . . . . [173]
17L . . . . . A . . . . . [173]
43C . . . T . . G . . . . . C . . . . . T [132]
43D . . . T . . . . . C . . . . . T . . . . . [132]
43E . . . T . . . . . C . . . . . T . . . . . [132]
43F . . . T . . . T . . . . . C . . . . . [132]
43G . . . T . . . . . C . . . . . T . . C . . . . . A . . [132]
43H . . . T . . . . . C . . . . [132]
43I . . . T . . G . . . . . C . . . . . T [132]
43J . . . T . . G . . . . . C . . . . . T [132]
43K . . . T . . G . . . . . C . . . . . T [132]
43L . . . T . . . . . C . . . . [132]
43M . . . T . . G . . . . . C . . . . . T [132]
43N . . . T . . . . . C . . . . . T . . . . . [132]
43O . . . T . . G . . . . . C . . . . . T [132]
43P . . . T . . . . . G T . . . . C . . . . . [132]
43R . . . T . . . . . C . . . . . [132]
43S . . . T . . G . . . . . C . . . . . T [132]
24A . . . T . . . . . A . . . . . C . . . [132]
24AA . . . T . . . . . C . . . . . [132]
24AB . . . T . . . . . C . . . . . C . . . . . [132]
24AD . . . T . . . . . C . . . . . C . A . . . . . [132]
24B . . . T . . . . . C . . . . . [132]
24G . . . T . . . . . C . . . . . C . . . . . C . . . . . [132]
24I C . . T . . . T . . T . C . . . . . C . . . . . [132]
24K . . . T . . G . . . . . C . . . . . T [132]
24L . . . T . . . . . C . . . . [132]
24M . . . T . . G . . . . . C . . . . . T [132]
24N . . . T . . G . . . . . C . . . . . T [132]
24Q . . . T . . . . . C . . . . . [132]
24T . . . T . . G . . . . . C . . . . . T [132]
24U . . . T . . . . . C . . . . . A C . . . . C . . . . . [132]
24V . . . T . . G . . . . . C C . . . . . T [132]
24W . . . T . . G . . . . . C . . . T . . . . . T [132]
45A . . . T . . . . . C . . . . . [132]
45B . . . T . . G . . . . . T [132]
45C . . . T . . G . . . . . C . . . . . T [132]
45D . . . T . . G . . . . . C . . . . . T [132]
45E . . . T . . G . . . . . C . . . . . T [132]
45F . . . T . . . . . C . . T . . . . . C . . . . . [132]
NL 45G . . . T . . G . . . . . C . . . . . T [132]
45H . . . T . . . . . C . . . . . C . A . . . . T . . . . . [132]
45I . . . T . . G . . . . . C . . . . . T [132]
45J . . . T . . . . . C . . . . . C . . . . . [132]
45K . . . T . . G . . . . . C . . . . . T [132]
45L . . . T . . . . . T . . . . . C . . . . . [132]
45M . . . T . . G . . . . . C . . . . . T [132]
45N . . . T . . G . A . . . . . C . . . . . T [132]
45O . . . T . . G . . . . . C . . . . . T [132]
45P . . . T . . G . . . . . C . . . . . T [132]
45R . . . T . . . . . C . . . . . [132]
45S . . . T . . G . . . . . C . . . . . T [132]
45T . . . T . . . . . C . . . . . C . . . . . [132]
45U . . . T . . G . . . . . C . . . . . T [132]
N1 . . . T . . . T . . . . . C . . A T . . . . . G . . . [132]
N2 . . . T . . . . . C . . . . . [132]
N4 . . . T . . . . . C . . . . . T . . . . . [132]
N5 . . . T . . G . . . . . C . . . . . T [132]
N6 . . . T . . G . . . . . T . . . . . C . . . . . T [132]
N7 . . . T . . . . . C . . . . . T C . . . . . [132]
N8 . . . T . . . T . C . . . . . C . . . . . T . . . . . [132]
N9 . . . T . . G . . . . . C . . . . . T [132]
N10 . . . T . . . . . T . . C . . . . . [132]
PC 46A . . . T . . G . . . . . C . . . . . T [132]
KR 29A . . . T . . . . . C . . . . . [132]
WH
—


17M [173]
37A [173]
37B [173]
37C [173]
37D [173]
37E [173]
37F [173]
37G [173]
37H [173]
38A [173]
38B [173]
38C [173]
38D [173]
38E [173]
38F [173]
38G [173]
38H [173]
38I [173]
38J [173]
38K [173]
38L [173]
TP
NN1 [173]
33A [173]
33B [173]
33C [173]
33D [173]
33E [173]
36A [173]
36B [173]
36C [173]
40A [173]
40B [173]
41A [173]
41B [173]
41C [173]
41D [173]
41E [173]
41F [173]
41G [173]
41H [173]
42A [173]
42B [173]
42C [173]
42D [173]
42E [173]
42F [173]
42G [173]
44A [173]
44B [173]
44C [173]
44D [173]
AKL
39A [173]
39B [173]
39C [173]
CR
34A [173]
34B [173]
34C [173]
34D [173]
34E [173]
34F [173]
34G [173]
34H [173]
34I [173]
34J [173]
34K [173]
34L [173]
34M [173]
34N [173]
34O [173]
TR
47A [173]


