Population genetics and connectivity in *Paphies subtriangulata* and *Paphies australis* (Bivalvia: Mesodesmatidae)

Danielle Amelia Hannan

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

Victoria University of Wellington

Te Whare Wānanga o te Ūpoko o te Ika a Māui

2014

Abstract

Understanding the different types of genetic population structure that characterise marine species, and the processes driving such patterns, is crucial for establishing links between the ecology and evolution of a species. This knowledge is vital for management and conservation of marine species. Genetic approaches are a powerful tool for revealing ecologically relevant insights to marine population dynamics. Geographic patterns of genetic population structure are largely determined by the rate at which individuals are exchanged among populations (termed 'population connectivity'), which in turn is influenced by conditions in the physical environment. The complexity of the New Zealand marine environment makes it difficult to predict how physical oceanographic and environmental processes will influence connectivity in coastal marine organisms and hence the type of genetic structure that will form. This complexity presents a challenge for management of marine resources but also makes the New Zealand region an interesting model system to investigate how and why population structure develops and evolves over time.

Paphies subtriangulata (tuatua) and *P. australis* (pipi) are endemic bivalve 'surf clams' commonly found on New Zealand surf beaches and harbour/estuary environments, respectively. They form important recreational, customary and commercial fisheries, yet little is known about the stock structure of these species. This study aimed to use genetic techniques to determine population structure, levels of connectivity and 'seascape' genetic patterns in *P. subtriangulata* and *P. australis*, and to gain further knowledge of common population genetic processes operating in the New Zealand coastal marine environment.

Eleven and 14 novel microsatellite markers were developed for *P. subtriangulata* and *P. australis*, respectively. Samples were collected from 10 locations for *P. subtriangulata* and 13 locations for *P. australis* (35-57 samples per location; total sample size of 517 for *P. subtriangulata* and 674 for *P. australis*). Geographic patterns of genetic variation were measured and rates of migration among locations were estimated on recent and historic time scales. Both species were characterised by genetic population structure that was

consistent with their habitat. For *P. subtriangulata*, the Chatham Island population was strongly differentiated from the rest of the sampled locations. The majority of mainland locations were undifferentiated and estimated rates of migration among locations were high on both time scales investigated, although differentiation among some populations was observed. For *P. australis*, an overall isolation by distance (IBD) pattern was likely to be driven by distance between discrete estuary habitats. However, it was difficult to distinguish IBD from hierarchical structure as populations could be further subdivided into three significantly differentiated groups (Northern, South Eastern and South Western), providing evidence for barriers to dispersal. Further small scale patterns of genetic differentiation were observed in some locations, suggesting that complex current patterns and high self-recruitment drive small scale genetic population structure in both *P. subtriangulata* and *P. australis*.

These patterns of genetic variation were used in seascape genetic analyses to test for associations with environmental variables, with the purpose of understanding the processes that might shape genetic population structure in these two species. Although genetic population structure varied between the two species, common physical and environmental variables (geographic distance, sea surface temperature, bed slope, tidal currents) are likely to be involved in the structuring of populations. Results suggest that local adaptation, in combination with restricted dispersal, could play a role in driving the small scale patterns of genetic differentiation seen among some localities.

Overall, the outcomes of this research fill a gap in our knowledge about the rates and routes by which populations are connected and the environmental factors influencing such patterns in the New Zealand marine environment. Other studies have highlighted the importance of using multi-faceted approaches to understand complex processes operating in the marine environment. The present study is an important first step in this direction as these methods are yet to be widely applied to New Zealand marine species. Importantly, this study used a comparative approach, applying standardised methodology to compare genetic population structure and migration across species. Such an approach is necessary if we wish to build a robust understanding of the spatial and temporal complexities of population dynamics in the New Zealand coastal marine environment, and to develop effective management strategies for our unique marine species.

Acknowledgements

This PhD thesis would not have been possible without the guidance and support of many people. I thank my primary supervisor, Dr Peter Ritchie, for allowing me the freedom to develop and manage my research project, and for always being available in times of need. I am also grateful to my co-supervisors, Professor Jonathan Gardner, Associate Professor James Bell and Associate Professor Joe Zuccarello for their advice and valuable feedback on the content of this thesis. The feedback and knowledge that I have received from all of my supervisors has allowed me to mature as both a researcher and an academic.

I would like to thank the other PhD students involved in the 'coastal connectivity' research group (Heather Constable and Catarina Silva) and those that shared the KK811 lab space with me, for their advice, help and friendship. In particular, Heather's help was valuable in the early stages of developing the field and laboratory methods for this research. I have many fond memories of our great sample collecting road trip where we partook in many beers and ice creams at the beach. I would also like to thank everyone else who assisted me in sample collection: Bridgit Bretherton-Jones, Nick Devereaux, Douglas Long and James Williams (NIWA).

This thesis was greatly improved by comments from Dr Monica Gruber and Darcy Webber. In particular Monica has been a fantastic office mate over the past few years and I am very grateful to her for sharing her knowledge and experience with me. I am also thankful to the administrative and technical staff at the School of Biological Sciences for their cheerful faces and assistance throughout my PhD.

I would like to acknowledge Dr Andrew Munkacsi, Dr Sharyn Goldstien and Dr Chris Burridge for their time and effort spent in examining my thesis. They set a challenging and thorough test of my knowledge and skills, and certainly made me work hard to deservedly achieve my PhD. I appreciate the feedback they gave on my thesis, which will add value to the publications that arise from this research. I would also like to thank the staff of the Faculty of Graduate Research for their assistance with the examination process.

Funding for this research was supplied via the Ministry for Primary Industries Biodiversity Research Advisory Group project ZBD2009-10. I am thankful for the financial support that they have supplied and it was a privilege as a post-graduate student to have the opportunity to participate in BRAG meetings and interact with some of New Zealand's most prominent marine scientists. I am also grateful to Victoria University of Wellington for their financial assistance via a Doctoral Submission Scholarship.

Finally I would like to thank all my family, friends and flatmates for their support and understanding over the last three years. I would like to thank the beautiful and talented ladies of the Victoria University Netball Club (NBHFL) for their incredible comradeship and for providing many a welcome distraction.

My sister Adrienne and her partner Scott gave me a place to stay upon arrival in Wellington and have supplied a poor student with many a roast dinner.

To my parents, Janet and Brian, I would not be where I was today if it wasn't for the love and opportunities that you have supplied me with, and the encouragement and belief that I can always achieve and succeed beyond my own expectations in whatever I set my mind to.

Table of Contents

List of Figures

List of Tables

Abbreviations

°C	Degrees Celsius
°C km^{-1}	Degrees Celsius per kilometre
ACC	Antarctic Circumpolar Front
AIC	Akaike information criterion
AMOVA	Analysis of molecular variance
BEST	Biological environmental stepwise
BLU	Bluff
bp	Base pair
CGP	Chaotic genetic patchiness
CHA	Chatham Island
CI	Credible Interval
cm	Centimetre
COL	Collingwood
DBT	Doubtful Sound
DC	D'Urville Current
ddH_2O	Double distilled water
dNTP	Deoxynucleotide triphosphate
E	East
EAUC	East Auckland Current
ECC	East Cape Current
ECE	East Cape Eddy
EEC	Estuarine Environment Classification
EEZ	Exclusive economic zone
F0	First generation migrant
FDR	False discovery rate
GLZ	Generalised linear model
HAK	Hakahaka Bay
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HUI	Huia

HWE	Hardy-Weinberg equilibrium
IBD	Isolation by distance
KAK	Kakamatua
KAR	Karamea
km	Kilometre
LGM	Last glacial maximum
LYT	Lyttelton
µ	Mutation rate
µl	Microlitre
µM	Micromole
m	Metre
m	Migration rate
MAR	Marfell Beach
MCMC	Markov chain Monte Carlo
MEC	Marine Environment Classification
mg	Milligram
$MgCl_2$	Magnesium chloride
mL	Millilitre
mm	Millimetre
mM	Millimole
m s^{-1}	Metres per second
n	Sample size
N	Census population size
N_a	Number of alleles
NAP	Napier
NCE	North Cape Eddy
N_e	Effective population size
N_em	Effective number of migrants
N_F	Frequency of null alleles
ng	Nanogram
NGS	Next generation sequencing
OAK	Oakura
OKU	Okuru
PA	Private alleles

PAP	Papamoa
PCA	Principal component analysis
PCR	Polymerase chain reaction
PET	Petone
$P_{(HWE)}$	Hardy-Weinberg equilibrium probability value
PKR	Paekakariki
PLD	Pelagic larval duration
QMA	Quota management area
QMS	Quota management system
R_a	Allelic richness
RAG	Raglan
R_S	Spearman rank co-efficient
RUA	Ruakaka
S	South
SAF	SubAntarctic Front
SC	Southland Current
SRS	Sweepstakes reproductive success
SST	Sea surface temperature
STF	Subtropical Front
θ	Population size parameter 'theta'
t	Tonnes
TACC	Total allowable commercial catch
TAP	Tapotupotu Bay
TAU	Tauranga
TF	Tasman Front
W	West
WAI	Waiwera
WAUC	West Auckland Current
WC	Westland Current
WCC	Wairarapa Coastal Current
WE	Wairarapa Eddy
W m^{-2}	Watts per metre squared
WMR	Waimarama
WPK	Waipapakauri

1 Introduction

1.1 Genetic population structure and connectivity in the marine environment

1.1.1 The link between genetic population structure, connectivity and larval dispersal

Understanding the different types of genetic population structure that characterise marine species, and the processes driving such patterns, is crucial for establishing links between the ecology and evolution of a species. The geographic distribution of individuals within a species' range (i.e., population structure) is largely determined by the rate at which individuals are exchanged between geographically separated populations, termed 'population connectivity' (Cowen *et al.* 2007). Many marine species are characterised by a bi-phasic life history, whereby adults are relatively sedentary compared to their larvae, which are released into the ocean environment. Therefore, it is the processes operating at the larval stage that largely determine the ability of larvae to disperse between geographically distinct locations and the type of population structure that will form. These processes are related to both the life history characteristics of the organism and the physical features of the ocean environment (Cowen & Sponaugle 2009). The small size of larvae relative to the large ocean environment poses a challenge when it comes to tracking dispersal routes among geographic locations. A range of methods are available to investigate population structure and connectivity (Levin 2006), of which population genetics is an increasingly powerful tool for revealing population processes operating at a range of spatial and temporal scales.

In the marine environment a continuum of genetic population structure exists, which ranges from 'open' to 'closed', and is attributed to relative levels of connectivity. In 'open' populations individuals are received and exported among populations, gene flow is high, random genetic drift is limited and there should be little opportunity for accumulation of genetic differences among populations. In 'closed' populations there is limited exchange of individuals among geographically separated populations, leading to reduced gene flow and

differentiation of populations via genetic drift or natural selection. Due to the inter-connected nature of the marine environment and perceived lack of barriers to larval dispersal, it has been predicted that most species should be characterised by 'open' population structure. Early studies using allozyme markers tended to support this view (Levin 2006).

The recent development of molecular markers with higher resolution has revealed that marine populations can possess genetic structure that is consistent with a 'closed' population model, meaning that barriers to larval dispersal do exist in the ocean and that populations may rely on self-recruitment to varying extent (Swearer *et al.* 2002). This phenomenon has been particularly well documented for reef fish (Jones *et al.* 1999; Swearer *et al.* 1999) as well as an increasing number of invertebrate taxa such as crabs, oysters, mussels, tiger shrimp, starfish, copepods and snails (Palumbi 2003). In the past much has been made of the role of pelagic larval duration (PLD) in shaping population structure, based on the assumption that longer PLD will allow larvae to disperse greater distances and populations will have greater connectivity (Bohonak 1999). However, re-examination of this relationship has provided conflicting views as to whether population connectivity can be considered a simple function of PLD (Bradbury *et al.* 2008b; Shanks 2009; Shanks *et al.* 2003; Weersing & Toonen 2009). Although there is some value in considering PLD as a simplistic proxy for dispersal ability, it seems that larvae cannot be viewed as passive particles at the mercy of ocean currents. Instead a number of factors, such as spawning output, oceanographic features, larval behaviour, predator/prey interactions, availability of suitable habitat, local adaptation and post-settlement survival (Cowen & Sponaugle 2009) can interact to modify larval distribution, promote self-recruitment and shape the genetic structure of populations.

1.1.2 Theoretical considerations of genetic population structure and connectivity

The inherent variability of the physical marine environment poses a challenge for analysing the genetic attributes of a species. For this reason population genetic analyses often involve matching measured genetic population structure to a few models that are based on simplified and stable population states. The theoretical basis of these models dates back to the work of Sewall Wright (1931, 1951), who proposed an 'island' model where populations of equal size are linked by equal numbers of migrants being exchanged

at a constant rate per generation. The island model is unlikely to be a realistic representation of the structure of most natural populations. The 'stepping-stone' model (Slatkin 1993) is thought to more realistically represent the 'linear' distribution of marine populations along a coastline. Under the stepping-stone model, populations exchange migrants with neighbouring populations at a greater rate than they do with more geographically distant populations. This migration model should result in an isolation by distance (IBD) pattern of genetic population structure, where geographically proximate populations are more genetically similar. Alternatively, populations that have unrestricted dispersal and lack geographic patterns of genetic structure are considered to have a single 'panmictic' gene pool.

The classical view of population genetics is that the type of population structure that forms is determined by the relative contribution and interaction of four forces: random genetic drift, gene flow, mutation and selection. Of these forces gene flow, or the effective exchange of migrants among populations per generation, is usually of most interest to population geneticists when determining connectivity among populations. Mutation and selection often receive little attention and are considered to be negligible, whereas the impact of genetic drift on allele frequencies should only become significant when gene flow and population sizes are relatively low (Hellberg *et al.* 2002). It is estimated that an effective migration rate of only one individual per generation is sufficient to counteract the effects of drift (Slatkin 1987). However, it is important to remember that low levels of gene flow might not be 'demographically' significant and populations may not be ecologically connected until the migration rate is greater than 10% of the population size (Hastings 1993).

As our knowledge of population genetics in marine species develops it is becoming increasingly evident that these simple theoretical models cannot sufficiently account for the complexity of processes and patterns that are observed. A greater understanding of the physical and ecological processes experienced by marine species has highlighted the stochastic and unpredictable nature of these processes (Siegel *et al.* 2008). It is becoming increasingly recognised that multi-disciplinary approaches combining genetic and ecological information are required to fully understand how population genetic structure is formed and evolves over time (Selkoe *et al.* 2010).

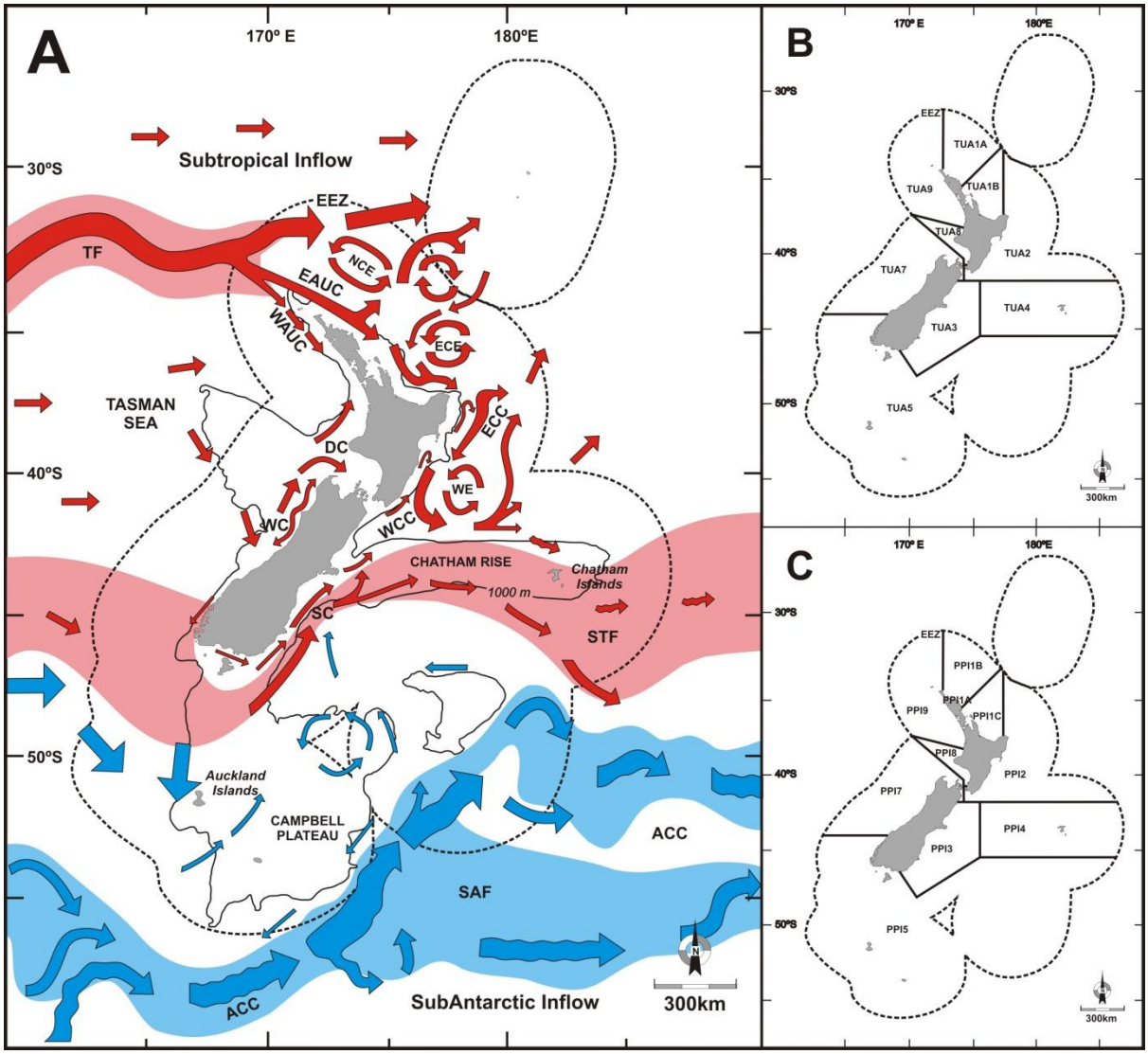


Figure 1.1: A: The New Zealand marine environment, showing the location of notable landforms, major current patterns and water masses, bathymetry and the New Zealand exclusive economic zone (EEZ). Abbreviations used for currents and water masses are as follows: Tasman Front (TF), East Auckland Current (EAUC), West Auckland Current (WAUC), North Cape Eddy (NCE), East Cape Eddy (ECE), East Coast Current (ECC), Wairarapa Eddy (WE), D'Urville Current (DC), Wairarapa Coastal Current (WCC), Westland Current (WC), Southland Current (SC), Subtropical Front (STF), Sub-Antarctic Front (SAF), Antarctic Circumpolar Current (ACC). Redrawn after Laing & Chiswell (2003). B: Location of quota management areas (QMA) used for management of the commercial and recreational *Paphies subtriangulata* fishery. C: QMAs used for management of the commercial and recreational *Paphies australis* fishery.