47B [173]
47C [173]
47D [173]
47E [173]
47F [173]
32B [173]
32D [173]
32E [173]
32F [173]
32G [173]
NIA [173]
NIB [173]
NIC [173]
NID [173]
NIE [173]
NIF [173]
NIG [173]
43A [173]
43B [173]
43C [173]
43D [173]
43E [173]
43F [173]
43G [173]
43H [173]
WL 43I [173]
43J [173]
43K [173]
43L [173]
43M [173]
43N [173]
43O [173]
43P [173]
43R [173]
43S [173]
24A [173]
24AA [173]
24AB [173]
24AD [173]
24B [173]
24G [173]
24I [173]
24K [173]
24L [173]
24M [173]
24N [173]
24Q [173]
24T [173]
24U [173]
24V [173]
24W [173]
45A [173]
NL 45B [173]
45C [173]
45D [173]
45E [173]
45F [173]
45G [173]
45H [173]
45I [173]
45J [173]
45K [173]
45L [173]
45M [173]
45N [173]
45O [173]
45P [173]
45R [173]
45S [173]
45T [173]
45U [173]
N1 [173]
N2 [173]
N4 [173]
N5 [173]
N6 [173]
N7 [173]
N8 [173]
N9 [173]
N10 [173]
PC 46A [173]
KR 29A [173]

Supplementary Table 11 Haplotype frequencies for the mtDNA genes by mere counting in populations. The IDs are individual names. [+]Population codes. *Sample size.

COI							Cytb							CONCATENATED						
	[+]WH	TP	AK	TA	WL	NL		WH	TP	AK	TA	WL	NL		WH	TP	AK	TA	WL	NL
ID	5*	20	30	33	18	45	ID	5	20	30	33	18	45	ID	5	20	30	33	18	45
17B	1	0	0	0	0	0	17B	1	0	0	0	0	0	17B	1	0	0	0	0	0
17J	1	0	0	0	0	0	17I	1	0	0	1	0	0	17J	1	0	0	0	0	0
17I	1	0	0	0	0	0	17J	1	0	0	0	0	0	17I	1	0	0	0	0	0
17L	1	0	0	0	0	0	17L	1	2	5	8	1	4	17L	1	0	0	0	0	0
17M	1	0	0	0	0	0	17M	1	1	1	1	0	1	17M	1	0	0	0	0	0
37A	0	1	2	2	0	0	37A	0	2	7	3	3	10	37A	0	1	1	1	0	0
37B	0	1	1	1	0	0	37C	0	1	0	0	0	0	37B	0	1	1	1	0	0
37C	0	2	0	0	0	0	37E	0	2	0	0	1	0	37C	0	1	0	0	0	0
37D	0	1	0	0	0	0	37F	0	1	0	0	0	0	37D	0	1	0	0	0	0
37E	0	1	0	0	0	0	37G	0	1	0	0	0	0	37E	0	1	0	0	0	0
37F	0	2	3	3	1	7	38A	0	1	1	2	0	0	37F	0	1	0	0	0	0
37G	0	2	1	0	0	0	38B	0	1	1	2	0	1	37G	0	1	0	0	0	0
38A	0	1	2	4	1	2	24V	0	1	0	0	0	1	37H	0	1	3	2	1	5
38B	0	1	0	0	0	0	38D	0	1	0	0	0	0	38A	0	1	1	1	0	0
38C	0	1	0	0	0	0	38E	0	1	1	2	3	4	38B	0	1	0	0	0	0
24M	0	1	0	1	0	6	38F	0	2	1	0	0	2	38C	0	1	0	0	0	0
38G	0	1	0	0	0	0	38G	0	1	0	0	0	0	38D	0	1	0	0	0	0
38H	0	1	0	0	0	0	38J	0	1	0	0	0	0	38E	0	1	0	0	0	3
38I	0	1	0	0	0	0	38L	0	1	0	0	0	0	38F	0	1	1	0	0	0
38J	0	1	0	0	2	0	NN1	0	0	1	0	0	0	38G	0	1	0	0	0	0
38K	0	1	0	0	0	0	33B	0	0	1	0	0	0	38H	0	1	0	0	0	0
38L	0	1	0	0	0	0	33D	0	0	1	0	0	0	38I	0	1	0	0	0	0
NN1	0	0	1	0	0	0	40B	0	0	3	0	0	0	38J	0	1	0	0	0	0
33B	0	0	1	0	0	0	41C	0	0	1	0	0	0	38K	0	1	0	0	0	0
33C	0	0	1	0	0	0	41D	0	0	1	0	0	0	38L	0	1	0	0	0	0
33D	0	0	1	0	0	0	41H	0	0	1	0	0	0	NN1	0	0	1	0	0	0
33E	0	0	1	0	0	0	42B	0	0	1	0	0	0	33B	0	0	1	0	0	0
36B	0	0	1	0	0	0	42D	0	0	1	0	0	0	33C	0	0	1	0	0	0
36C	0	0	1	0	0	0	42F	0	0	1	1	0	0	33D	0	0	1	0	0	0
40A	0	0	1	0	0	0	44A	0	0	1	0	0	0	33E	0	0	1	0	0	0
45D	0	0	1	0	0	1	34A	0	0	0	1	0	0	36B	0	0	1	0	0	0
41B	0	0	1	0	0	0	34E	0	0	0	1	0	0	36C	0	0	1	0	0	0
41C	0	0	1	0	0	0	34G	0	0	0	1	0	0	40A	0	0	1	0	0	0
41D	0	0	1	0	0	0	34I	0	0	0	1	0	0	40B	0	0	1	0	0	0
41G	0	0	1	0	0	0	34J	0	0	0	1	0	0	41A	0	0	1	2	0	0
41H	0	0	1	0	0	0	47A	0	0	0	1	0	0	41B	0	0	1	0	0	0
42A	0	0	1	0	0	0	47B	0	0	0	1	0	0	41C	0	0	1	0	0	0
42B	0	0	1	0	0	0	47C	0	0	0	1	0	0	41D	0	0	1	0	0	0
42C	0	0	1	1	0	4	47E	0	0	0	1	0	0	41G	0	0	1	0	0	0
42D	0	0	1	0	0	0	32D	0	0	0	1	0	0	41H	0	0	1	0	0	0
42E	0	0	1	0	0	0	32E	0	0	0	1	0	1	42A	0	0	1	0	0	0

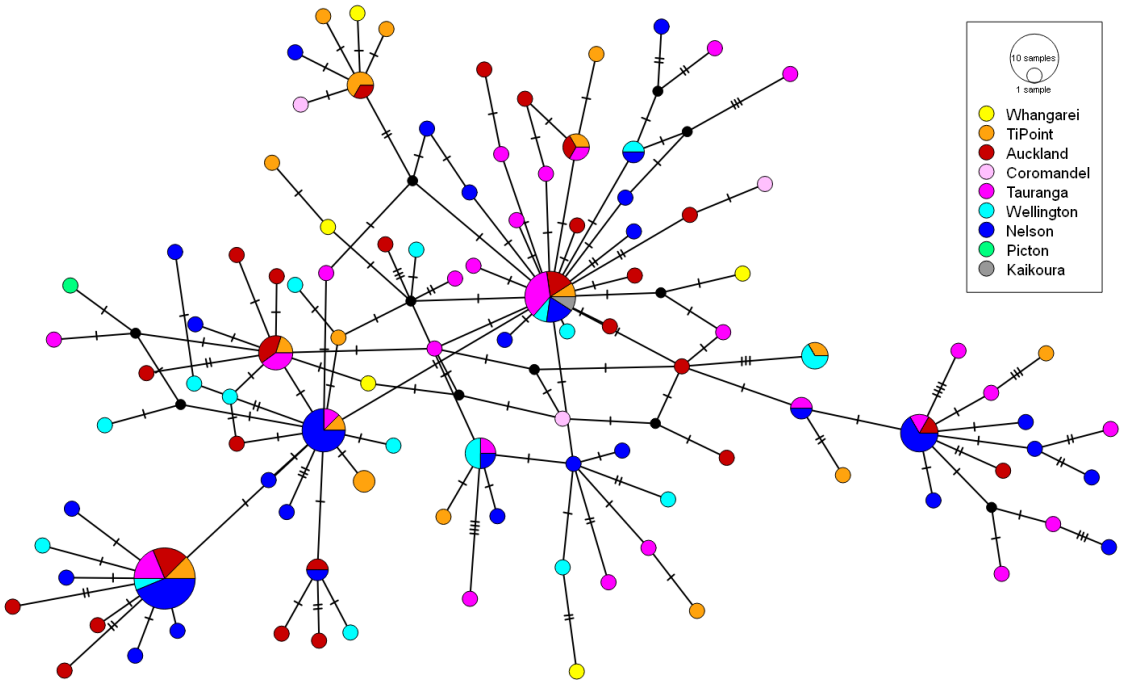
42F	0	0	1	0	0	0
44C	0	0	1	0	0	0
34A	0	0	0	1	0	0
34E	0	0	0	1	0	0
34F	0	0	0	1	0	0
34G	0	0	0	1	0	0
34H	0	0	0	1	0	0
34I	0	0	0	1	0	0
34J	0	0	0	1	2	1
34K	0	0	0	1	0	0
34L	0	0	0	1	0	0
34M	0	0	0	1	0	0
34O	0	0	0	1	0	0
47A	0	0	0	1	0	0
47C	0	0	0	1	0	0
47D	0	0	0	1	0	0
32B	0	0	0	1	0	0
32E	0	0	0	1	0	0
32F	0	0	0	1	0	0
NI-A	0	0	0	1	0	0
NI-B	0	0	0	1	0	0
NI-E	0	0	0	1	0	0
NI-F	0	0	0	1	0	0
43A	0	0	0	0	1	0
43C	0	0	0	0	1	0
43D	0	0	0	0	1	0
43F	0	0	0	0	1	1
43I	0	0	0	0	1	0
43J	0	0	0	0	1	0
43K	0	0	0	0	1	0
43N	0	0	0	0	1	0
43O	0	0	0	0	1	0
43P	0	0	0	0	1	0
43R	0	0	0	0	1	0
43S	0	0	0	0	1	0
24A	0	0	0	0	0	1
24AA	0	0	0	0	0	1
24AD	0	0	0	0	0	1
24B	0	0	0	0	0	1
24I	0	0	0	0	0	1
24K	0	0	0	0	0	1
24N	0	0	0	0	0	1
24W	0	0	0	0	0	1
45A	0	0	0	0	0	1
45C	0	0	0	0	0	1
45E	0	0	0	0	0	1