1.2 The New Zealand context

New Zealand is an archipelago of over 700 islands, stretching from the subtropical Kermadec Islands at 29° S to sub-Antarctic Campbell Island at 52° S (Figure 1.1A). The oceanography of the region is complex; it straddles the subtropical convergence and is influenced by the inflow of both warm sub-tropical and cold sub-Antarctic water masses (Laing & Chiswell 2003). Major westerly flowing current systems reach the continental shelf and are divided into numerous coastal currents and eddies. The latitudinal extent of the country means that significant north-south gradients exist in the environmental variables that characterise the physical ocean environment (Francis & Nelson 2003). It is difficult to predict the effect this complexity might have on the populations of marine organisms inhabiting the coastal marine environment (Ross *et al.* 2009). However, it is the complexity of the region that makes it ideal for investigating how current systems and variability in environmental characteristics can influence patterns of genetic structure and connectivity in coastal marine species. New Zealand's large exclusive economic zone (EEZ) contains a huge diversity of marine organisms, many of which are endemic and commercially valuable (Gordon *et al.* 2010). Understanding the patterns of connectivity that can exist among coastal marine species in the New Zealand region, and how they might arise, is essential for sustainable fisheries management, conservation and for biosecurity purposes.

To understand the effect that complex coastal marine environments can have on genetic population structure it is often useful to compare population genetic patterns across many species in the same region. This can be a powerful way to identify common or unexpected geographic barriers and patterns of connectivity, and can suggest hypotheses about the processes that form such patterns. Studies from the north-eastern Pacific have demonstrated the merit of this approach. For example, Kelly and Palumbi (2010) compared the genetic population structure of 50 rocky intertidal species from Alaska to California and found evidence for unexpected regional patterns of genetic variation. Furthermore, they found that habitat (in terms of height in the intertidal zone) had a strong influence on levels of genetic subdivision within a species. A further example from Pacific reef fish demonstrates how, by comparing connectivity patterns among several species, unexpected cases of gene flow could be detected across a well-established biogeographic barrier, which was previously thought to be largely impermeable (Lessios & Robertson 2006). These

studies demonstrate the value of comparing patterns of genetic population structure between species to understand connectivity at the ecosystem wide level required for large-scale management efforts. The attributes of the New Zealand coastal marine environment (as described above) make it an ideal region in which to repeat similar multi-species investigations.

The patterns of genetic population structure observed in New Zealand coastal marine species have been reviewed elsewhere (Gardner *et al.* 2010; Ross *et al.* 2009). Gardner *et al.* (2010) identified common patterns of genetic population structure by reviewing 58 studies of 42 coastal marine organisms, and summarised the patterns of observed genetic structure into five categories. Sixteen of these studies reported no genetic structure among populations, 9 reported isolation by distance, 12 reported divergence within and/or among populations, one reported east-west divergence and 19 north-south divergence. In studies that reported genetic breaks among populations, sampling effort was often insufficient to determine the geographic location of the barrier but studies that described north-south differentiation commonly reported the location of that break at around 41-42° S, the approximate location of Cook Strait. The review also identified habitat types (such as estuarine, soft substrate and open coast habitats) and geographic locations that had been poorly surveyed, and recommended that these knowledge gaps be filled. Choice of study species for the present research is therefore based on the recommendations of Gardner *et al.* (2010).

1.3 Background to the study species

The tuatua (*Paphies subtriangulata*; Wood 1928) and pipi (*P. australis*; Gmelin 1791) are bivalve molluscs belonging to the family Mesodesmatidae. In New Zealand the genus *Paphies* is comprised of large, edible 'surf clams', which also include the toheroa (*P. ventricosa*) and deep water tuatua (*P. donacina*). All four species form important recreational and customary fisheries and some support modest commercial catches (< 205 tonnes; Hooker 1997; New Zealand Ministry for Primary Industries 2013a; b; Redfearn 1987).

Paphies subtriangulata and *P. australis* have a widespread distribution around all of New Zealand and on offshore islands, including the Chatham Islands (both species) and

Auckland Island (*P. australis*) (Powell 1979). *Paphies subtriangulata* is one of the most abundant infaunal bivalves on fine-sand, open coast beaches and can be found from the low intertidal to depths of about four metres (Redfearn 1987). *Paphies australis* is characteristic of coarse shell and sand substrates in sheltered harbour and estuary environments, and can be found from the intertidal to depths of seven metres in channel areas (Morton & Miller 1968). The two species are known to coexist where these two habitats overlap, such as harbour entrances or sheltered sandy beaches (Grant *et al.* 1998).

1.3.1 Biology and ecology of the study species

1.3.1.1 Paphies subtriangulata

Paphies subtriangulata is gonochoristic and reproduces sexually by free-spawning of gametes, followed by external fertilisation. In the north-east of the North Island spawning is known to occur from February to April, followed by regeneration of the gonad from May to late August and resumption of spawning activity from September to November (Grant & Creese 1995). Only a small proportion of the population spawns at one time and there are large variations in spawning activity over the spawning period (Grant & Creese 1995). Due to similarities in their reproductive behaviour, *P. subtriangulata* and *P. australis* have been observed spawning synchronously in areas where their habitat overlaps, suggesting the possibility of hybridisation. *In vitro* experiments suggest that hybridisation can occur between these two species at high sperm concentrations, although in the natural environment it is likely to be minimal due to potential prezygotic reproductive incompatibility (Grant *et al.* 1998).

The pelagic larval period lasts two to three weeks before settlement to the adult habitat (Redfearn 1987). Settlement occurs high in the intertidal, but spat and juveniles are highly mobile, moving around in the swash zone and quickly reburying themselves. Mortality at this stage is thought to be high, as is common with most surf clam species (New Zealand Ministry for Primary Industries 2013b). Growth is rapid but variable, with a shell length of 40 to 70 mm reached in about 3 years. Maximum length varies among areas, ranging from about 50 to 80 mm (New Zealand Ministry for Primary Industries 2013b). As individuals grow larger they migrate down the beach to occupy the lower intertidal and shallow subtidal. Adults are filter feeders and can be found wedged a few centimetres

below the surface of the sand, with the siphon end often exposed and discoloured by a green or brown algal film (Richardson *et al.* 1982). Maximum age is probably about five or more years (New Zealand Ministry for Primary Industries 2013b).

There has been some confusion over the status of tuatua in New Zealand due to the morphological similarity of the tuatua (*P. subtriangulata*) and the deep water tuatua (*P. donacina*). Initially it was proposed that the two species represented two ends of a geographic cline as *P. subtriangulata* is more commonly found on northern beaches and *P. donacina* more common on southern beaches (Beu & de Rooij-Schuiling 1982). Their status was resolved by examination of shell shape and colour, adductor muscle colour, and an electrophoretic marker which clearly revealed that they were separate species (Richardson *et al.* 1982). Variability in *P. subtriangulata* morphology has been noted from different populations around New Zealand. Shells from the north-east coast of the North Island have a different shape to those from Wellington beaches, large animals (> 80 mm) are found on east coast South Island beaches (Smith *et al.* 1989), and shells from the Chatham Islands have been considered a third species of tuatua, *P. porrecta* (Beu & de Rooij-Schuiling 1982). An investigation of allozyme polymorphism at four loci from 13 locations around New Zealand indicated the presence of three geographical groups of *P. subtriangulata*: north, central, and Chatham Island (Smith *et al.* 1989). Samples from Stewart Island may represent a fourth southern group owing to this population's geographical isolation and genetic differences with the central group (Smith *et al.* 1989).

1.3.1.2 Paphies australis

Paphies australis has a similar reproductive cycle to *P. subtriangulata*. Separate sexes release gametes into the water column, followed by external fertilisation. Sexual maturity occurs at a shell length of about 40 mm (Hooker & Creese 1995a). Gametes begin to form in autumn and by late winter gonads are mature and ready to spawn (Hooker & Creese 1995b). The gametes are released over an extended spawning period from late winter to late summer, usually as a series of partial spawning events (Hooker & Creese 1995a). The fertilised eggs develop into planktotropic larvae, which spend approximately three weeks in the pelagic environment before metamorphosis and settlement to the adult habitat (Hooker 1997). After settlement juveniles are generally sedentary, but there is

evidence that juveniles can actively migrate via drifting by buoying themselves up using mucus threads secreted by their siphons (Hooker 1995).

Juvenile growth appears to be rapid at first, with a length of 40 mm reached in approximately 18 months, and 55 to 60 mm reached by three to four years of age (Hooker 1997). Adults can reach up to 90 mm shell length. There appears to be a strong seasonal difference in growth rates, with rapid growth occurring in spring and summer, and slower growth rates during autumn and winter (Hooker 1997). Adults are suspension feeders with short siphons, and can be found in quite high densities just below the surface of the sand with their posterior end protruding (Williams *et al.* 2007).

Little is known about mortality and longevity in *P. australis* but it has been suggested that they can live up to 10 years (Williams *et al.* 2007). Post-recruitment mortality is known to be high in juveniles, and appears to be related to extremely localised recruitment and high population densities on a small spatial scale (Cole *et al.* 2000). Salinity is also known to affect mortality as prolonged exposure to low salinity environments results in increased mortality (McLeod & Wing 2008). *Paphies australis* inhabits a dynamic environment that is subject to continual erosion and sedimentation processes caused by both natural and anthropogenic events. Prolonged exposure will result in predation mortality, whereas their short siphons mean that burial will result in starvation and a lack of oxygen. An investigation of burrowing behaviour in this species suggests that in their natural orientation, they are able to rebury and can survive sediment inundations of up to 10 cm per day (Hull *et al.* 1998). Burrowing ability is slow compared to other New Zealand clams, such as *P. subtriangulata* and the trough shell (*Mactra discors*), but they have the ability to support anaerobic energy production whilst buried (Carroll & Wells 1995).

1.3.2 Management of *Paphies subtriangulata* and *Paphies australis* fisheries

Paphies subtriangulata and *P. australis* are both managed under the quota management system (QMS) yet there is little information on the stock structure of these species. The fisheries are primarily recreational and customary with shellfish being gathered by hand, but the level of recreational/customary harvest is unknown and there is no minimum legal size (New Zealand Ministry for Primary Industries 2013a; b). The recreational daily catch limit for both species is 150 per person. Many stocks are assumed to be near

virgin biomass but there is concern for depletion of some local stocks, especially in the Auckland-Coromandel region where the limit has been reduced to 50 per person per day. High variability in larval recruitment is likely to mean that biomass is also highly variable (New Zealand Ministry for Primary Industries 2013a; b).

Commercial harvest of *P. subtriangulata* is a dredge fishery restricted to the Kaipara Harbour in the TUA 9 quota management area (QMA) with a total allowable commercial catch (TACC) of 43 t per year (Figure 1.1B; New Zealand Ministry for Primary Industries 2013b). Commercial harvest of *P. australis* primarily comes from Mair Bank Whangarei, in the PPI 1A QMA, with a TACC of 200 t per year (Figure 1.1C; New Zealand Ministry for Primary Industries 2013a). It has been estimated that *P. australis* is harvested below the maximum sustainable yield but it is unknown if these levels of harvest are sustainable in the long term (New Zealand Ministry for Fisheries 2010). The status of all *P. subtriangulata* stocks is unknown (New Zealand Ministry for Primary Industries 2013b).

1.4 Research needs and predictions

While a previous study of allozyme variation in *P. subtriangulata* suggests that some population differentiation might be present in this species (Smith *et al.* 1989), it is unknown if genetically distinct populations of *P. australis* exist and how patterns of genetic variation might be geographically structured. Further investigation of *P. subtriangulata* genetic population structure using microsatellite markers is worthwhile as microsatellites can provide more resolution than allozyme markers. Patterns and rates of larval movement among populations, and the physical/environmental variables that influence genetic population structure are also unknown for both species. Information regarding a species' genetic structure and population connectivity is required for sustainable management of fisheries. Additionally, such information will contribute to a growing body of literature on patterns of genetic population connectivity in the New Zealand marine environment, particularly as these species are characteristic of habitats that have been under-represented in past studies (Gardner *et al.* 2010). As patterns of marine connectivity in New Zealand become better understood, this will assist with marine biosecurity, fisheries management and identification of appropriate sites for marine protected areas.

Based on what is known about the ecology and reproductive biology of *P. subtriangulata* and *P. australis*, some predictions can be made about the type of genetic population structure likely to be observed in this study. Both species have pelagic larval durations of approximately three weeks, allowing larvae to potentially disperse among populations separated by several hundred kilometres (Shanks 2009; Shanks *et al.* 2003). The primary way in which these two species differ is in their habitat. The open sandy coastlines inhabited by *P. subtriangulata* are thought to be characterised by fewer barriers to larval dispersal. Several studies of bivalves inhabiting such environments, such as the soft-shell clam (*Mya arenaria*; Strasser & Barber 2008), the Arctic surf clam (*Mactromeris polynyma*; Cassista & Hart 2007) and the surf clams *Donax serra* and *D. deltoides* (Laudien *et al.* 2003; Murray-Jones & Ayre 1997), have revealed little genetic structure among populations separated by thousands of kilometres. Other studies have revealed some structure among populations but usually associated with a biogeographic break (*Merceneria merceneria*; Baker *et al.* 2008) a physical barrier (*Coelomactra antiquate*; Kong *et al.* 2007) or isolation by distance (*Mactra veneriformis*; Hou *et al.* 2006). As mentioned previously, an earlier study of *P. subtriangulata* suggests that some genetic differentiation among populations may exist, implying that there are barriers to dispersal for the larvae of this species (Smith *et al.* 1989).

In comparison, the estuarine habitats favoured by *P. australis* often represent discrete environments separated from each other by ecological and geographic barriers. Therefore it is expected that the exchange of individuals between estuaries will be limited, leading to the genetic differentiation of populations (Bilton 2002). However, on a geological timescale estuaries can be transitory. The origin of many present day estuaries is likely to be recent, so it is possible that there has been little time for the accumulation of genetic differentiation among populations in such habitats (Williams *et al.* 2008). Presently the only study of an estuarine bivalve in New Zealand has been of the common cockle (*Austrovenus stutchburyi*), which reported genetic differentiation among populations consistent with isolation by distance (Ross *et al.* 2011). There are also examples of genetic differentiation among populations of estuarine bivalves outside New Zealand waters, such as the lagoon cockle *Cerastoderma glaucum* (Tarnowska *et al.* 2010), the oyster *Crassostrea ariakensis* (Xiao *et al.* 2010), as well as among estuarine amphipods (Kelly *et al.* 2006) and fish (Bradbury *et al.* 2008a; McCraney *et al.* 2010).

If habitat differences between the two species are considered, then based on results of the studies cited above it seems likely that connectivity among *P. australis* populations may be limited in comparison to connectivity among *P. subtriangulata* populations. This should result in higher levels of genetic differentiation among *P. australis* populations compared to *P. subtriangulata* populations. However, while habitat considerations are important, the cited studies often highlight the role that physical oceanographic features, environmental variability, historical events and natural selection can have on shaping the genetic population structure of marine species. It is difficult to predict the extent to which these factors will influence genetic population structure in the study species.

To date studies of genetic population structure in New Zealand marine organisms have primarily focussed on determining the geographic patterns of genetic diversity. With the exception of Wei *et al.* (2013a), no attempts have been made to quantify rates or patterns of migration among populations. Instead population connectivity is implied based on genetic similarities or differences among populations. Many studies often speculate on the physical or environmental processes that could be responsible for the observed patterns of genetic structure but few have attempted to quantify these variables and test for associations with the observed genetic patterns.

Recent advances in computational power and theoretical frameworks have made it possible to measure levels of connectivity among populations on a variety of temporal and spatial scales, allowing for better understanding of the genetic population structure uncovered by more traditional population genetic methods (Hauser & Carvalho 2008; Pearse & Crandall 2004). The value of these methods have been demonstrated by studies from outside the New Zealand region (e.g., Fraser *et al.* 2007; Harris *et al.* 2012; Jolly *et al.* 2009; Weetman *et al.* 2006) but are yet to be widely applied to the New Zealand context. As more is understood about how the biological characteristics of a species interacts with its physical environment to produce observed patterns of genetic population structure, there is a growing awareness of the need for multi-disciplinary analyses that incorporate genetic and ecological information. This has led to a rapidly developing field of 'seascape' genetics, which attempts to use environmental variation to shed light on complex genetic patterns (Galindo *et al.* 2010; Riginos & Liggins 2013; Selkoe *et al.* 2010). Furthermore, there is growing awareness of the need for comparative studies incorporating multiple species

(Riginos & Liggins 2013). By using standardised methods to compare population genetic patterns to environmental patterns across multiple species we gain an understanding of commonalities in the population processes occurring in a region (Bohonak 1999). It may be possible to generalise these findings to other ecologically similar species within that region. These recent developments in the field of marine population genetics lead us closer to our goal of understanding the links between the ecology and evolution of marine species.

Currently we have a good understanding of the types of genetic structure that can be observed in New Zealand marine organisms (Gardner *et al.* 2010; Ross *et al.* 2009) but there is a lack of knowledge about rates of connectivity among populations and the environmental variables responsible for driving patterns of population structure. There is a need to move beyond simplistic investigations of genetic population structure and instead focus on multi-faceted analyses that can provide a more realistic picture of the complex processes shaping marine populations over time.

1.5 Aims of the present research

The primary goal of this research was to determine patterns of genetic population structure and genetic connectivity in *P. subtriangulata* and *P. australis,* and to determine the location of any potential barriers to connectivity. This section of the study tested a null hypothesis of panmixia (i.e., no genetic population structure), which implies high connectivity and no barriers to dispersal. This information will be valuable for management of *P. subtriangulata* and *P. australis* fisheries, and may also be of relevance for management of surf clam fisheries as a whole. Management of these species is not currently backed up by information on the location of discrete stocks or knowledge of recruitment sources and sinks for such stocks. Furthermore, this study adds to our knowledge on the types of genetic population structure that can exist in New Zealand coastal marine species. This study also begins to fill a gap that has been identified in our knowledge of New Zealand marine population connectivity, in terms of estimating the routes and rates by which populations are connected.

Secondly, this study aimed to identify any environmental variables that were likely to explain the observed patterns of genetic structure and connectivity. This section of the study tested the null hypothesis that there was no significant correlation between genetic

and environmental variation. Next, this study sought to determine patterns of genetic population structure, genetic connectivity, barriers to dispersal and environmental variables that were common between the two study species. Finally, the patterns of genetic population structure and connectivity, and the barriers to dispersal observed in the present study were compared to those seen in other New Zealand coastal marine species, to identify patterns that are common across multiple species. By identifying common patterns between the two study species and other New Zealand coastal marine species it could be possible to understand the features of these species that might be responsible for driving their genetic population structure. Likewise, identifying the reasons for differences in genetic population structure between species can be equally informative. Comparisons of genetic population structure and connectivity in this study were based on the hypothesis that differences between the two study species are likely to be a result of their habitat differences, whereas any similarities are likely to be a result of their similar reproductive strategies. If we can understand how and why genetic population structure forms and changes over time this will allow for the development of effective management and conservation strategies in the present and in the face of impending climate change expected in the near future.