NI-C	0	0	0	1	0	0
NI-E	0	0	0	1	0	0
43B	0	0	0	0	2	0
43C	0	0	0	0	1	0
43D	0	0	0	0	1	1
43E	0	0	0	0	1	0
43F	0	0	0	0	1	0
43K	0	0	0	0	1	0
43L	0	0	0	0	1	0
43N	0	0	0	0	1	0
43P	0	0	0	0	1	0
24A	0	0	0	0	0	1
24AA	0	0	0	0	0	1
24AB	0	0	0	0	0	1
24L	0	0	0	0	0	1
24Q	0	0	0	0	0	1
24U	0	0	0	0	0	1
24W	0	0	0	0	0	1
45B	0	0	0	0	0	1
45C	0	0	0	0	0	1
45D	0	0	0	0	0	1
45F	0	0	0	0	0	1
45H	0	0	0	0	0	1
45T	0	0	0	0	0	1
45U	0	0	0	0	0	1
N1	0	0	0	0	0	1
N2	0	0	0	0	0	1
N4	0	0	0	0	0	1
N5	0	0	0	0	0	1
N6	0	0	0	0	0	1
N7	0	0	0	0	0	1

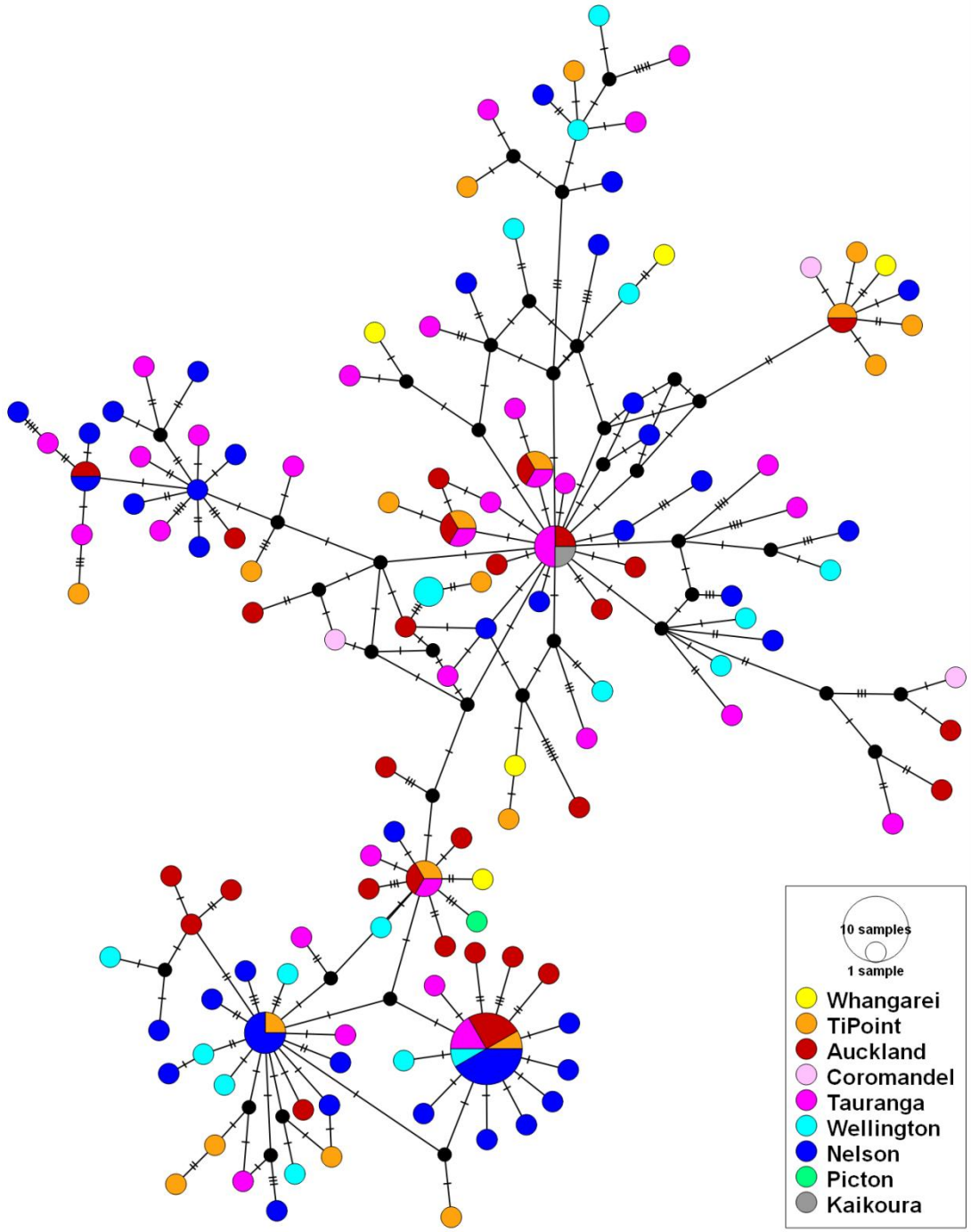
42B	0	0	1	0	0	0
42C	0	0	1	0	0	1
42D	0	0	1	0	0	0
42E	0	0	1	0	0	0
42F	0	0	1	0	0	0
44A	0	0	1	0	0	0
44C	0	0	1	0	0	0
34A	0	0	0	1	0	0
34E	0	0	0	1	0	0
34F	0	0	0	1	0	0
34G	0	0	0	1	0	0
34H	0	0	0	1	0	0