To achieve these aims this thesis is divided into a further five chapters, focussing on the following topics:

Chapter 2 described novel microsatellite markers developed for *P. subtriangulata* and *P. australis* that were used in all population genetics and connectivity analyses.

Chapter 3 measured genetic diversity and determined spatial patterns of genetic structure among ten *P. subtriangulata* locations. Rates and patterns of migration among these locations were estimated at both recent and historical time scales.

Chapter 4 measured genetic diversity and determined spatial patterns of genetic structure among thirteen *P. australis* locations. Rates and patterns of migration among these locations were estimated at both recent and historical time scales.

Chapter 5 used the levels of genetic variation measured for *P. subtriangulata* and *P. australis* populations in chapters 3 and 4 to test for associations with a number of geospatial

and environmental variables. The purpose of these analyses was to determine if genetic variation among locations was correlated with variation in environmental features. This is a first step to understanding how key environmental processes might shape the evolution of populations of marine species.

Chapter 6 summarises the major findings of the research and presents a synthesis of results from each chapter. Genetic population structure and connectivity patterns from the two study species were compared and contrasted to determine the processes involved in shaping population structure in New Zealand's coastal marine species. The results from the present study were also compared to genetic population structure observed in other New Zealand marine species to identify common patterns that may assist with development of management and conservation strategies. The implications for management and conservation of marine species are discussed and directions for future research are presented.

2 Characterisation of novel microsatellite markers for *Paphies subtriangulata* and *Paphies australis*

2.1 Introduction

Microsatellite markers are increasingly becoming the genetic tool of choice for population genetics studies due to their ease of use and power to answer a wide range of ecologically relevant questions (Selkoe & Toonen 2006). Microsatellite markers consist of short motifs, typically two to six base pairs in length, which are repeated in tandem. Due to the nature of these repeated DNA motifs they are thought to be prone to slippage during DNA replication, resulting in high variation in repeat number between individuals. This high mutation rate, along with their co-dominant nature, means that they are ideal for fine-scale investigation of population structure. In particular, they can be used to identify patterns of genetic variation between populations within a species on relatively recent time scales, which may be of relevance to contemporary demographic processes.

Recent advances in next generation sequencing (NGS) technologies mean that it is now cheap and efficient to develop microsatellite markers for a range of non-model organisms (Abdelkrim *et al.* 2009). Microsatellite development can be particularly challenging for marine invertebrates as high fecundity is often associated with high mutation rates, meaning that large numbers of null alleles (alleles that fail to amplify via polymerase chain reaction) can be present in the genome (Cruz *et al.* 2005; Hedgecock *et al.* 2004). Furthermore, such species often fail to meet idealised Hardy-Weinberg equilibrium (HWE) expectations that form the underlying basis of many population genetics analyses (Addison & Hart 2005). By systematically screening loci and measuring potential sources of error, it is possible to mitigate these problems or at least be aware of the scale to which they might influence results.

In addition, the NGS data used for microsatellite development represents a significant genetic resource that can be mined for other purposes. In particular, it is possible to identify single nucleotide polymorphisms (SNPs) that can also be used in population genetic studies. SNPs are found in high frequency throughout the genome and are highly informative in large numbers (Morin *et al.* 2004). However, the expense associated with identifying enough SNPs for meaningful population genetics analysis has, until recently, been prohibitive. Furthermore, if candidate loci associated with local adaptation to environmental variation are identified and homologous sequences from closely related species are known, then these loci can be identified for the target species from available NGS data.

The primary aim of this chapter was to describe novel microsatellite markers that were developed for the New Zealand surf clams *Paphies subtriangulata* (tuatua) and *P. australis* (pipi). Markers were developed for the purposes of investigating genetic population structure, connectivity patterns and associations with environmental variation for these two species. This knowledge will aid in the management of these species and lead to a better understanding of the overall patterns of gene flow and connectivity among populations of coastal marine species in New Zealand. In addition, 454 sequence data was screened and numbers of putative SNP markers reported.

2.2 Methods

Genomic DNA was extracted from the foot tissue of *P. subtriangulata* collected from Paekakariki (-40.99° S, 174.95° E) and *P. australis* collected from Petone (-41.23° S, 174.86° E) using a standard proteinase K digestion followed by phenol-chloroform purification and ethanol precipitation. DNA extracts were used for two 1/8 run on the Roche 454 GS-FLX platform. Resulting sequences were searched for microsatellite repeats with motifs of between two and six base pairs (bp) in iQDD 1.3 (Meglécz *et al.* 2010), using default parameters. For *P. australis* the Phobos 3.3.12 plugin (http://www.rub.de/spezzoo/cm/cm_phobos.htm) as implemented in GENEIOUS 5.5.6 (Biomatters) was also used to search for microsatellite repeats of between two and eight base pairs, using default parameters. Primers were designed for microsatellite regions containing five or more repeat units using Primer3 1.1.4 (Rozen & Skaletsky 2000).

Primers were screened via polymerase chain reaction (PCR) using a GeneAmp 2700 (Life Technologies^{TM}) thermocycler in a final reaction volume of 10 µl containing: ~50 ng template DNA (as quantified by a Implen Nanophotometer), 1X Reaction Buffer (Fisher Biotec), 3 mM $MgCl_2$, 0.4 mg/mL Bovine Serum Albumen, 50 µM of each dNTP, 0.4 µM of a fluorescently labelled M13 primer (Schuelke 2000), 0.4 µM reverse primer, 0.1 µM forward primer, 1 unit of *Taq* polymerase (Fisher Biotec) and ddH_2O to volume. Cycle conditions were as follows: 94°C for 5 minutes; followed by 22 cycles of 94°C for 30 seconds, 62°C for 45 seconds, 74°C for 60 seconds; followed by 8 cycles of 94°C for 30 seconds, 53°C for 45 seconds, 74°C for 60 seconds; followed by a final extension at 74°C for 10 minutes. PCR products were visualised on a 1.5% agarose gel using ethidium bromide staining and, if successfully amplified, PCR products were then size separated and the fluorescently labelled amplicons detected using an ABI3730XL (Life Technologies^{TM}) automated capillary sequencer. Allele sizes were determined using Peak Scanner 1.0 (Life Technologies^{TM}) and binned using Autobin (http://www4.bordeaux-aquitaine.inra.fr/biogeco/Media/Ressources/Logiciels/Autobin).

Eleven loci for *P. subtriangulata* and fourteen loci for *P. australis* were found to be polymorphic and amplify consistently. Markers were tested for their ability to cross-amplify with non-target species within the genus *Paphies* (i.e., *P. australis*, *P. subtriangulata* and *P. donacina*). However, none of the markers could be consistently amplified or scored for any species other than the one that they were developed for. The 5' ends of the forward primers were tagged with the fluorescent labels FAM, VIC, NED or PET (Table 2.2). Multiplex Manager 1.2 (Holleley & Geerts 2009) was used to arrange loci into two multiplex PCR panels for each species using the same PCR conditions as above with the following modifications: (1) Removal of the M13 primer, equal amounts of forward and reverse primer were used at concentrations listed in Table 2.2; (2) Removal of the M13 annealing step, instead 34 cycles with an annealing temperature of 62°C were used .

For each species 10 individuals from five locations (50 individuals in total) were genotyped and basic statistics were calculated for each locus. This sample size was deemed sufficient to determine if levels of across population polymorphism were adequate for loci to be informative in further population genetic analyses. Loci were checked for scoring errors and large allele dropout using MICROCHECKER 2.2.3 (van Oosterhout *et al.* 2004).

GENEPOP 4.2 (Rousset 2008) was used to test for linkage disequilibrium among loci and deviation from HWE expectations (Markov-chain parameters: 10,000 dememorization steps, 1,000 batches and 10,000 iterations per batch). Comparisons were considered significant if $p < 0.05$ after implementing the false discovery rate (FDR) correction for multiple tests (Verhoeven *et al.* 2005). The F_{ST} outlier method, implemented in LOSITAN (Antao *et al.* 2008; Beaumont & Nichols 1996) was used to identify potential outlier loci that may be influenced by selection (50,000 simulations using a stepwise mutation model). FreeNA (Chapuis & Estoup 2007) was used to estimate proportions of null alleles at each locus using 1,000 bootstrap replicates. FSTAT 2.9.3.2 (Goudet 1995) was used to calculate the number of alleles and allelic richness. ARLEQUIN 3.5 (Excoffier & Lischer 2010) was used to calculate observed (H_O) and expected (H_E) heterozygosity for each location and total for each locus. GENEPOP 4.2 (Rousset 2008) was used to calculate the inbreeding co-efficient F_{IS}.

A *de novo* assembly of the 454 sequences was performed for each species using default parameters in GENEIOUS 5.5.6 (Biomatters). Contigs were screened for putative SNPs using the 'Find Variations/SNPSs' tool with a minimum coverage of 5 sequences and a minimum variant frequency of 0.25.

2.3 Results

A 1/8 run on the Roche 454 GS-FLX platform generated 153,761 sequences (mean length 546 bp) for *P. subtriangulata* and 108,871 sequences (mean length 246 bp) for *P. australis* (Table 2.1). From the available microsatellite containing sequences, primers were designed for 47 *P. subtriangulata* loci and 73 *P. australis* loci (Table 2.1). Loci were discarded if they failed to amplify in PCR, displayed high levels of allelic drop out, were confounded by artefacts or were monomorphic. The screening process resulted in the final panels of 11 *P. subtriangulata* loci and 14 *P. australis* loci that were polymorphic and amplified consistently in multiplex (Table 2.2).

No evidence was found for scoring errors, large allele drop out or linkage dis-equilibrium among loci. Significant HWE deviation was observed for two *P. subtriangulata* loci (*Psub_5* and *Psub_7*) and eight *P. australis* loci (*Paus_2, Paus_3, Paus_4, Paus_6, Paus_8, Paus_9, Paus_11, Paus_14*; Table 2.2). LOSITAN indicated that two *P. australis* loci (*Paus_6* and *Paus_9*) were potential F_{ST} outliers that may be influenced by selection. Four *P.*

subtriangulata loci and eight *P. australis* loci had levels of null alleles (>10%; Chapuis & Estoup 2007) that may affect population differentiation estimates (Table 2.2).

The number of alleles ranged from 4-13 for *P. subtriangulata* and 5-19 for *P. australis* (Table 2.2). Allelic richness ranged from 2.39-6.80 for *P. subtriangulata* and 2.63-10.45 for *P. australis* (Table 2.2). Total observed heterozygosity (H_O) for each locus ranged from 0.16-0.62 for *P. subtriangulata* and 0.20-0.70 for *P. australis*. Total expected heterozygosity (H_E) ranged from 0.19-0.83 for *P. subtriangulata* and 0.30-0.89 for *P. australis* (Table 2.2). F_{IS} ranged from -0.11-0.77 for *P. subtriangulata* and -0.04-0.95 for *P. australis* (Table 2.2).

A *de novo* assembly of the 454 sequences obtained 278 contigs for *P. subtriangulata* and 832 contigs for *P. australis* with a coverage of 10 or more reads. A total of 745 and 613 putative SNPs were identified for *P. subtriangulata* and *P. australis*, respectively. These SNPs represent heterozygote variation in the individuals used for 454 sequencing are a potential genetic resource that could be developed for these species in the future.

Table 2.1: Summary of 454 sequencing reads and microsatellite primer design for *Paphies subtriangulata* and *P. australis.*

	Paphies subtriangulata	***Paphies australis***
Number of Reads	153,761	108,871
Average Length	546	246
Microsatellite Containing Reads	13,366	1,618
Number of Reads with Primers Designed	713	161
Dinucleotide Repeats (%)	74.9	79.9
Trinucleotide Repeats (%)	16.8	14.8
Tetranucleotide Repeats (%)	7.1	3.4
Pentanucleotide Repeats (%)	1.0	0.7
Other Repeats (%)	0.0	1.2
Number of Loci Tested	47	73
Number of Loci Selected	11	14

Table 2.2 (over page): Characteristics of the microsatellite markers developed for *Paphies subtriangulata* and *P. australis*, showing arrangement of multiplexes, fluorescent tag, primer concentration, forward and reverse primer sequence, and repeat motif. Genetic diversity indices (N_a: number of alleles, R_a: allelic richness, H_O: observed heterozygosity, H_E: expected heterozygosity, $P_{(HWE)}$: HWE *p*-value, N_F: estimated frequency of null alleles, F_{IS}: inbreeding co-efficient) were calculated for *n* number of samples in total and 10 individuals from each of the following locations: for *P. subtriangulata* Waipapakauri (WPK -35.04° S, 173.17° E), Paekakariki (PKR -40.99° S, 174.95° E) Marfell Beach (MAR -41.72° S, 174.20° E), Waimarama (WMR -39.83° S, 177.00° E) and Papamoa (PAP -37.70° S, 176.29° E); and for *P. australis* Waiwera (WAI -36.54° S, 174.71° E), Huia (HUI -37.00° S, 174.57° E), Raglan (RAG -37.80° S, 174.87° E), Lyttelton (LYT -43.64° S, 172.75° E) and Doubtful Sound (DBT -45.28° S, 166.91° E). Bold values represent significant HWE *p*-values after false discovery rate correction for multiple tests ($\alpha = 0.05$).

Table 2.2.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
***P. subtriangulata* Multiplex 1**														
Psub_2	NED	0.30	F: GCGAATGACTTTGTTTGGGT	$(GA)_7$	50	13	5.92	118-138	WPK	0.400	0.558	0.356	0.066	-0.186
			R: TTAAGTCCACCGAATTTGCC						PKR	0.400	0.511	0.235	0.094	0.324
									MAR	0.300	0.505	0.080	0.113	0.002
									WMR	0.400	0.505	0.653	0.072	0.025
									PAP	0.300	0.816	**0.002**	0.275	0.131
									Total	0.360	0.591	0.009	0.160	0.063
Psub_4	FAM	0.20	F: ATGTGGTCGAGTGTCATGGA	$(CAT)_5$	49	4	2.39	207-216	WPK	0.400	0.337	1.000	0.000	-0.200
			R: AGAACCTGGTCGTCATGCTT						PKR	0.000	0.000	1.000	0.001	0.000
									MAR	0.200	0.195	1.000	0.000	-0.047
									WMR	0.100	0.100	1.000	0.000	0.000
									PAP	0.111	0.307	0.059	0.178	0.419
									Total	0.162	0.189	0.461	0.052	0.203
Psub_5	FAM	0.40	F: TAATGTAAGCCGGTGTGCAA	$(AG)_7$	49	8	5.33	93-107	WPK	0.500	0.663	0.413	0.048	-0.067
			R: TGATGGCACCTTTCAACAAA						PKR	0.100	0.563	**0.001**	0.300	0.928
									MAR	0.200	0.611	**0.004**	0.260	0.851
									WMR	0.600	0.837	0.158	0.108	0.621
									PAP	0.000	0.758	**<0.001**	0.419	1.000
									Total	0.280	0.709	**<0.001**	0.256	0.729
Psub_7	NED	0.60	F: ATGAGAAATGATTGGTCGGC	$(TA)_8$	47	12	6.80	167-239	WPK	0.556	0.850	0.016	0.135	0.489
			R: CGCCGAAATCCAATCTAAAA						PKR	0.800	0.847	0.694	0.000	0.008
									MAR	0.444	0.830	0.009	0.187	-0.018
									WMR	0.400	0.858	**0.001**	0.226	0.686
									PAP	0.556	0.660	0.161	0.092	0.740
									Total	0.551	0.825	**<0.001**	0.149	0.131

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Psub_8	PET	0.60	F: TTCCTACGGCAATTCTGACC	$(TA)_8$	47	7	4.72	202-238	WPK	0.444	0.752	0.022	0.158	0.942
			R: CCGTCTACGCAACACAACAA						PKR	0.400	0.611	0.116	0.143	0.434
									MAR	0.889	0.621	0.276	0.000	-0.143
									WMR	0.500	0.426	1.000	0.000	-0.040
									PAP	0.444	0.621	0.095	0.101	0.332
									Total	0.536	0.610	0.038	0.082	0.770
P. subtriangulata **Multiplex 2**														
Psub_1	FAM	0.35	F: TCCGAACCAAGCAATAGGAC	$(CATA)_5$	50	7	4.62	148-216	WPK	0.600	0.711	0.399	0.083	0.130
			R: GGACTGACTCTGCCTTAGCG						PKR	0.600	0.684	0.738	0.004	0.230
									MAR	0.500	0.605	0.162	0.037	0.145
									WMR	0.500	0.695	0.231	0.135	-0.012
									PAP	0.400	0.600	0.109	0.068	0.789
									Total	0.520	0.667	0.200	0.094	0.261
Psub_3	PET	0.30	F: GTTGATCCACAGAGGCAGGT	$(TTC)_6$	50	4	3.58	189-198	WPK	0.500	0.537	1.000	0.004	0.227
			R: CTGCCAAATATCGCTGGTTT						PKR	0.700	0.616	0.863	0.000	-0.400
									MAR	0.600	0.605	0.635	0.000	-0.107
									WMR	0.600	0.574	1.000	0.000	0.000
									PAP	0.700	0.595	0.734	0.000	-0.410
									Total	0.620	0.583	0.998	0.000	-0.110
Psub_6	VIC	0.40	F: CCAGAAAGCCACAGGTTGTT	$(TA)_7$	50	5	3.37	224-240	WPK	0.200	0.611	0.017	0.251	0.969
			R: TTCGGTTTTGGAGGTGTAGC						PKR	0.300	0.511	0.133	0.128	0.147
									MAR	0.400	0.511	0.431	0.034	0.213
									WMR	0.700	0.753	0.900	0.000	-0.332
									PAP	0.300	0.521	0.246	0.131	0.438
									Total	0.380	0.592	0.078	0.127	0.089