34I	0	0	0	1	0	0
34J	0	0	0	1	0	0
34K	0	0	0	1	0	0
34L	0	0	0	1	0	0
34M	0	0	0	1	0	0
34N	0	0	0	1	0	0
34O	0	0	0	1	0	0
47A	0	0	0	1	0	0
47B	0	0	0	1	0	0
47C	0	0	0	1	0	0
47D	0	0	0	1	0	0
47E	0	0	0	1	0	0
32B	0	0	0	1	0	0
32D	0	0	0	1	0	0
32E	0	0	0	1	0	0
32F	0	0	0	1	0	0
NI-A	0	0	0	1	0	0
NI-B	0	0	0	1	0	0
NI-C	0	0	0	1	0	0
NI-E	0	0	0	1	0	0
NI-F	0	0	0	1	0	0
43A	0	0	0	0	1	0
43B	0	0	0	0	2	0
43C	0	0	0	0	1	0
43D	0	0	0	0	1	0
43E	0	0	0	0	1	0
43F	0	0	0	0	1	0
43H	0	0	0	0	1	0
43I	0	0	0	0	1	0
43J	0	0	0	0	1	0
43K	0	0	0	0	1	0
43L	0	0	0	0	1	0
43N	0	0	0	0	1	0
43O	0	0	0	0	1	0

45F	0	0	0	0	0	1
45G	0	0	0	0	0	1
45H	0	0	0	0	0	1
45K	0	0	0	0	0	1
45L	0	0	0	0	0	1
45N	0	0	0	0	0	1
45R	0	0	0	0	0	1
45T	0	0	0	0	0	1
N2	0	0	0	0	0	1
N4	0	0	0	0	0	1
N7	0	0	0	0	0	1
N8	0	0	0	0	0	1

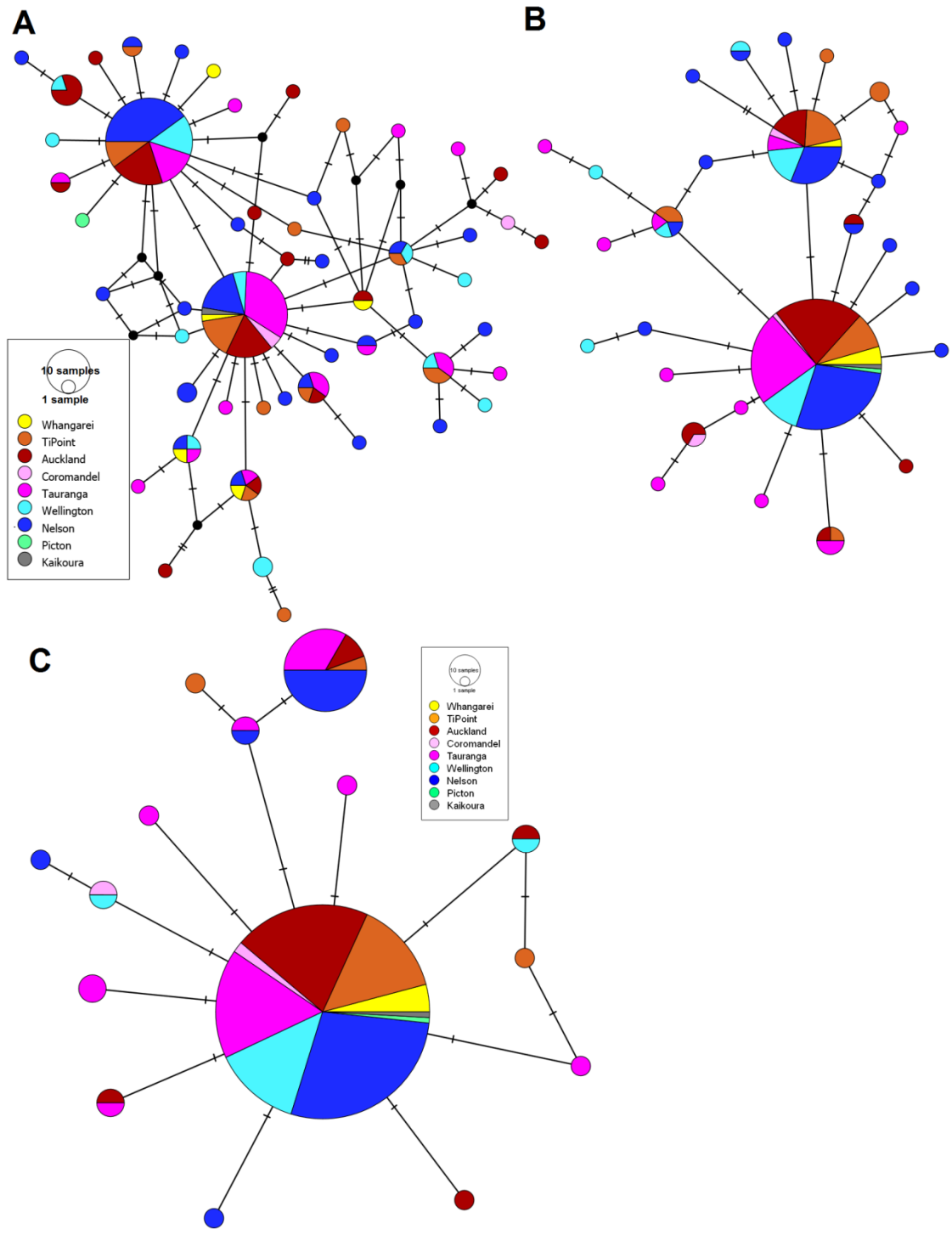
43P	0	0	0	0	1	0
43R	0	0	0	0	1	0
43S	0	0	0	0	1	0
24A	0	0	0	0	0	1
24AA	0	0	0	0	0	1
24AB	0	0	0	0	0	1
24AD	0	0	0	0	0	1
24B	0	0	0	0	0	1
24I	0	0	0	0	0	1
24K	0	0	0	0	0	1
24L	0	0	0	0	0	1
24N	0	0	0	0	0	1
24Q	0	0	0	0	0	1
24U	0	0	0	0	0	1
24V	0	0	0	0	0	1
24W	0	0	0	0	0	1
45A	0	0	0	0	0	1
45B	0	0	0	0	0	1
45C	0	0	0	0	0	1
45D	0	0	0	0	0	1
45E	0	0	0	0	0	1
45F	0	0	0	0	0	1
45G	0	0	0	0	0	1
45H	0	0	0	0	0	1
45J	0	0	0	0	0	1
45K	0	0	0	0	0	1
45L	0	0	0	0	0	1
45N	0	0	0	0	0	1
45R	0	0	0	0	0	1
45T	0	0	0	0	0	1
45U	0	0	0	0	0	1
N1	0	0	0	0	0	1
N2	0	0	0	0	0	1
N4	0	0	0	0	0	1
N5	0	0	0	0	0	1
N6	0	0	0	0	0	1
N7	0	0	0	0	0	1
N8	0	0	0	0	0	1
N10	0	0	0	0	0	1