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Psub_9	VIC	0.20	F: GGCTTTTCTAGTTGGTCACCC	$(CT)_7$	50	5	2.68	133-153	WPK	0.000	0.189	0.053	0.204	1.000
			R: GACCATCGACCGAGTTGTTT						PKR	0.400	0.363	1.000	0.000	0.052
									MAR	0.100	0.100	1.000	0.000	0.000
									WMR	0.300	0.353	0.307	0.000	-0.404
									PAP	0.000	0.189	0.053	0.204	1.000
									Total	0.160	0.243	0.078	0.089	0.117
Psub_10	NED	0.25	F: TAGTCCGTTTCGACTTGCCT	$(TG)_8$	50	12	5.80	181-209	WPK	0.700	0.642	0.849	0.000	-0.391
			R: GATCATCAGACAGACGTCGC						PKR	0.700	0.674	0.720	0.000	-0.067
									MAR	0.200	0.574	**0.003**	0.233	0.861
									WMR	0.300	0.511	0.053	0.131	0.025
									PAP	0.400	0.674	0.079	0.153	0.561
									Total	0.460	0.623	0.008	0.095	0.267
Psub_11	PET	0.25	F: GGCATTTACAAATCTGCGGT	$(ATCC)_5$	50	4	2.97	96-108	WPK	0.300	0.395	0.480	0.064	0.250
			R: AAATACGAATGAACGGTCGG						PKR	0.500	0.563	1.000	0.014	-0.091
									MAR	0.700	0.616	0.855	0.000	-0.410
									WMR	0.400	0.432	0.481	0.000	0.486
									PAP	0.500	0.532	1.000	0.004	-0.179
									Total	0.480	0.509	0.975	0.029	-0.035
P. australis **Multiplex 1**														
Paus_4	FAM	0.60	F: ACTCCTGCGTGCATGTATGT	$(AAT)_7$	47	7	5.68	273-300	WAI	0.333	0.712	**0.010**	0.219	0.895
			R: AAAAAGGTTTCAAATGAAGTACGA						HUI	0.400	0.821	**0.010**	0.209	0.920
									RAG	0.000	0.523	**0.001**	0.343	1.000
									LYT	0.300	0.674	0.033	0.225	0.799
									DBT	0.111	0.830	**<0.001**	0.379	0.981
									Total	0.229	0.741	**<0.001**	0.294	0.949

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Paus_5	VIC	0.50	F: CGAACGGATCCAGATTTCAG	$(ACAG)_8$	50	8	4.96	190-218	WAI	0.700	0.632	0.633	0.000	0.213
			R: AAAGGGGCTGTTGGCTATTC						HUI	0.500	0.437	1.000	0.000	0.053
									RAG	0.300	0.368	0.308	0.000	0.391
									LYT	0.600	0.605	0.817	0.000	0.184
									DBT	0.900	0.826	0.311	0.000	0.489
									Total	0.600	0.572	0.814	0.000	0.316
Paus_7	FAM	0.60	F: ACTATCATCCGCCAGAATGG	$(TA)_7$	50	8	4.86	86-106	WAI	0.400	0.442	0.219	0.000	-0.109
			R: GGCAAACCAACAAATTCAGG						HUI	0.700	0.721	0.645	0.000	0.210
									RAG	0.400	0.505	0.236	0.094	0.091
									LYT	0.400	0.674	0.079	0.153	0.425
									DBT	0.500	0.647	0.098	0.106	0.111
									Total	0.480	0.604	0.086	0.091	0.197
Paus_10	FAM	0.60	F: GGGGACCACTAGGCAATTCT	$(AAACCAAC)_5$	50	9	6.19	166-210	WAI	0.700	0.805	0.138	0.042	0.100
			R: CACGGGTTTTCGGGTGTA						HUI	0.800	0.795	1.000	0.000	0.167
									RAG	0.800	0.821	0.350	0.000	-0.219
									LYT	0.600	0.821	0.083	0.119	0.126
									DBT	0.600	0.779	0.131	0.074	0.407
									Total	0.700	0.810	0.129	0.069	0.104
Paus_11	PET	0.60	F: CAGCGCGAAATAAAGGAAAA	$(ATC)_8$	50	5	3.77	243-255	WAI	0.400	0.663	0.052	0.150	-0.059
			R: CGGTGCGAGAGTGAGTGATA						HUI	0.400	0.574	0.346	0.084	0.046
									RAG	0.500	0.716	0.044	0.106	0.200
									LYT	0.300	0.284	1.000	0.000	0.018
									DBT	0.200	0.563	**0.015**	0.210	0.579
									Total	0.360	0.571	**0.012**	0.139	0.166

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Paus_13	VIC	0.20	F: TAAGCCGGACATGCCTGTAT	$(AG)_{10}$	50	5	2.63	114-122	WAI	0.100	0.395	0.047	0.217	0.757
			R: GCGCAATCTCATCCTCTACC						HUI	0.100	0.100	1.000	0.000	0.000
									RAG	0.100	0.100	1.000	0.000	0.000
									LYT	0.600	0.468	1.000	0.000	-0.248
									DBT	0.400	0.363	1.000	0.000	-0.119
									Total	0.260	0.287	0.408	0.038	-0.044
Paus_14	NED	0.40	F: CCATCAAATTGTCGTTGGTCT	$(ACACTC)_{12}$	49	13	7.27	147-257	WAI	0.800	0.837	0.792	0.000	0.390
			R: GATCAGGTATTTCATTCTGAAGCA						HUI	0.600	0.916	**0.006**	0.165	0.506
									RAG	0.500	0.779	0.046	0.125	0.191
									LYT	0.556	0.784	0.348	0.092	-0.104
									DBT	0.200	0.805	**<0.001**	0.321	0.859
									Total	0.531	0.841	**<0.001**	0.168	0.438
***P. australis* Multiplex 2**														
Paus_1	VIC	0.25	F: CCCACGACGGACACTGTA	$(ATCC)_7$	50	7	4.71	169-193	WAI	0.600	0.626	0.642	0.000	0.022
			R: GGCATTGTCACATGAAGAGC						HUI	0.500	0.700	0.180	0.068	0.051
									RAG	0.600	0.768	0.214	0.078	0.521
									LYT	0.700	0.716	0.407	0.000	0.314
									DBT	0.600	0.663	0.392	0.033	-0.011
									Total	0.600	0.700	0.352	0.038	0.219
Paus_2	FAM	0.60	F: TGCTGAGAATCGAAACGATG	$(AG)_{15}$	50	13	7.64	232-286	WAI	0.500	0.879	**0.007**	0.187	0.319
			R: GTTGGGCTACGTCGACATTT						HUI	0.200	0.832	**<0.001**	0.327	0.362
									RAG	0.400	0.784	**0.005**	0.186	0.226
									LYT	0.300	0.747	**0.019**	0.244	0.751
									DBT	0.600	0.826	0.105	0.094	0.206
									Total	0.400	0.837	**<0.001**	0.241	0.364

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Paus_3	NED	0.45	F: TTTATGTTGCTGTTGCTGACG	$(ATC)_{25}$	49	16	8.34	135-190	WAI	0.500	0.732	0.045	0.067	-0.685
			R: ATCGCCATTATCACCATCG						HUI	0.444	0.693	0.035	0.114	0.394
									RAG	0.400	0.863	**0.005**	0.237	0.510
									LYT	0.300	0.816	**<0.001**	0.260	0.096
									DBT	0.800	0.732	0.780	0.000	0.358
									Total	0.489	0.783	**<0.001**	0.164	0.145
Paus_6	FAM	0.60	F: GCAGGATCCAGATTTTCAGC	$(ACAG)_8$	48	19	10.45	78-140	WAI	0.500	0.879	0.036	0.187	0.425
			R: TGCGGTAATTAGTCCCCTACC						HUI	0.111	0.725	**<0.001**	0.349	0.996
									RAG	0.778	0.889	0.175	0.009	0.459
									LYT	0.400	0.932	**<0.001**	0.261	0.884
									DBT	0.600	0.889	**0.006**	0.141	0.484
									Total	0.478	0.885	**<0.001**	0.230	0.626
Paus_8	PET	0.60	F: AGCTGGCATGTTGTCTCCTT	$(AAT)_{10}$	49	11	6.82	219-276	WAI	0.400	0.732	**0.016**	0.172	0.886
			R: ACATGGAAGAGTTGGGCTTG						HUI	0.000	0.889	**<0.001**	0.457	1.000
									RAG	0.300	0.774	**0.004**	0.262	0.764
									LYT	0.300	0.647	**0.008**	0.222	0.961
									DBT	0.000	0.800	**<0.001**	0.432	1.000
									Total	0.200	0.801	**<0.001**	0.336	0.945
Paus_9	PET	0.50	F: CAGTAAATCACACCCCACCA	$(AAT)_8$	47	11	6.83	112-142	WAI	0.000	0.732	**<0.001**	0.410	1.000
			R: GGGAAACCCAAATGTGTAGC						HUI	0.000	0.358	**0.003**	0.285	1.000
									RAG	0.400	0.632	0.047	0.099	-0.540
									LYT	0.444	0.882	**0.002**	0.216	0.223
									DBT	0.333	0.895	**<0.001**	0.276	0.670
									Total	0.236	0.727	**<0.001**	0.298	0.470

Table 2.2 continued.

Locus	Dye	Primer Conc. (µM)	Sequence 5' - 3'	Motif	*n*	N_a	R_a	Size Range (bp)	Location	H_O	H_E	P_{HWE}	N_F	F_{IS}
Paus_12	FAM	0.50	F: AGGCAGAAACAATCCACTCG	$(AT)_{11}$	50	8	4.36	183-201	WAI	0.700	0.574	0.519	0.000	0.126
			R: ATGTGCACCACTGGCATAAA						HUI	0.600	0.826	0.092	0.108	0.701
									RAG	0.800	0.632	0.603	0.000	0.150
									LYT	0.700	0.595	**0.013**	0.037	0.611
									DBT	0.400	0.526	0.563	0.067	0.250
									Total	0.640	0.630	0.075	0.052	0.442

2.4 Discussion

These microsatellite markers represent the first novel markers developed for *P. subtriangulata* and *P. australis*. The markers have been developed in a way that will allow for easy multiplexing and genotyping of further samples. They can be used to investigate and compare aspects of genetic population structure in these two species; for example, the degree of differentiation among populations, levels of gene flow and environmental factors that are associated with the observed patterns of genetic population structure. In addition, a number of putative SNPs have been identified that could be further tested and used in future studies if required.

Identifying loci that were unaffected by HWE deviations and null alleles was challenging. Heterozygote deficiencies were observed for most of the locations that were sampled but different patterns in HWE deviation were observed across loci, indicating that several processes could be contributing to the observed HWE deviations. For some loci most of the sampled locations deviate from HWE (i.e., *Paus_2, Paus_4, Paus_8, Paus_9*) suggesting that it is characteristics of the locus itself that are problematic. For other loci only single locations deviate from HWE (i.e., *Psub_2, Psub_7, Psub_10, Paus_11*) suggesting that population-level processes might be responsible for the observed deviations. Some processes are expected to have locus-wide effects (e.g., null alleles, genotyping error, selection), whereas others are expected to have population specific effects (e.g., non-random mating, Wahlund effect; DeWoody *et al.* 2006; Selkoe & Toonen 2006).

Loci with widespread HWE deviation could be affected by genotyping error and null alleles (DeWoody *et al.* 2006). MICROCHECKER found no evidence that genotyping error was an issue in this case. Null alleles are often observed in fecund species with high levels of early mortality, such as bivalves, as a high mutation load is expected (Hedgecock *et al.* 2004). In microsatellite loci this can result in alleles that fail to amplify due to mutations in the primer annealing sites. Alleles may also fail to amplify due to other technical problems with the PCR reaction, such as low DNA quality/quantity and preferential amplification of smaller alleles (Chapuis & Estoup 2007), the latter of which could be a problem for multiplex PCR. Estimates of null allele frequency indicate that most of the loci described here are affected by null alleles to a varying extent. Loci with widespread HWE deviation could also be influenced by selection: LOSITAN indicated that two loci (*Paus_6* and *Paus_9*) might be affected by positive selection and most locations tested for these loci are not in HWE.

Other potential causes of the observed heterozygote deficiencies include non-random mating and spatial/temporal genetic patchiness. Non-random mating is unlikely in free-spawning species with large population sizes, especially among individuals sampled from the same location. However, when unrecognised population structure exists, genetically distinct groups can be inadvertently joined into a single sampling unit, ('Wahlund effect'; Broquet & Petit 2009; Selkoe & Toonen 2006). The effect can occur spatially due to the isolation of discrete population units, and temporally due to the variable nature of larval dispersal in the marine environment (Hedgecock 1994; Hellberg 2009). It is unknown if the small number of samples genotyped for this marker development are a representative population sample so it is difficult to draw conclusions as to whether or not the observed heterozygote deficiencies are a result of unrecognised population structure.

Genetic differences among populations will be further investigated in chapters 3 and 4, and this will help to clarify the relationship between genetic population structure and heterozygote deficiency in *P. subtriangulata* and *P. australis*. However, similar to what has been reported for other bivalve species (e.g., Arias-Pérez *et al.* 2012; Benzie & Smith-Keune 2006; Kenchington *et al.* 2006), null alleles were widespread among the microsatellite markers developed for *P. subtriangulata* and *P. australis*. Further population genetic analyses will need to be interpreted with consideration of the effect that null alleles might be having on observed patterns of genetic structure and connectivity.

3 Spatially variable patterns of genetic population structure and connectivity in *Paphies subtriangulata*

3.1 Introduction

The seemingly continuous nature of the marine environment means that marine species can often inhabit large geographic ranges. To effectively manage and/or conserve a marine species we require knowledge of how individuals are distributed and move around within the species' range. For ease of interpretation, the distribution of a species is often divided into spatially discrete 'populations', forming the basic evolutionary units on which to test hypotheses (Waples & Gaggiotti 2006). A fundamental question of population genetics is how many of these 'population units' exist within a species and what is their structure, i.e., how is genetic diversity distributed across a landscape?

In practice, levels of genetic diversity are measured in terms of the allele variants observed within a population and genetic population structure determined by the way frequencies of variants change over a species range. In chapter 1 I introduced the concept of opposing evolutionary forces (i.e., drift, gene flow, mutation, selection), which should equilibrate over time to determine the geographic distribution of allelic variation within a species (Hedgecock *et al.* 2007). In an evolutionary sense, population units can be defined by significant differences in the frequency at which alleles are observed at different geographic locations. When we consider the spatial and temporal scales at which variation is observed, it is possible to make inferences about the rates and routes that connect geographically separate populations or the locations where barriers to dispersal might exist.

The type of genetic population structure that is observed in a species is dependent on the ability of individuals to disperse among geographic locations. For many marine species opportunities for dispersal are limited to the larval stage. Therefore it is the life history characteristics of the organism (effective population size, reproductive output,

pelagic larval duration, post-settlement mortality) and the physical characteristics of the ocean environment (currents, land form features, distance between suitable habitat) that determine the genetic population structure of a species (Pineda *et al.* 2007). The traditional view of the coastal marine environment as a simple linear model characterised by large scale, unstructured processes is no longer considered relevant (Selkoe *et al.* 2008). Instead physical processes in the near shore environment are stochastic, unpredictable and can change over short time scales in comparison to the offshore ocean environment (Cowen & Sponaugle 2009; Siegel *et al.* 2008). It is exactly this type of open coast environment that *Paphies subtriangulata* (tuatua) inhabits. While certain characteristics of this species (large populations, high fecundity, moderate larval duration) suggest that populations could be well connected it is unclear how physical processes will interact to modify connectivity and shape genetic population structure.

Among New Zealand's coastal marine organisms a variety of genetic population structures have been observed and patterns of connectivity inferred from this information. These patterns were summarised into five categories by Gardner *et al.* (2010): (i) no structure, (ii) isolation by distance, (iii) divergence within and/or among populations, (iv) north-south divergence, (v) east-west divergence. Often these studies have relied on simple summary statistics and analyses based on the F_{ST} fixation index. While these types of analyses are valuable and allow for comparison among studies (Neigel 2002), recent theoretical and computation advances allow us to deduce much more about population demography from the available data than just simple summary statistics (Meirmans & Hedrick 2010; Pearse & Crandall 2004).

One limitation of many of the population genetics studies reviewed by Gardner *et al.* (2010) is that they only go as far as inferring population connectivity and don't attempt to estimate rates of migration. Estimating migration can be challenging when dealing with the high gene flow situations that commonly characterise marine species. Furthermore, the models that are currently available for estimating migration are often based on unrealistic and restrictive assumptions, meaning that results must be interpreted with caution. However it has been shown that it is possible to distinguish situations where low but significant patterns of genetic patchiness have arisen because gene flow is intermittent (e.g., Knutsen *et al.* 2011; Selkoe *et al.* 2006; Stenseth *et al.* 2006). Furthermore variation in

oceanographic conditions and larval production mean that gene flow among populations is often not temporally consistent (Hedgecock 1994; Hedgecock & Pudovkin 2011; Hedrick 2005). When determining levels of gene flow, it helps to use complementary methods that will provide estimates over different time scales. It is important to quantify the rate and direction of migration among localities, and how it varies over time, if we want to gain an understanding of the ecological importance of connections within a species. Ideally we want to build up a picture of the types of genetic population structure and connectivity that can occur within a region across multiple species. This approach allows for a more integrative, eco-system wide understanding of genetic population structure and connectivity, which is more effective from a management perspective than working on a per-species basis.

The purpose of this chapter was to determine genetic population structure and estimate rates of migration for *P. subtriangulata*, testing the null hypothesis of panmixia (i.e., no genetic population structure) and the high rates of migration implied by such structure. A previous study of genetic structure in *P. subtriangulata* using allozyme markers suggested north-south divergence of populations (Smith *et al.* 1989). However, allozymes often do not have the power to detect fine scale patterns of genetic structuring, so re-examination using higher resolution microsatellite markers is required. Comparison of the genetic population structure observed between the two studies may also provide an indication as to whether this structure has been stable over time. Additionally, patterns of migration among populations have not been previously investigated in *P. subtriangulata*. In this chapter Bayesian frameworks were used to estimate levels of migration among the sampled localities, something that has been done for few other coastal marine organisms in New Zealand (but see Wei *et al.* (2013a) for one example). This study allows for a comprehensive understanding of genetic diversity, structure and connectivity in *P. subtriangulata*, and will form the basis for further multi-disciplinary and multi-species investigations into environmental drivers of genetic population structure and connectivity.

The panel of microsatellite markers described in chapter 2 were used to gather genetic data for *P. subtriangulata* from ten locations, to specifically:

1. Quantify levels of genetic diversity in *P. subtriangulata* sampled from ten locations;

2. Investigate levels of differentiation among the sampled locations to determine which of the five categories from Gardner *et al.* (2010) best describes the genetic population structure of *P. subtriangulata*;
3. Estimate the rate and direction of migration among *P. subtriangulata* locations over both recent and long-term time scales.

3.2 Methods

3.2.1 Sample collection and laboratory methods

Samples of *Paphies subtriangulata* were collected between June 2010 and January 2013 from ten locations (Figure 3.1; Table 3.1). The biogeographic classification scheme of Shears *et al.* (2008) and the New Zealand fishery quota management areas were used as a basis for selecting sampling sites. The aim was to sample at least one site within each biogeographic/fishery area, dependent on being able to locate suitable populations for sampling within those areas. Samples were gathered by hand at low tide, photographed, shell length was measured to the nearest 0.5 cm and a sub-sample of foot tissue was taken and stored in 80% ethanol at 4°C.

For each sample a piece of tissue approximately 25 mm^2 was used for DNA extraction, using one of the following methods:

1. A standard proteinase K digestion followed by phenol-chloroform purification and ethanol precipitation;
2. A Qiagen DNeasy Blood and Tissue extraction kit, following the manufacturer's protocols;
3. A Geneaid Genomic DNA Mini Kit, following the manufacturer's protocols;
4. A Zygem prepGEM extraction kit, following the manufacturer's protocols.

DNA extracts were quantified using an Implen Nanophotometer and diluted in ddH_2O to a concentration of approximately 100 ng/µl for use in a PCR reaction.

Eleven *P. subtriangulata* microsatellite loci were amplified in multiplex polymerase chain reaction (PCR) following the protocols listed in chapter 2. PCR products were size separated and the fluorescently labelled amplicons detected using an ABI3730XL (Life Technologies™) automated capillary sequencer. Allele sizes were determined using Peak

Scanner 1.0 (Life Technologies™) and binned using Autobin (http://www4.bordeaux-aquitaine.inra.fr/biogeco/Media/Ressources/Logiciels/Autobin). For each plate of 96 samples genotyped, four previously genotyped samples were included as a positive control and to estimate levels of genotyping error. The mean error rate per allele was calculated following Pompanon *et al.* (2005).

3.2.2 Data quality checks and genetic diversity analyses

Genotype data was checked for scoring errors and large allele drop out using MICROCHECKER 2.2.3 (van Oosterhout *et al.* 2006). Loci were checked for linkage disequilibrium and conformance to Hardy-Weinberg equilibrium (HWE) using GENEPOP 4.2 (Rousset 2008; Markov-chain parameters: 10,000 dememorization steps, 1,000 batches and 10,000 iterations per batch). Comparisons were considered significant if $p < 0.05$ after false discovery rate (FDR) correction for multiple tests (Verhoeven *et al.* 2005). To determine if sufficient allelic variation had been sampled allele discovery curves were calculated using the 'jackmsatpop' function in the PopGenKit package (Rousset 2008) implemented in R (R core team 2012). Jackknife resampling was used to calculate the number of sampled alleles for a given constant increase in sample size; 1000 replicates with an interval of 1 were used.

LOSITAN (Antao *et al.* 2008) was used to detect outlier loci (50,000 simulations using a stepwise mutation model). The F_{ST} outlier method is based on the relationship between F_{ST} and expected heterozygosity (H_E) to identify loci that have excessively high or low F_{ST} compared to neutral expectations. A 95% confidence interval was used to determine which loci were outliers. The number of alleles, allelic range and allele frequencies for each locus were calculated in GENEPOP 4.2 (Rousset 2008). The proportion of null alleles at each locus was estimated using the algorithm of Dempster *et al.* (1977) as implemented in FreeNA (Chapuis & Estoup 2007) using 1,000 bootstrap replicates.

For each population the mean number of alleles (N_a) and allelic richness (R_a) were calculated using FSTAT 2.9.3.2 (Goudet 1995). Allelic richness is a measure of allele diversity corrected for the smallest sample size (n = 40 for *P. subtriangulata*). Observed (H_O) and expected (H_E) heterozygosity were calculated using ARLEQUIN 3.5 (Excoffier & Lischer 2010). The inbreeding co-efficient F_{IS} was calculated using GENEPOP 4.2 (Rousset 2008). Private alleles for each location were calculated by hand.

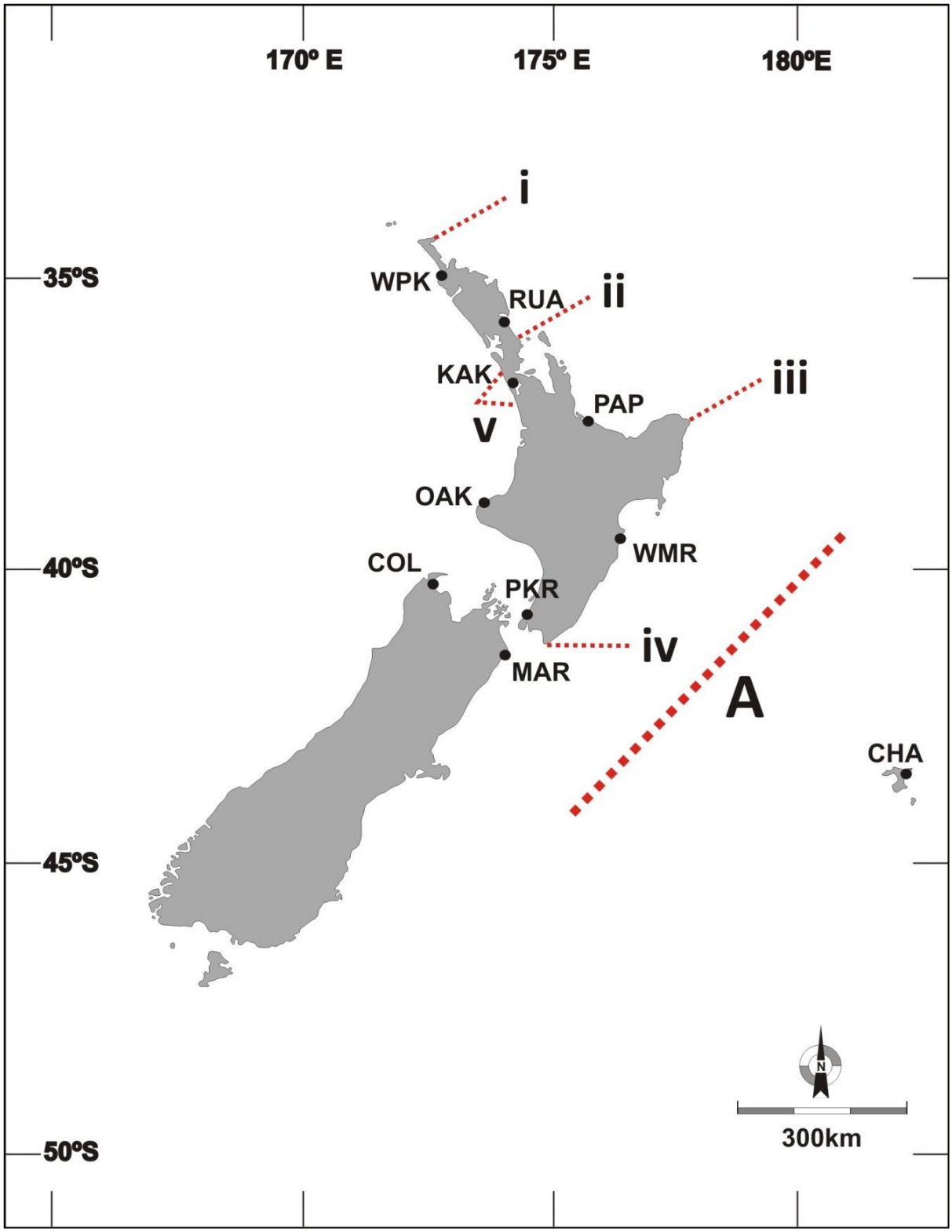


Figure 3.1: Collection locations for *Paphies subtriangulata.* Location abbreviations refer to those listed in Table 3.1. Red dashed lines represent the location of putative genetic breaks among populations detected by the present study. A major genetic break was detected between Chatham Island and mainland New Zealand locations (A). Five other possible locations where gene flow is limited among *P. subtriangulata* populations are indicated by the thinner dashed lines (i-v).

Table 3.1: Location, number of samples and geographical co-ordinates for *Paphies subtriangulata* samples used in this study.

Location	Abbreviation	Sample Size	Geographical Co-ordinates	
Ruakaka	RUA	53	-35.91° S	174.46° E
Papamoa	PAP	51	-37.70° S	176.29° E
Waimarama	WMR	51	-39.82° S	177.00° E
Marfell Beach	MAR	56	-41.72° S	174.20° E
Paekakariki	PKR	55	-40.99° S	174.95° E
Collingwood	COL	52	-40.68° S	172.69° E
Oakura	OAK	40	-39.12° S	173.95° E
Kakamatua	KAK	57	-37.01° S	174.60° E
Waipapakauri	WPK	54	-35.04° S	173.17° E
Chatham Island	CHA	48	-43.80° S	-176.35° W

3.2.3 Genetic population differentiation analyses

All population differentiation analyses were first performed using all loci, then were repeated excluding loci identified as outliers by LOSITAN. This was to assess the impact that these outlier loci were having on the patterns of genetic differentiation that were observed. Global F_{ST} was calculated in FreeNA (Chapuis & Estoup 2007) both with and without correction for null alleles to determine the impact that null alleles might have on population differentiation estimation. Pairwise F_{ST} was calculated using GENEPOP 4.2 (Rousset 2008). An exact test of population differentiation using the G log likelihood ratio test (Goudet *et al.* 1996) was performed in GENEPOP 4.2 (Markov-chain parameters: 10,000 dememorization steps, 1,000 batches and 10,000 iterations per batch). Comparisons were considered significant if $p < 0.05$ after implementing the FDR correction for multiple tests (Verhoeven *et al.* 2005). Modified pairwise phi-statistics (ϕ'_{ST}) were calculated in GENODIVE 2.0b23 (Meirmans & van Tienderen 2004). The ϕ'_{ST} index, based on the AMOVA (analysis of molecular variance) framework (Excoffier *et al.* 1992), is designed for use with multi-allelic data and is standardised by the maximum possible ϕ_{ST} value. This should allow for more accurate estimation of population differentiation than F_{ST} and standardisation allows for better comparison among species. The significance of pairwise ϕ'_{ST} values was assessed

using the AMOVA ϕ_{ST} test for population differentiation with 1,000 permutations, implemented in GENODIVE 2.0b23 (Meirmans & van Tienderen 2004). Comparisons were considered significant if $p < 0.05$ after FDR correction.

Tests were conducted to determine if the patterns of genetic population structure observed for other New Zealand coastal marine species, as reported by Gardner *et al.* (2010), were present for *P. subtriangulata* populations (i.e., no structure, IBD, divergence within and/or among populations, north-south divergence, east-west divergence). Mantel tests were used to determine if an IBD pattern was present, as implemented in FSTAT 2.9.3.2 (Goudet 1995). Correlation between measures of genetic differentiation (F_{ST}) and geographic distance (km) was assessed using 10,000 randomisations and was considered significant if $p < 0.05$. F_{ST} was calculated in two different ways: (1) $F_{ST}/1\text{-}F_{ST}$ calculated following Weir and Cockerham (1984); and (2) $F_{ST}/1\text{-}F_{ST}$ calculated using the ENA correction method to account for null alleles, as described in Chapuis and Estoup (2007). The shortest distance between sampling sites was determined using Google Maps. IBD tests were conducted both including and excluding the Chatham Island population.

Patterns of genetic differentiation among locations were visualised using a principal component analysis (PCA) implemented in PCA-Gen 1.2 (http://www2.unil.ch/popgen/softwares/pcagen.htm). This program determines the inertia of each axis and tests for significance of each axis. One thousand randomisations were used and axes were considered significant if $p < 0.05$. AMOVA tests were used to determine if the groupings observed in PCA analyses were significant and to test for north-south/east-west differentiation. AMOVA tests were conducted in ARLEQUIN 3.5 (Excoffier & Lischer 2010) with 10,000 permutations (significant if $p < 0.05$ after FDR correction).

The genetic structure of *P. subtriangulata* locations was assessed using two clustering methods: STRUCTURE 2.3.4 (Pritchard *et al.* 2000) and AWclust (Gao & Starmer 2008). STRUCTURE is a Bayesian clustering algorithm that attempts to assign each individual into *K* number of clusters to minimise departures from HWE and gametic disequilibrium. Although the data violates some of the assumptions of this analysis (i.e., departure from HWE), by comparing STRUCTURE results to those obtained from AWclust (which does not assume HWE) it is possible to determine the sensitivity of this analysis to these violations.

Ten independent STRUCTURE runs were performed for values of *K* ranging from 1 to 10, using 100,000 iterations and a burn-in length of 10,000. The 'admixture' model and 'correlated allele frequencies' options were used (Falush *et al.* 2003), and sampling locations were used as a prior as this is known to improve the performance of the program when genetic population structure is weak (Hubisz *et al.* 2009). The Evanno method (Evanno *et al.* 2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2011) was used to determine the value of *K* that best fit the data. This method is based on the rate of change in the log probability of data between successive *K* values, i.e., the most likely *K* value precedes the greatest rate of decline in the log probability. The optimal arrangement of clusters from the 10 replicates was determined in CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) using the 'Greedy' algorithm with 1,000 random input orders. Genetic population structure was visualised using DISTRUCT 1.2 (Rosenberg 2003).

The non-parametric clustering algorithm AWclust was used to determine genetic population structure without relying on the assumptions of HWE and linkage disequilibrium (Gao & Starmer 2008). Microsatellite data was converted into a presence/absence matrix where each allele was treated as a locus and allocated a 0, 1 or 2 to indicate the number of copies of that allele for each individual. The AWclust package was implemented in R (R core team 2012) to firstly calculate an allele sharing distance matrix then assign each individual to one cluster. A gap statistic that compares the pooled within-cluster sum of squares with expectations from a null reference distribution was calculated for values of *K* ranging from 1 to 8 (maximum range of *K* allowed by AWclust) using 100 null simulations. The *K* value with the largest deviation from the expected distribution was determined to be the optimal value (Gao & Starmer 2008). Genetic population structure was visualised using DISTRUCT 1.2 (Rosenberg 2003). Both clustering analyses were first conducted including all sampling locations then repeated excluding the Chatham Island population to examine genetic population structure among mainland New Zealand locations only. Spearman rank correlation tests were used to determine if there were any significant latitudinal or longitudinal gradients in the distribution of clusters identified by STRUCTURE and AWclust ($p < 0.05$ after FDR correction).

3.2.4 Estimation of migration rates

Contemporary migration among locations was estimated using the 'detection of first generation migrants' option as implemented in GENECLASS2 (Piry *et al.* 2004). This method detects individuals in the current generation (F0) that are in genotypic disequilibrium relative to their sampled population and are therefore likely to have originated from a population other than where they were sampled. The test statistic *Lh/Lhmax* (ratio of the likelihood of drawing an individual's genotype from the population in which it was sampled to the maximum likelihood observed for this genotype in any population; Paetkau *et al.* 2004) and the Bayesian approach of Rannala and Mountain (1997) were used to detect migrants. The probability of an individual's multilocus genotype originating from each location was calculated using the Markov chain Monte Carlo (MCMC) resampling method of Paetkau *et al.* (2004). Ten thousand individuals were simulated and a probability detection threshold of $\alpha = 0.01$ was used, below which an individual was assigned as a migrant.

Contemporary migration among locations was also estimated using the Bayesian assignment approach implemented in BAYESASS 3.0.3 (Wilson & Rannala 2003). BAYESASS is similar to GENECLASS2 in that it detects the temporary genetic disequilibrium that recent migrant genotypes are expected to show relative to their sampled population and uses this information to infer rates of migration among populations. The approaches differ in that GENECLASS2 can only detect F0 migrants, whereas BAYESASS detects migrants within the past few generations, typically equivalent to the number of cohorts present in the population (approximately five years for *P. subtriangulata*). Approximately 10 runs of BAYESASS were conducted to determine the number of iterations required for the MCMC chain to converge and appropriate values for the mixing parameters Δa, Δf and Δm, which determine rate at which parameters are accepted for each iteration. The optimal acceptance rate should be in the range of 40-60%, but this may not be possible when the log-likelihood surface is relatively flat, which can be the case when population structure is weak (Wilson & Rannala 2003). However, as long as the MCMC chain was observed to converge then it can be assumed that sufficient mixing has occurred (Wilson & Rannala 2003). Final results were based on the average of three independent runs using 1×10^7 iterations, the first 1×10^6 iterations being discarded as burn-in. Samples were taken every 500 iterations. The mixing parameters used for each run were Δa 0.8, Δf 0.8 and Δm 0.6.

Long-term migration rates among locations were estimated using MIGRATE 3.5.1 (Beerli 2006). MIGRATE estimates two parameters: the population size parameter θ, which is equivalent to four times the effective population size scaled by mutation rate ($4N_e\mu$) and a mutation scaled migration rate M ($m\mu$). A Bayesian MCMC strategy was used with a static heating scheme (temperatures: 1.0, 1.5, 3.0, 100,000; swapping interval 1). Uniform prior distributions were used for θ and M. Preliminary runs determined that the best prior intervals to use were 0-10 for θ and 0-10,000 for M. Three replicates were run for 1×10^5 iterations with a sample taken every 100 steps, resulting in a total of 30,000 recorded steps over the three replicate runs. The first 10,000 steps of each replicate were discarded as burn-in. Values of θ, M and their credible intervals (CI) were averaged over the three runs. Chain convergence was assessed by comparing the correlation between parameter estimates between independent runs. θ and M were multiplied to give the effective number of migrants per generation (N_em) between each pair of sampling locations. The effective size of each population (N_e) can also be estimated from θ if the mutation rate of the markers used is known. Mutation rate per generation for microsatellite loci is estimated to be in the range of 1×10^{-4} to 1×10^{-3} (Ellegren 2000, 2004).

Based on the results of genetic population differentiation analyses, all three migration estimation methods were carried out excluding outlier loci. In some cases the data violated some of the assumptions of the analyses (e.g., HWE, F_{ST}). However by comparing results across multiple types of analyses it was possible to determine the level of sensitivity to these violations. Because analyses that estimated migration among all population pairs were unlikely to have high enough F_{ST} to detect significant levels of migration the analyses were repeated to estimate migration among population groupings identified as significantly differentiated by exact G and AMOVA tests.

3.3 Results

3.3.1 Data quality checks and summary statistics

A total of 517 *Paphies subtriangulata* individuals from ten locations were genotyped (Figure 3.1; Table 3.1). Despite reports that *P. subtriangulata* is distributed throughout the South Island (Powell 1979), it was not possible to find this species further south than -42° S on the New Zealand mainland in the present study. Therefore the majority of *P.*

subtriangulata locations sampled in this study were from the North Island except for two locations at the top of the South Island (Collingwood and Marfell Beach) and one Chatham Island location, approximately 700 km east of mainland New Zealand.

Thirty eight *P. subtriangulata* samples (7.4% of samples) were repeat genotyped and the mean rate of genotyping error was calculated at 4.8%. The majority of error (63%) was associated with allele drop out, rather than mistyping of alleles. MICROCHECKER found no evidence for genotype scoring errors or large allele drop out and there was no significant linkage disequilibrium among loci. Only two loci were in HWE for the total sample (*Psub_3* and *Psub_4*; Table 3.2). Two loci (*Psub_5* and *Psub_7*) were not in HWE for any of the locations tested and LOSITAN detected one outlier locus (*Psub_2*; Figure 3.2). Although this locus was a statistically significant outlier, removal of the locus from analyses had little effect on estimates of genetic population structure. However analyses both including and excluding this locus are presented for completeness. Allele frequency graphs and discovery curves can be seen in appendices 2 and 3. For all loci allele discovery curves look to have reached or are approaching an asymptote, suggesting that apart from a few rare alleles, most of the allelic diversity has been sampled for this species. The number of alleles per locus ranged from 5-28 (Table 3.3). For most loci observed heterozygosity was lower than expected, but for *Psub_3* observed heterozygosity was similar to expected (Range H_O: 0.105-0.551; H_E: 0.121-0.791; Table 3.3). The frequency of null alleles ranged from 0.035-0.278 and F_{IS} ranged from 0.007-0.610 (Table 3.3).