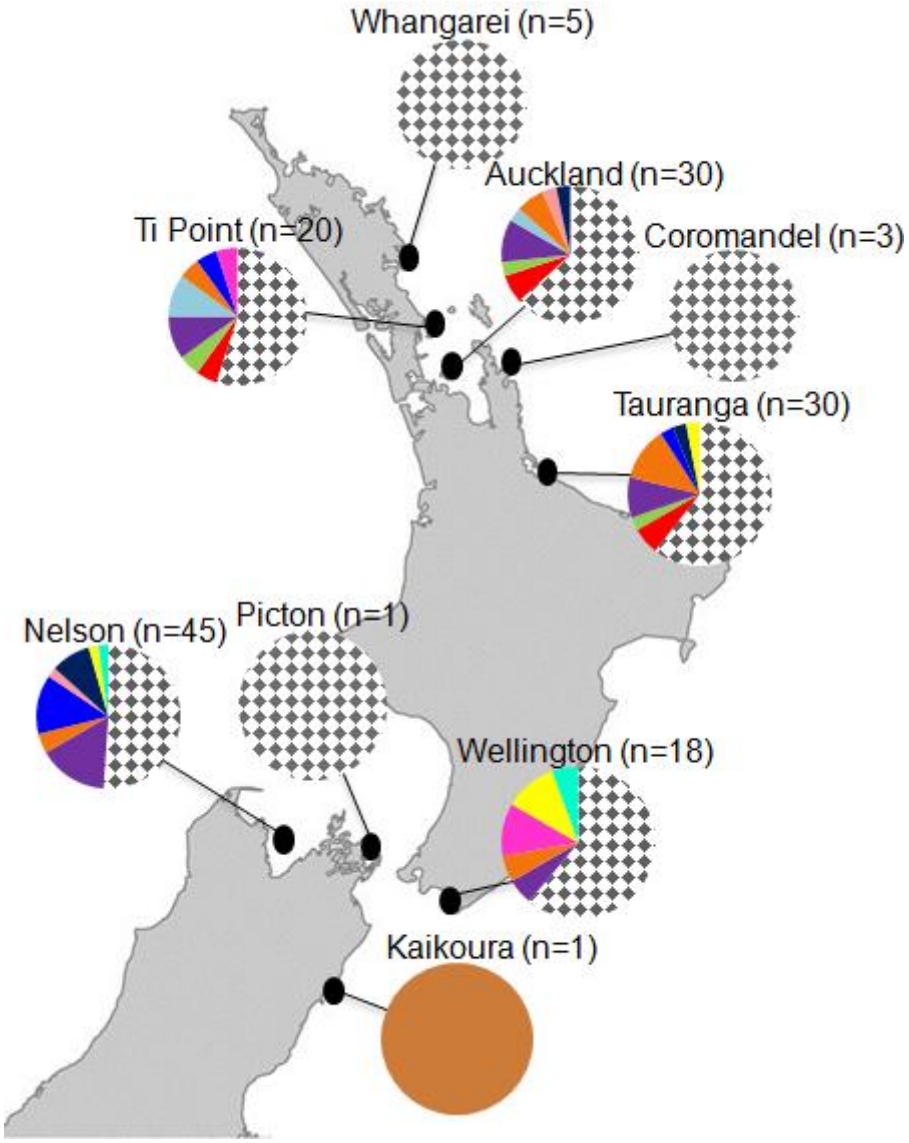
Supplementary Figure 3 Median joining network of the COI haplotypes from 156 individuals. The network was coloured according to the sampling locations. The diameters of the circles are proportional to the frequency of the haplotypes. The hashes indicate the mutational steps between the haplotypes. The black nodes represent the imaginary haplotypes necessary to create a bridge between the present haplotypes.