The mean number of alleles ranged from 6-8.091, allelic richness ranged from 5.290-6.566 and private alleles ranged from 0-13 (Table 3.4). The frequency of private alleles was generally low (<2%; Appendix 1); i.e., these alleles can be considered ‘rare’. The Chatham Island population had a much larger number of private alleles than other *P. subtriangulata* populations and some of these alleles were observed in high frequency (up to 14%; Appendix 1). Observed heterozygosity was lower than expected for all locations (Range H_O: 0.298-0.399; H_E: 0.521-0.588) and no locations were in HWE (Table 3.4). F_{IS} was above zero for all locations due to an excess of homozygotes (Range: 0.242-0.442; Table 3.4).

Table 3.2: Hardy-Weinberg equilibrium (HWE) *p*-values for each locus and location for *Paphies subtriangulata*. Values in bold represent significant departures from HWE expectation after false discovery rate (FDR) correction for multiple tests ($p < 0.05$). Location abbreviations as per Table 3.1.

Locus	Location										
	RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	Total
Psub_1	**<0.001**	**0.001**	**0.008**	**0.001**	0.136	**0.019**	0.036	0.095	0.329	**<0.001**	**<0.001**
Psub_2	**0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.616	**<0.001**
Psub_3	0.313	0.163	0.914	0.841	0.208	0.043	0.800	0.914	0.355	0.657	0.493
Psub_4	1.000	0.147	1.000	0.093	1.000	**0.001**	1.000	1.000	0.116	1.000	0.060
Psub_5	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.020**	**<0.001**
Psub_6	**<0.001**	**<0.001**	**<0.001**	**0.001**	**<0.001**	**<0.001**	0.077	**0.002**	**<0.001**	**<0.001**	**<0.001**
Psub_7	**<0.001**	**0.009**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Psub_8	**<0.001**	**<0.001**	0.078	**<0.001**	0.032	**<0.001**	**0.012**	0.271	**<0.001**	0.557	**<0.001**
Psub_9	0.063	**0.003**	0.043	0.158	1.000	0.143	0.093	0.135	**<0.001**	0.223	**<0.001**
Psub_10	**<0.001**	**0.006**	**0.009**	**<0.001**	0.267	**<0.001**	**<0.001**	**<0.001**	0.052	1.000	**<0.001**
Psub_11	0.614	**0.022**	**0.015**	0.062	0.211	0.123	0.269	0.645	0.056	1.000	**0.003**

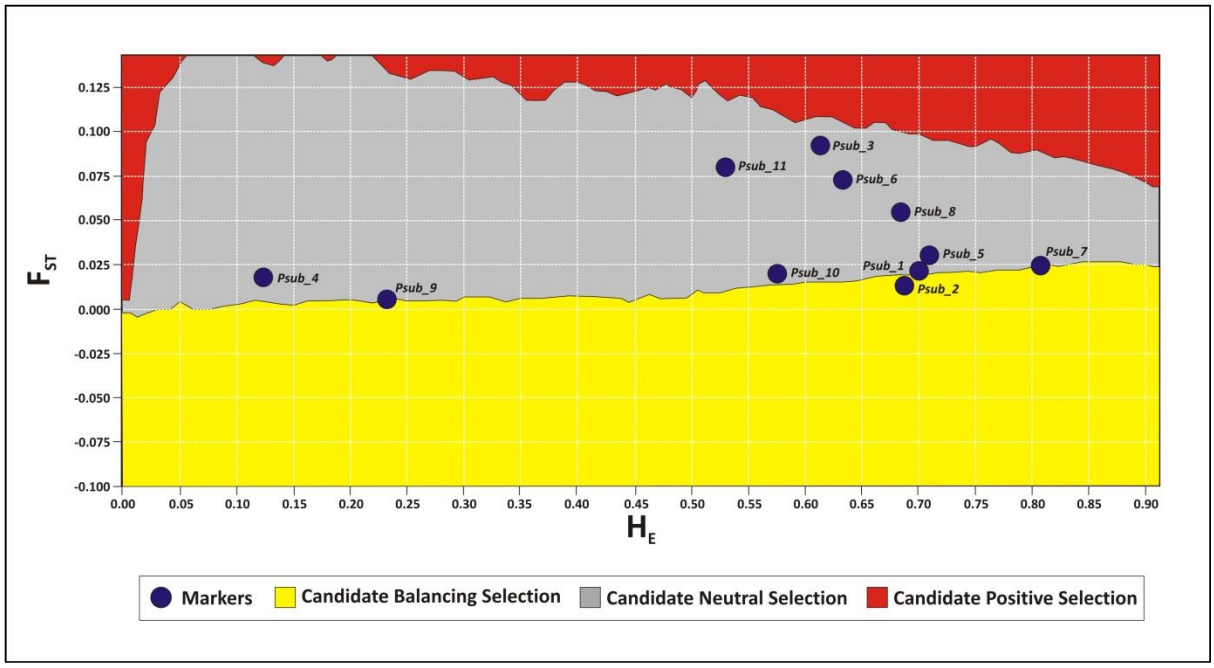


Figure 3.2: Output of LOSITAN analyses for outlier loci for *Paphies subtriangulata*. Blue dots represent the relationship between F_{ST} and H_E for each locus. Loci falling within the grey area represent 'neutral' loci, loci falling within the red or yellow areas represent F_{ST} outliers as determined by a 95% confidence interval.

Table 3.3: Genetic diversity statistics for *Paphies subtriangulata* loci used in this study (N_a: number of alleles, H_O: observed heterozygosity, H_E: expected heterozygosity, N_F: estimated frequency of null alleles, F_{IS}: inbreeding co-efficient).

Locus	Statistic					
	N_a	Allele Size Range (bp)	H_O	H_E	N_F	F_{IS}
Psub_1	20	140-244	0.542	0.683	0.101	0.209
Psub_2	28	107-148	0.405	0.680	0.173	0.411
Psub_3	6	186-201	0.551	0.556	0.035	0.007
Psub_4	7	204-222	0.105	0.121	0.040	0.132
Psub_5	12	89-113	0.269	0.693	0.257	0.610
Psub_6	13	212-250	0.327	0.590	0.190	0.446
Psub_7	23	103-247	0.303	0.791	0.278	0.610
Psub_8	12	202-248	0.402	0.649	0.160	0.338
Psub_9	12	133-155	0.197	0.231	0.046	0.149
Psub_10	21	171-235	0.405	0.565	0.119	0.283
Psub_11	5	96-112	0.404	0.518	0.065	0.168

Table 3.4: Genetic diversity statistics for each location and total for *Paphies subtriangulata* (N_a: mean number of alleles, R_a: allelic richness, PA: private alleles, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{IS}: inbreeding co-efficient). F_{IS} values in bold show significant departure from HWE expectations after FDR correction ($p < 0.05$). Location abbreviations as per Table 3.1.

Location	Statistic					
	N_a	R_a	PA	H_O	H_E	F_{IS}
RUA	7.818	6.489	5	0.353	0.588	**0.388**
PAP	7.091	5.928	4	0.358	0.559	**0.352**
WMR	7.182	6.044	2	0.352	0.553	**0.356**
MAR	7.636	5.797	5	0.322	0.541	**0.405**
PKR	7.636	5.647	3	0.399	0.550	**0.275**
COL	7.000	5.797	2	0.298	0.543	**0.442**
OAK	6.273	5.485	0	0.368	0.542	**0.318**
KAK	8.091	6.566	3	0.394	0.569	**0.301**
WPK	7.364	5.943	4	0.357	0.558	**0.339**
CHA	6.000	5.290	13	0.390	0.521	**0.242**
Total	14.455	6.542	40	0.355	0.553	**0.348**

3.3.2 Genetic population differentiation

There was no clear relationship between change in global F_{ST} and proportion of null alleles as correcting for null alleles increased F_{ST} for some loci and decreased F_{ST} for other loci. Global F_{ST} ranged from 0.004-0.090 for *P. subtriangulata* loci and 0.008-0.087 when corrected for null alleles (Figure 3.3). Pairwise F_{ST} among *P. subtriangulata* populations ranged from 0-0.140 when calculated using all loci and 0-0.153 when the outlier locus *Psub_2* was removed from the analysis (Table 3.5). F_{ST} values were generally higher with the outlier locus excluded, suggesting that this locus was masking population differentiation. The highest F_{ST} values were between Chatham Island and mainland New Zealand locations, and the Chatham Island population was found to be significantly differentiated from all other locations. Among mainland locations, Ruakaka and Waimarama were significantly differentiated from all other locations, and Kakamatua and Waipapakauri were differentiated from most locations. Pairwise ϕ'_{ST} showed a similar pattern to F_{ST} with significant differentiation of the Chatham Island and Ruakaka populations from all other

locations. Values ranged from 0-0.279 when calculated using all loci and 0-0.289 when the outlier locus *Psub_2* was removed from the analysis (Table 3.6).

Mantel tests using all sampling locations revealed a significant pattern of IBD with both measures of F_{ST} (*p*-value range 0.0001-0.0003; Figure 3.4A-D). When the Chatham Island population was excluded the IBD pattern among mainland locations was weaker; a significant pattern was only detected using F_{ST} corrected for null alleles (*p*-value range 0.0388-0.0835; Figure 3.5A-D). A PCA using all loci showed that the X axis explains 38.41% of variation (p = 0.001) and the Y axis 21.81% of variation (p = 0.024; Figure 3.6A). The Chatham Island and Ruakaka populations group separately from the remainder of the locations. When the locus *Psub_2* is removed from the analysis a similar pattern of population grouping is shown with the X axis explaining 40.20% of variation (p = 0.001) and the Y axis 22.69% of variation (p = 0.021; Figure 3.6B).

An AMOVA (Table 3.7) of all sampling locations using all loci showed that 63.2% of variation was partitioned among individuals and 33.6% of variation among individuals within populations. Exclusion of the *Psub_2* outlier locus had little effect, with 63.8% of variation partitioned among individuals and 32.7% of variation partitioned among individuals within populations. There were no significant north-south (All loci: {RUA, PAP, WMR, KAK, WPK} vs {MAR, PKR, COL, OAK}, p = 0.052; non-outlier loci {RUA, PAP, WMR, KAK, WPK} vs {MAR, PKR, COL, OAK}, p = 0.055) or east-west groupings (All loci: {RUA, PAP, WMR, MAR} vs {WPK, KAK, OAK, PKR, COL}, p = 0.466; non-outlier loci: {RUA, PAP, WMR, MAR} vs {WPK, KAK, OAK, PKR, COL}, p = 0.490). The arrangement of populations that explained the most variation among groups was {CHA} {RUA} {WMR} {KAK} {PAP, MAR, PKR, COL, OAK, WPK} (p = 0.001 both including and excluding *Psub_2*).

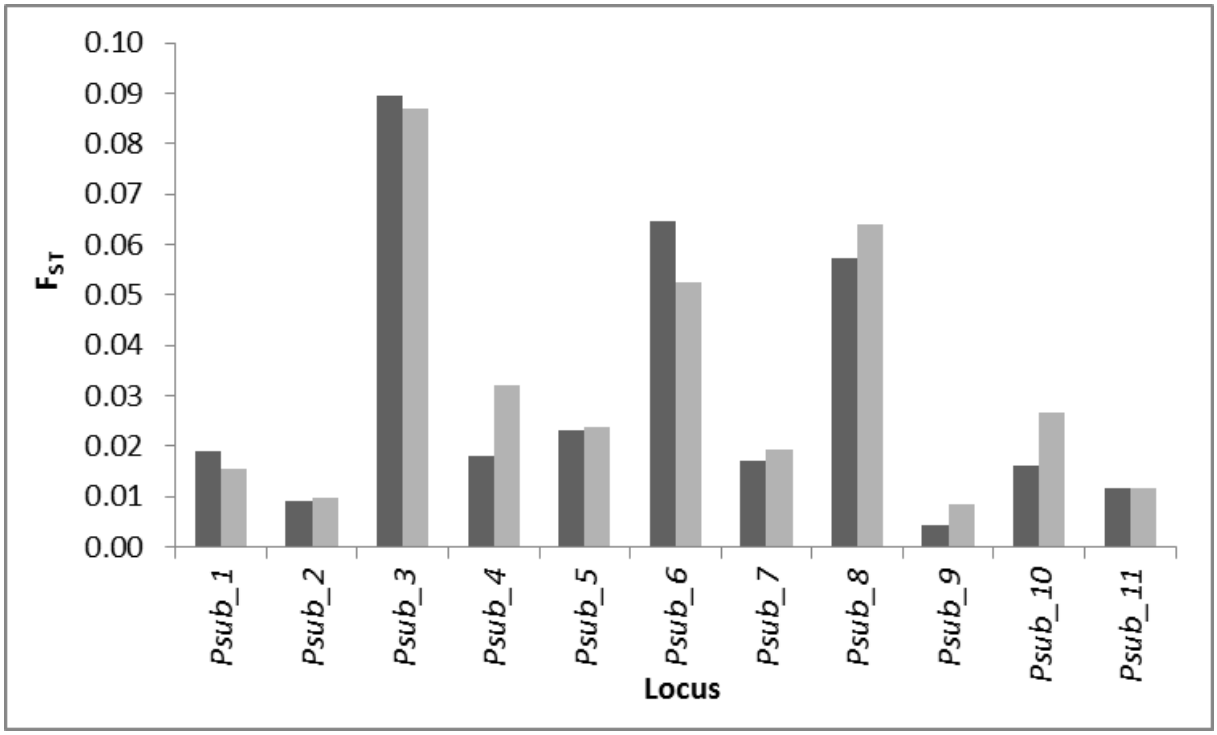


Figure 3.3: Global F_{ST} for each locus for *Paphies subtriangulata*. Dark grey bars represent F_{ST} uncorrected for null alleles, light grey bars represent F_{ST} corrected for null alleles using the algorithm described by Dempster *et al.* (1977).

Table 3.5: Pairwise F_{ST} values among *Paphies subtriangulata* locations using all loci (lower diagonal) and excluding the outlier locus *Psub_2* (upper diagonal). Values in bold represent significant differentiation after FDR correction ($p < 0.05$) as assessed by an exact G test (Goudet *et al.* 1996). Location abbreviations as per Table 3.1.

	RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA
RUA	-	**0.038**	**0.021**	**0.041**	**0.031**	**0.046**	**0.045**	**0.035**	**0.023**	**0.074**
PAP	**0.034**	-	**0.031**	0.000	0.006	0.000	0.005	**0.021**	**0.012**	**0.129**
WMR	**0.019**	**0.027**	-	**0.028**	**0.012**	**0.025**	**0.014**	**0.003**	**0.008**	**0.108**
MAR	**0.036**	0.000	**0.024**	-	0.003	0.000	0.002	**0.014**	**0.007**	**0.133**
PKR	**0.027**	0.005	**0.010**	0.002	-	**0.004**	0.000	0.004	**0.005**	**0.117**
COL	**0.040**	0.000	**0.021**	0.000	**0.003**	-	0.003	**0.014**	0.006	**0.144**
OAK	**0.040**	0.004	**0.012**	0.001	0.000	0.002	-	**0.005**	**0.007**	**0.153**
KAK	**0.031**	**0.018**	**0.002**	**0.011**	0.002	**0.011**	0.003	-	**0.004**	**0.130**
WPK	**0.023**	**0.015**	**0.010**	**0.009**	**0.007**	0.008	**0.007**	**0.006**	-	**0.116**
CHA	**0.069**	**0.114**	**0.099**	**0.121**	**0.105**	**0.130**	**0.140**	**0.117**	**0.116**	-

Table 3.6: Pairwise ϕ'_{ST} values among *Paphies subtriangulata* locations using all loci (lower diagonal) and excluding the outlier locus *Psub_2* (upper diagonal). Values in bold represent significant differentiation after FDR correction ($p < 0.05$) as assessed by an AMOVA ϕ'_{ST} test for population differentiation. Location abbreviations as per Table 3.1.

	RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA
RUA	-	**0.060**	**0.033**	**0.067**	**0.057**	**0.046**	**0.076**	**0.068**	**0.027**	**0.141**
PAP	**0.056**	-	**0.058**	0.000	0.007	0.000	0.000	**0.041**	**0.009**	**0.242**
WMR	**0.030**	**0.053**	-	**0.051**	**0.020**	**0.035**	**0.019**	0.000	**0.005**	**0.206**
MAR	**0.062**	0.000	**0.046**	-	0.002	0.000	0.000	**0.026**	0.001	**0.248**
PKR	**0.052**	0.006	**0.017**	0.000	-	0.000	0.000	0.005	0.002	**0.229**
COL	**0.042**	0.000	**0.030**	0.000	0.000	-	0.000	**0.020**	0.000	**0.238**
OAK	**0.069**	0.000	**0.017**	0.000	0.000	0.000	-	0.004	0.000	**0.289**
KAK	**0.061**	**0.035**	0.000	**0.022**	0.002	**0.017**	0.002	-	0.000	**0.259**
WPK	**0.028**	**0.017**	**0.009**	**0.005**	**0.007**	0.000	0.000	**0.005**	-	**0.217**
CHA	**0.139**	**0.228**	**0.199**	**0.238**	**0.218**	**0.230**	**0.279**	**0.245**	**0.225**	-

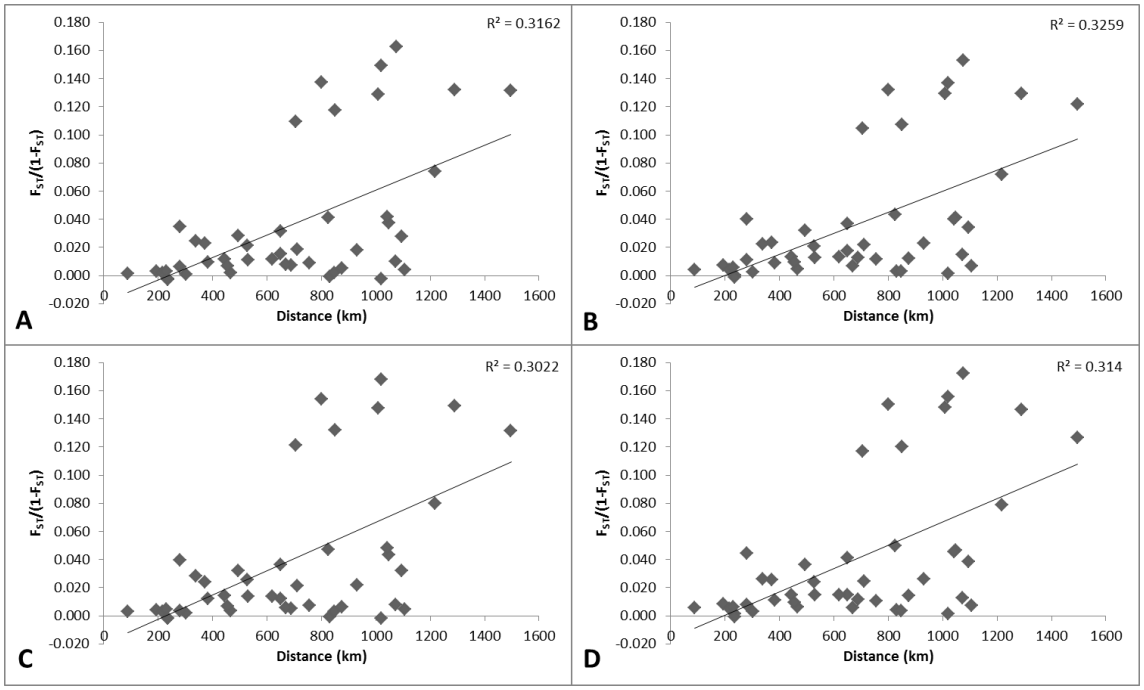


Figure 3.4: Results of Mantel tests for IBD for all *Paphies subtriangulata* locations (significant if $p < 0.05$). A: based on F_{ST} for all loci ($p = 0.0001$); B: based on F_{ST} corrected for null alleles using the ENA method described in Chapuis and Estoup (2007) for all loci ($p = 0.0001$); C: based on F_{ST} for non-outlier loci ($p = 0.0002$); D: based on ENA corrected F_{ST} for non-outlier loci ($p = 0.0003$).

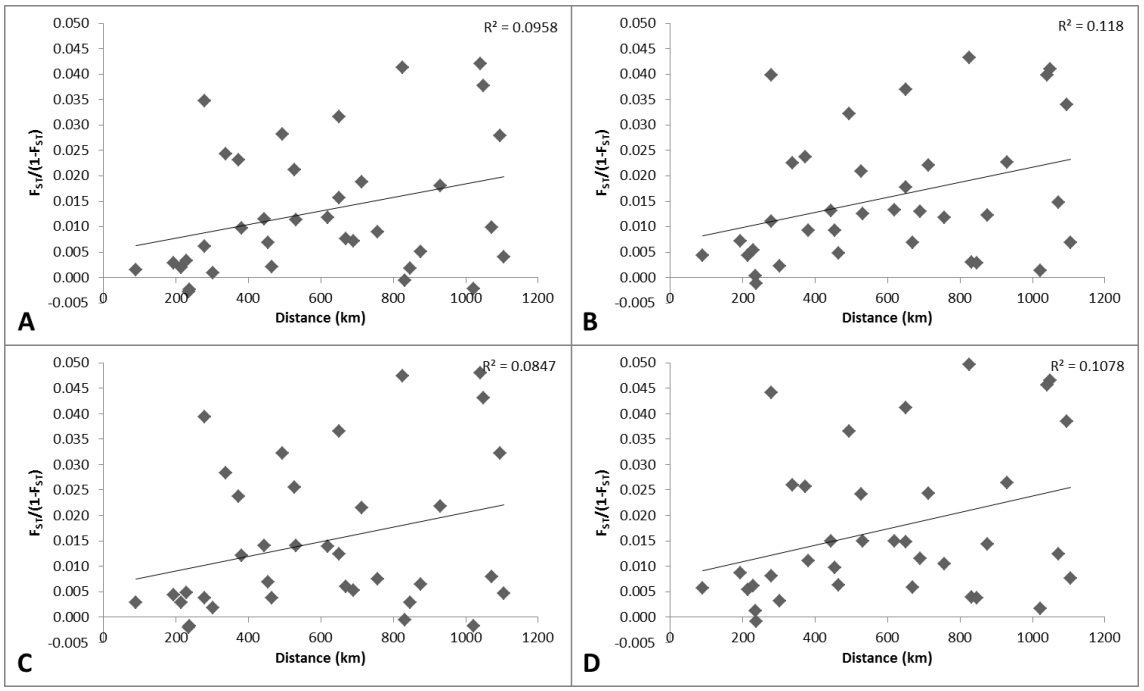


Figure 3.5: Results of Mantel tests for IBD for mainland *Paphies subtriangulata* locations, excluding the Chatham Island population (significant if $p < 0.05$). A: based on F_{ST} for all loci ($p = 0.0640$); B: based on F_{ST} corrected for null alleles using the ENA method described in Chapuis and Estoup (2007) for all loci ($p = 0.0388$); C: based on F_{ST} for non-outlier loci ($p = 0.0835$); D: based on ENA corrected F_{ST} for non-outlier loci ($p = 0.0488$).

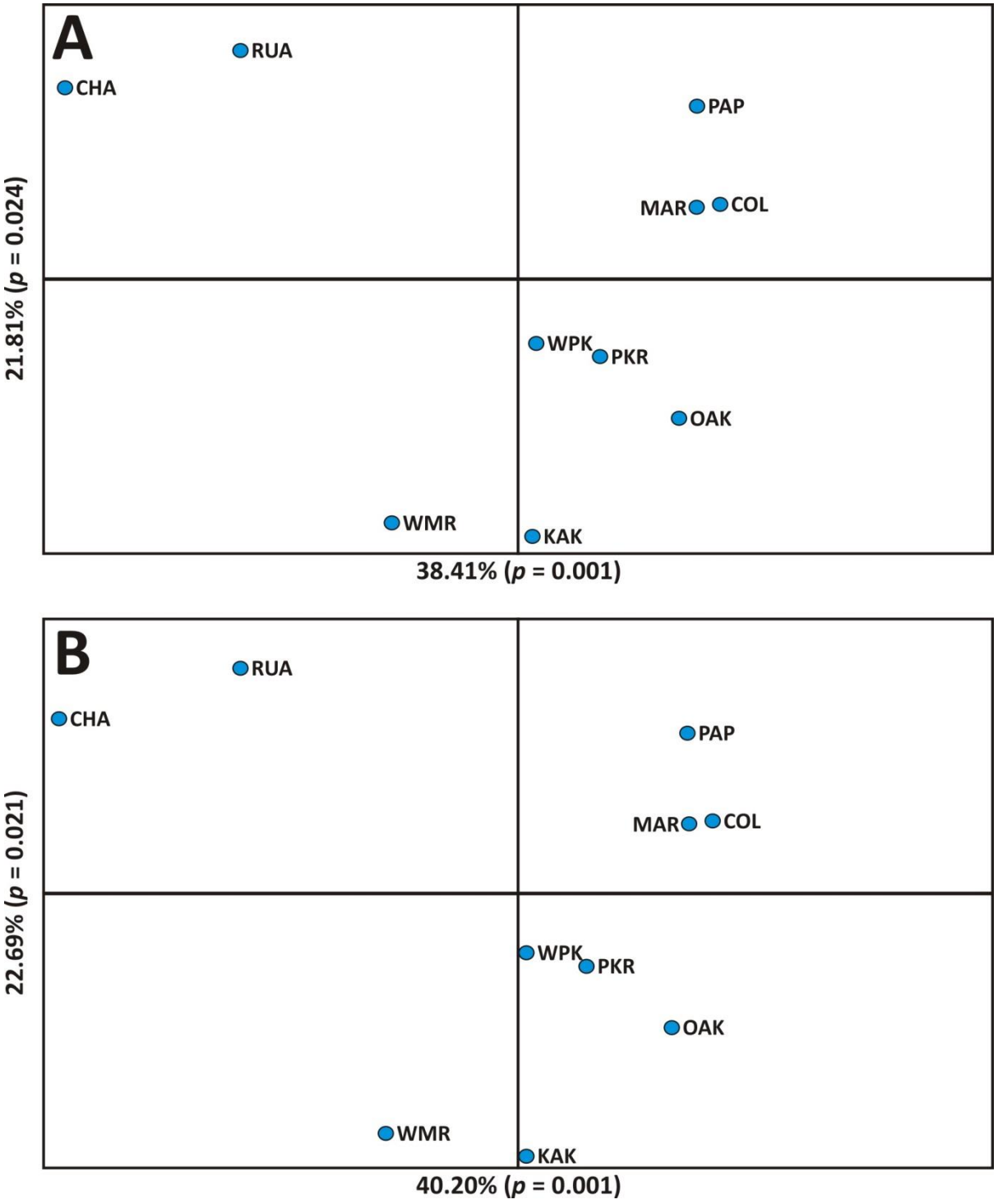


Figure 3.6: Principal component analysis (PCA) for *Paphies subtriangulata* showing patterns of genetic population differentiation A: using all loci; B: excluding outlier loci. The percentage of inertia explained by each axis and significance of the axis are displayed. An axis was considered significant if $p < 0.05$. Location abbreviations as per Table 3.1.

Table 3.7: AMOVA analyses for *Paphies subtriangulata*. Population structure was considered significant if $p < 0.05$ after false discovery rate correction for multiple tests.

Groupings Tested	**Group Configurations**	**Source of Variance**											
		Among groups			Among populations within groups			Among individuals within populations			Among individuals		
		% Var.	F_{CT}	*p*-value	% Var.	F_{SC}	*p*-value	% Var.	F_{IS}	*p*-value	% Var.	F_{IT}	*p*-value
All loci													
1 Group		-	-	-	3.210	0.032	**0.001**	33.578	0.347	**0.001**	63.212	0.368	**0.001**
North-South Division	{RUA, PAP, WMR, KAK, WPK}{MAR, PKR, COL, OAK}	0.340	0.003	0.052	0.036	0.012	**0.001**	35.258	0.358	**0.001**	63.228	0.368	**0.001**
East-West Division	{RUA, PAP, WMR, MAR}{WPK, KAK, OAK, PKR, COL}	-0.029	0.000	0.466	0.043	0.014	**0.001**	35.316	0.358	**0.001**	63.332	0.367	**0.001**
Differentiated populations	{CHA}{RUA}{WMR}{KAK}{PAP, MAR, PKR, COL, OAK, WPK}	4.082	0.041	**0.001**	0.404	0.004	**0.018**	33.135	0.347	**0.001**	62.378	0.376	**0.001**
Non-outlier loci													
1 Group		-	-	-	3.493	0.035	**0.001**	32.697	0.339	**0.001**	63.810	0.362	**0.001**
North-South Division	{RUA, PAP, WMR, KAK, WPK}{MAR, PKR, COL, OAK}	0.403	0.004	0.055	1.297	0.013	**0.001**	34.038	0.346	**0.001**	64.262	0.357	**0.001**
East-West Division	{RUA, PAP, WMR, MAR}{WPK, KAK, OAK, PKR, COL}	-0.023	0.000	0.490	1.536	0.015	**0.001**	34.103	0.346	**0.001**	64.383	0.356	**0.001**
Differentiated populations	{CHA}{RUA}{WMR}{KAK}{PAP, MAR, PKR, COL, OAK, WPK}	4.480	0.045	**0.001**	0.409	0.004	**0.021**	32.224	0.339	**0.001**	62.886	0.371	**0.001**

STRUCTURE analyses of all locations revealed two genetic clusters with similar results when the outlier locus *Paus_2* was included and excluded (Figure 3.7A, B; Figure 3.9A, B). Cluster 1 was most prevalent among the mainland New Zealand locations and cluster 2 was most prevalent in the Chatham Island population, although cluster 2 was found at all of the mainland locations to varying degrees, most notably at Ruakaka (Table 3.8). The strong genetic structure between Chatham Island and mainland New Zealand was supported by a Spearman rank correlation test, which showed a significant longitudinal difference in the distribution of clusters (Spearman's r -0.72-0.72; p = 0.018-0.022). AWclust analyses revealed eight genetic clusters when all loci were used and five clusters when *Paus_2* was excluded (Figure 3.8A, B). For both analyses cluster 1 was dominant at Chatham Island but rare among mainland locations (Table 3.8; Figure 3.9C, D). There was no significant correlation among the proportion of clusters and latitude or longitude. STRUCTURE and AWclust analyses both detected a difference in population structure between mainland New Zealand locations and Chatham Island, but AWclust also detected much more structure within the mainland New Zealand populations.

When STRUCTURE analyses were repeated for mainland New Zealand locations four genetic clusters were observed when *Psub_2* was both included and excluded from the analysis (Figure 3.7C, D). There was admixture of clusters within each location although Ruakaka was characterised by a high proportion of cluster 1 (Table 3.8; Figure 3.10A, B). AWclust analyses using only mainland New Zealand locations revealed six genetic clusters for all loci and three clusters when *Psub_2* was excluded (Figure 3.8C, D; Figure 3.10C, D). As with STRUCTURE analyses, admixture of clusters was present at all locations and there was little evidence for a pattern to the geographical distribution of clusters. No significant correlations between cluster proportion and latitude/longitude were found for mainland New Zealand locations with both STRUCTURE and AWclust analyses.

Figure 3.1 summarises the findings of analyses for genetic population structure in *P. subtriangulata*. The location of a major genetic break between mainland New Zealand populations and Chatham Island is illustrated (A) along with five other putative locations where dispersal among populations may be restricted (i-v).

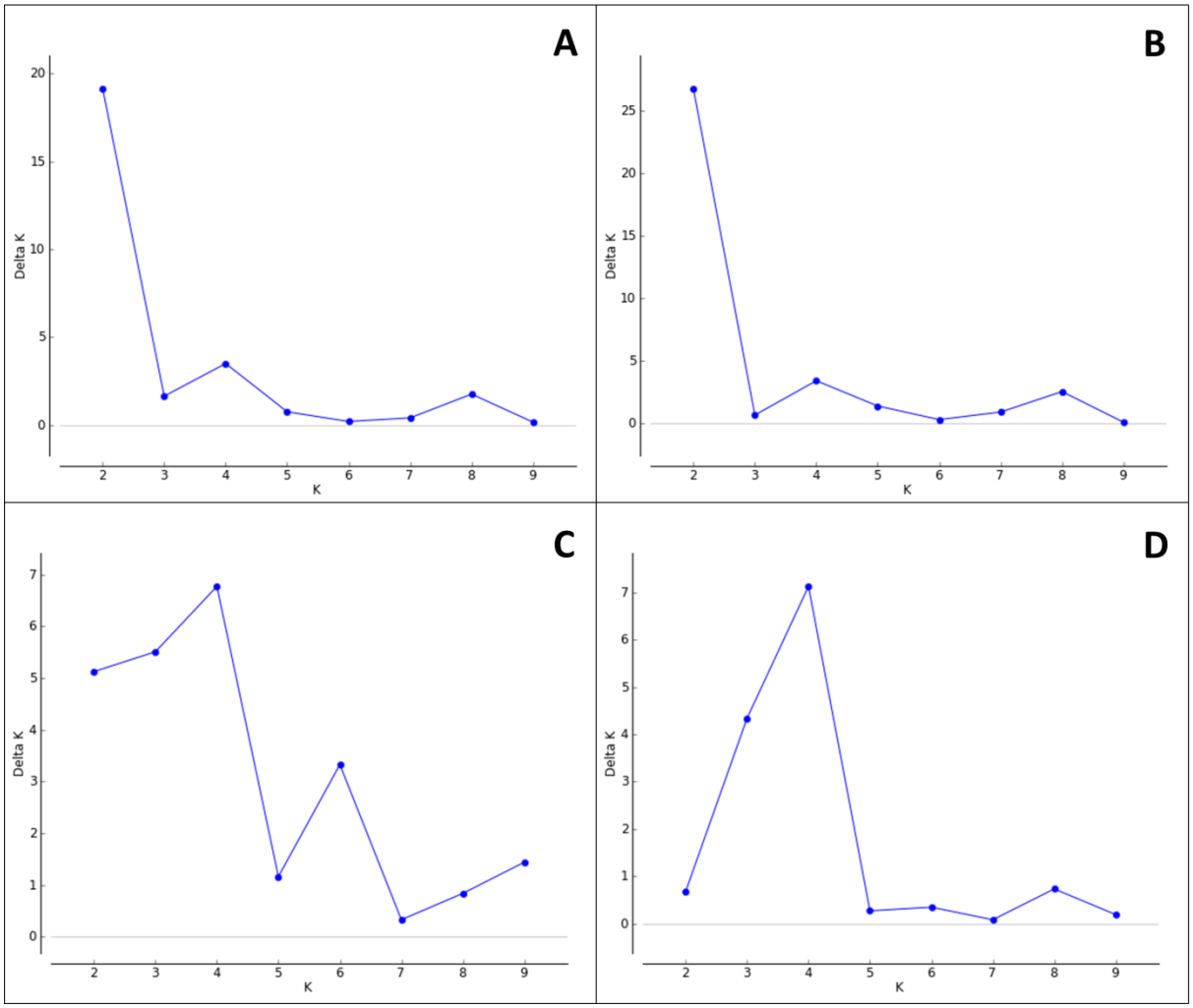


Figure 3.7: Plots to determine optimum *K* values for STRUCTURE analyses for *Paphies subtriangulata*. Delta*K* values were derived from the Evanno *et al.* (2005) method and generated in STRUCTURE HARVESTER (Earl & vonHoldt 2011) to determine the optimal value of *K* for A: all locations using all loci (*K* = 2); B: all locations using non-outlier loci (*K* = 2); C: excluding Chatham Island using all loci (*K* = 4); and D: excluding Chatham Island using non-outlier loci (*K* = 4).