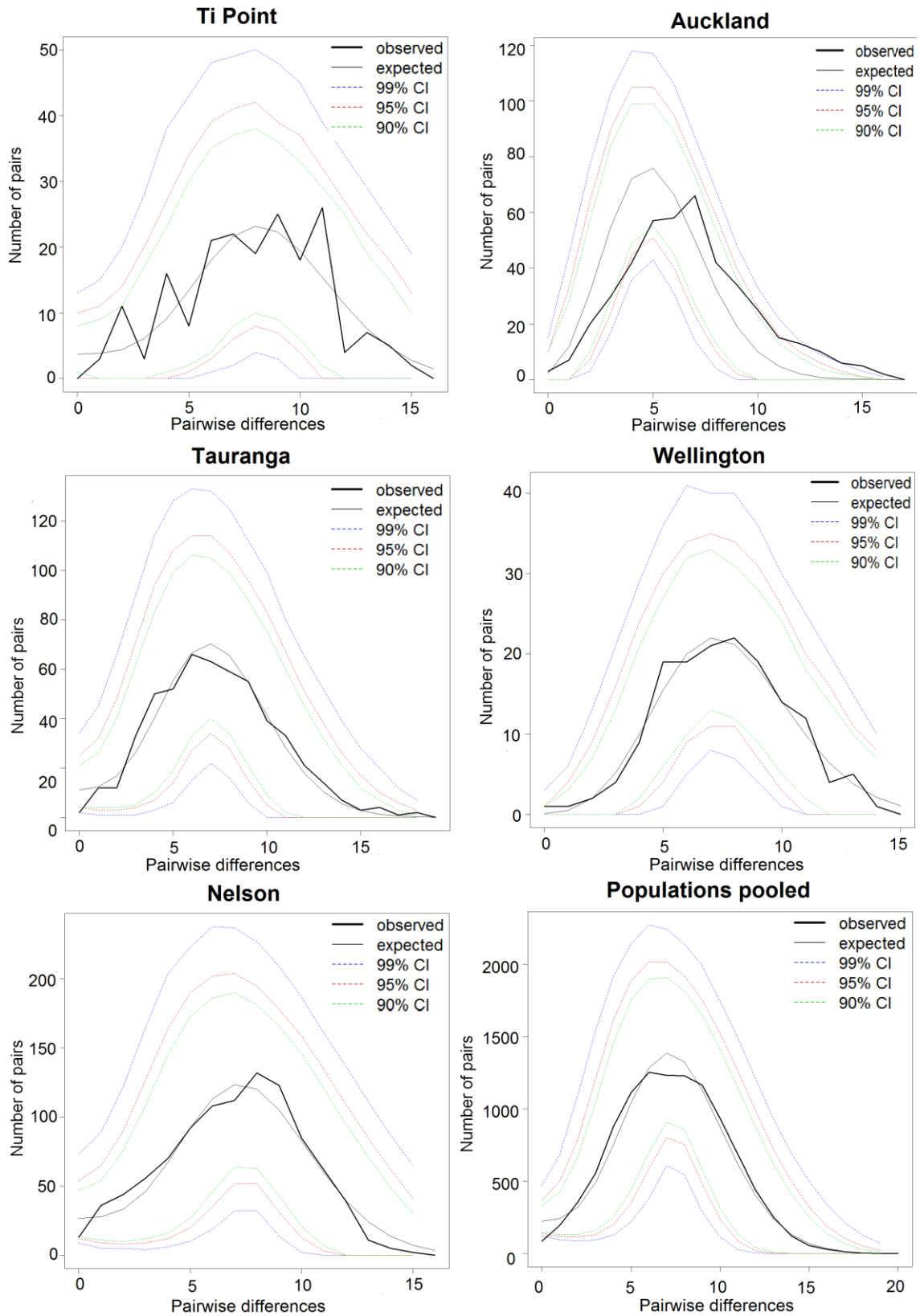
Supplementary Figure 4 Median joining network of the concatenated COI haplotypes from 156 samples. The network was coloured according to the sampling locations. The diameters of the circles are proportional to the frequency of the haplotypes. The hashes indicate the mutational steps between haplotypes. The black nodes represent the imaginary haplotypes necessary to create a bridge between the present haplotypes.



Supplementary Figure 5 Median joining networks drawn for quality error control or saturation at the third codons. The Cytb haplotypes were analysed in two fragments; **A)** 5' half (530 bp) amplified with one set of primers and **B)** 3' half (530 bp) amplified by a second set of primers; included the overlapping sequences with the first set of primers. **C)** COI haplotypes containing only 1st and 2nd positions of each codon. This analysis was performed to check whether the homoplasies at the third codon obscure any possible differentiation pattern or not. Population structure is not yet observed.



Supplementary Figure 6 The frequencies of the *P. maculata* COI haplotypes at each location. The pie segment represents the relative frequencies of Cytb haplotypes. Each colour corresponds to a different haplotype. The private haplotypes specific to each location are represented with a patterned segment.



Supplementary Figure 7 Mismatch distribution of pairwise base pair differences between the concatenated COI and Cytb haplotypes.

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Yeserin Yildirim

Name/Title of Principal Supervisor: Professor Paul B. Rainey

Name of Published Research Output and full reference:

Microsatellite development for a tetrodotoxin-containing sea slug (Pleurobranchaea maculata)

Yildirim, Y., S. Patel, C.D. Millar, and P.B. Rainey. 2014. Biochemical Systematics and Ecology. 55:342-345

Doi: 10.1016/j.bse.2014.04.001

In which Chapter is the Published Work: Chapter Two

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 80

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  The candidate developed the microsatellite markers after having obtained the partial genome sequence of the study organism. The genome sequencing process was performed by Selina Patel and Dr Craig Millar. Patel and Dr Millar provided practical and theoretical support during the development process. The candidate also wrote the manuscript.

Yeserin Yildirim Digitally signed by Yeserin Yildirim Date: 2015.07.08 10:24:21 +03'00'

Candidate's Signature

08/07/2015

Date

Paul Rainey Digitally signed by Paul Rainey DN: cn=Paul Rainey, o, ou=NZIAS, email=p.b.rainey@massey.ac.nz, c=NZ Date: 2015.07.09 14:12:12 +01'00'

Principal Supervisor's signature

09/07/2015

Date

GRS Version 3– 16 September 2011