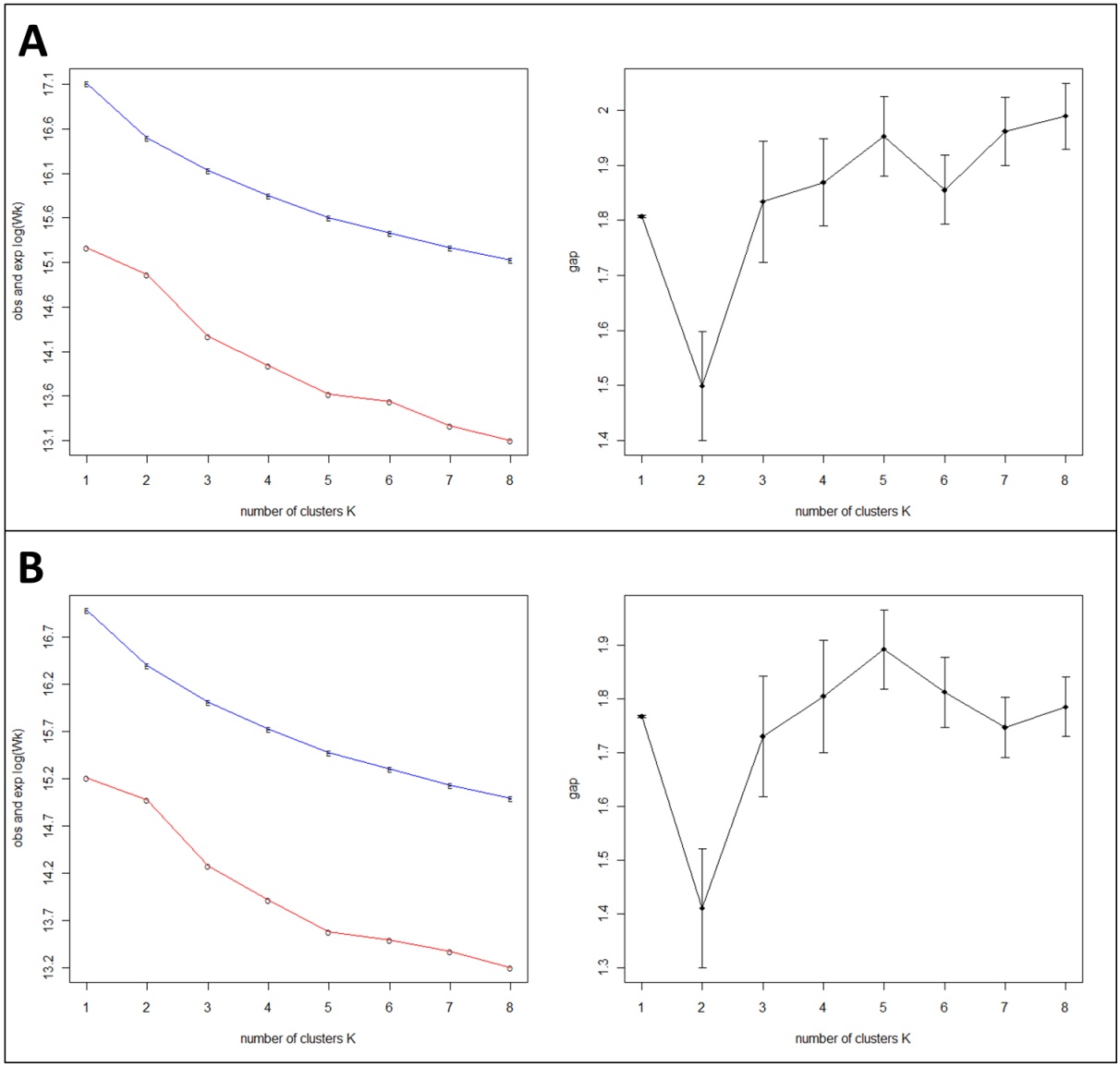


Figure 3.8: Plots to determine optimum *K* values for AWclust analyses for *Paphies subtriangulata*. Gap statistics generated from AWclust (Gao & Starmer 2008) to determine the optimal value of *K* for A: all locations using all loci (*K* = 8); B: all locations using non-outlier loci (*K* = 5); C: excluding Chatham Island using all loci (*K* = 6); and D: excluding Chatham Island using non-outlier loci (*K* = 3).

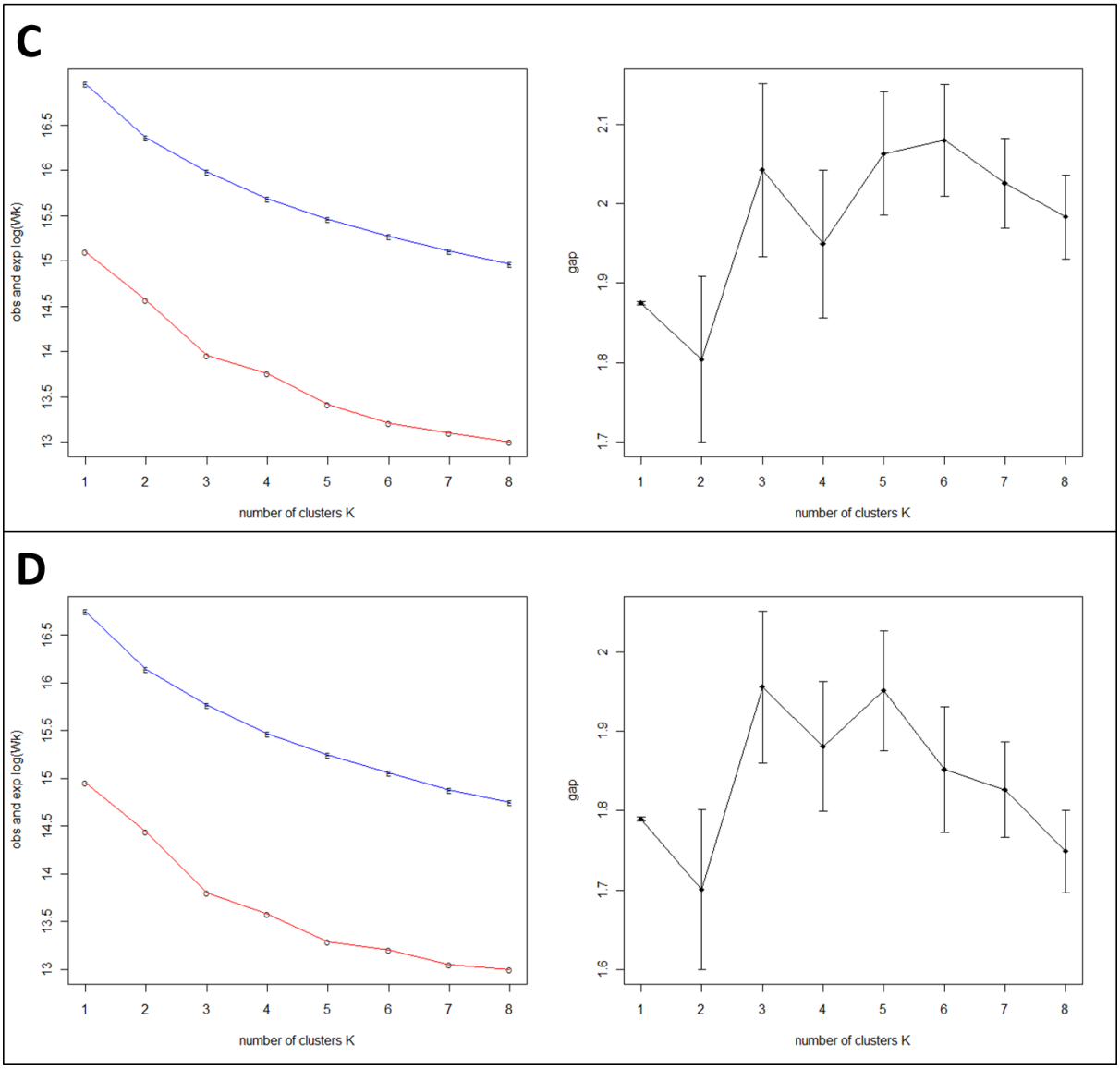


Figure 3.8 continued.

Table 3.8: Proportion of clusters found at *Paphies subtriangulata* locations from STRUCTURE and AWclust analyses. Cluster colours relate to those used in Figure 3.9 and Figure 3.10.

Analysis	Cluster	Location											
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	Total
STRUCTURE: all loci												
	1	0.663	0.924	0.876	0.933	0.964	0.976	0.994	0.923	0.950	0.019	0.822
	2	0.337	0.076	0.124	0.067	0.036	0.024	0.006	0.077	0.050	0.981	0.178
STRUCTURE: non-outlier loci												
	1	0.707	0.930	0.864	0.942	0.973	0.982	0.993	0.942	0.942	0.017	0.829
	2	0.293	0.070	0.136	0.058	0.027	0.019	0.007	0.058	0.058	0.983	0.171
STRUCTURE: excluding Chatham Islands - all loci												
	1	0.400	0.051	0.130	0.035	0.025	0.038	0.005	0.085	0.037	-	0.090
	2	0.277	0.579	0.405	0.600	0.609	0.534	0.635	0.490	0.394	-	0.503
	3	0.140	0.231	0.098	0.170	0.118	0.224	0.132	0.129	0.210	-	0.161
	4	0.183	0.139	0.367	0.195	0.248	0.204	0.228	0.297	0.359	-	0.246
STRUCTURE: excluding Chatham Islands - non-outlier loci												
	1	0.375	0.058	0.149	0.029	0.029	0.035	0.005	0.085	0.033	-	0.088
	2	0.270	0.551	0.354	0.570	0.585	0.493	0.613	0.442	0.346	-	0.469
	3	0.191	0.306	0.137	0.251	0.202	0.336	0.180	0.156	0.323	-	0.231
	4	0.165	0.086	0.361	0.150	0.183	0.137	0.202	0.318	0.298	-	0.211
AWclust: all loci												
	1	0.226	0.039	0.137	0.017	0.054	0.019	0.050	0.017	0.018	0.875	0.146
	2	0.113	0.254	0.176	0.250	0.218	0.076	0.225	0.140	0.092	0.083	0.163
	3	0.169	0.058	0.235	0.160	0.145	0.230	0.200	0.245	0.240	0.020	0.171
	4	0.094	0.176	0.000	0.107	0.090	0.211	0.025	0.087	0.166	0.020	0.098
	5	0.018	0.156	0.196	0.071	0.200	0.096	0.250	0.228	0.111	0.000	0.133
	6	0.056	0.137	0.058	0.107	0.072	0.153	0.075	0.035	0.129	0.000	0.083
	7	0.113	0.156	0.078	0.196	0.181	0.134	0.100	0.193	0.166	0.000	0.132
	8	0.207	0.019	0.117	0.089	0.036	0.076	0.075	0.052	0.074	0.000	0.075

Table 3.8 continued.

Analysis	Cluster	Location											
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	Total
AWclust: non-outlier loci												
	1	0.094	0.000	0.039	0.000	0.018	0.000	0.025	0.000	0.000	0.958	0.114
	2	0.150	0.235	0.235	0.250	0.236	0.211	0.200	0.193	0.203	0.020	0.194
	3	0.339	0.098	0.313	0.178	0.145	0.134	0.250	0.210	0.222	0.020	0.191
	4	0.207	0.274	0.137	0.196	0.236	0.211	0.125	0.228	0.296	0.000	0.191
	5	0.207	0.392	0.274	0.375	0.363	0.442	0.400	0.368	0.277	0.000	0.310
AWclust: excluding Chatham Islands - all loci												
	1	0.339	0.215	0.274	0.196	0.218	0.057	0.175	0.105	0.055	-	0.182
	2	0.169	0.058	0.215	0.160	0.145	0.192	0.200	0.245	0.222	-	0.179
	3	0.094	0.235	0.039	0.125	0.109	0.288	0.075	0.105	0.185	-	0.140
	4	0.075	0.176	0.215	0.125	0.163	0.115	0.250	0.210	0.148	-	0.164
	5	0.113	0.235	0.137	0.267	0.254	0.173	0.175	0.210	0.240	-	0.201
	6	0.207	0.078	0.117	0.125	0.109	0.173	0.125	0.122	0.148	-	0.134
AWclust: excluding Chatham Islands - non-outlier loci												
	1	0.528	0.294	0.529	0.303	0.327	0.288	0.500	0.315	0.370	-	0.384
	2	0.283	0.274	0.176	0.285	0.272	0.288	0.175	0.280	0.351	-	0.265
	3	0.188	0.431	0.294	0.410	0.400	0.423	0.325	0.403	0.277	-	0.350

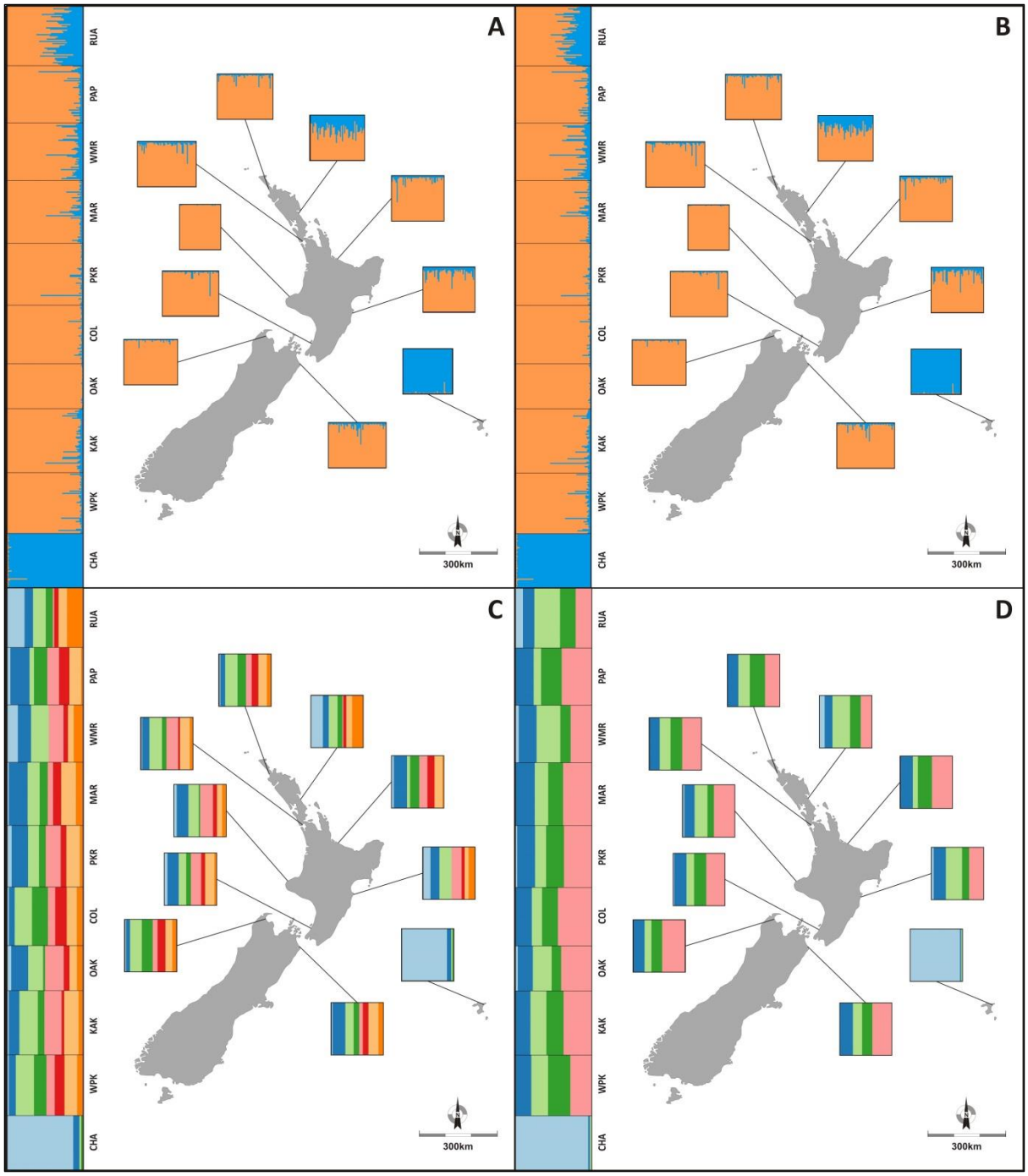


Figure 3.9: Output from cluster analyses for all *Paphies subtriangulata* locations showing the proportion of each cluster assigned to each location. Each colour denotes a different cluster as specified in Table 3.8. A: STRUCTURE analysis using all loci (*K* = 2); B: STRUCTURE analysis using non-outlier loci (*K* = 2); C: AWclust analysis using all loci (*K* = 8); D: AWclust analysis using non-outlier loci (*K* = 5). For STRUCTURE analyses each line represents an individual and individuals can belong to multiple clusters. For AWclust analyses individuals are assigned to a single cluster, the width of the bars indicates the proportion of each cluster allocated to that location.

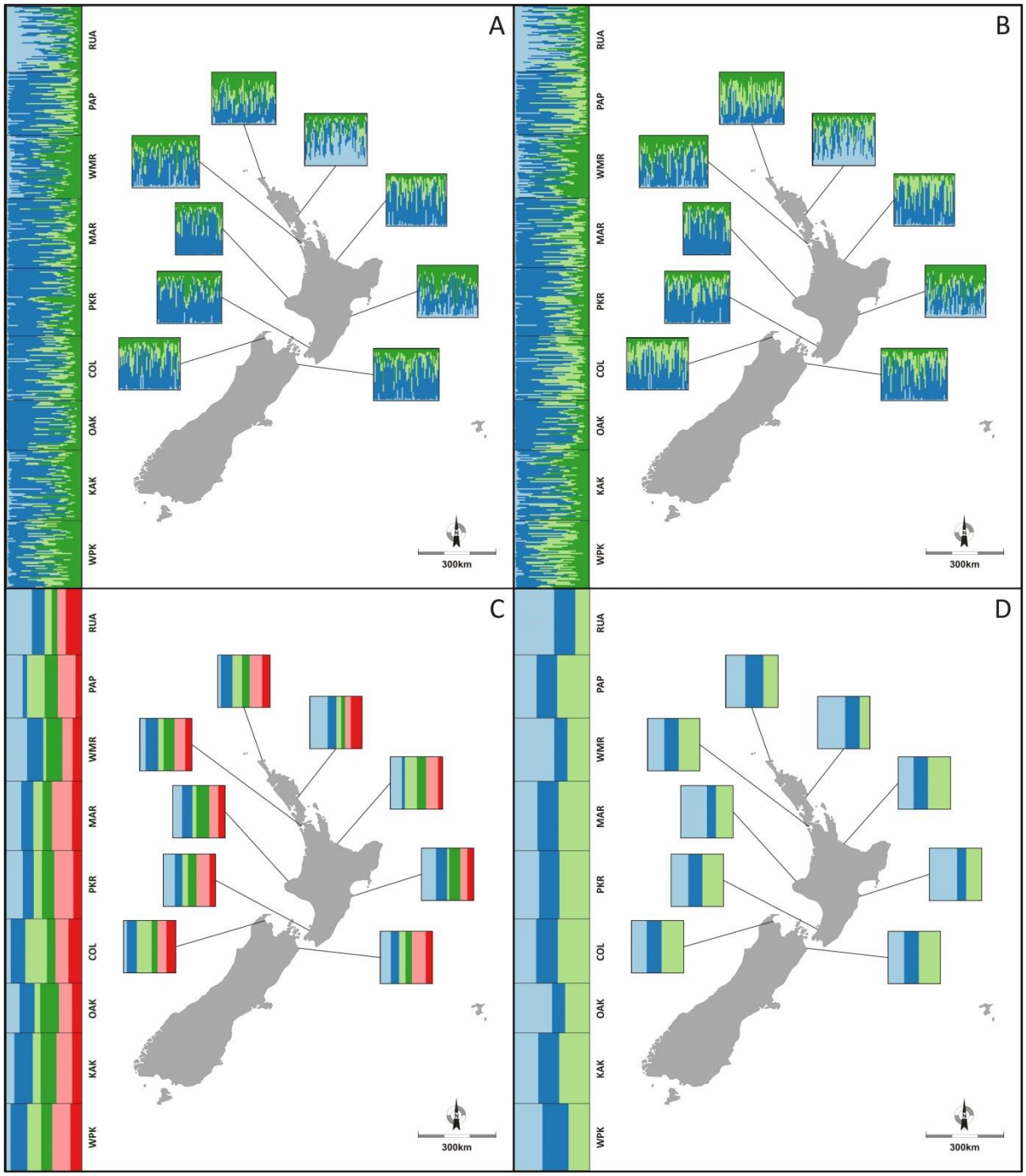


Figure 3.10: Output from cluster analyses for mainland *Paphies subtriangulata* locations (excluding Chatham Island) showing the proportion of each cluster assigned to each population. Each colour denotes a different cluster as specified in Table 3.8. A: STRUCTURE analysis using all loci (K = 4); B: STRUCTURE analysis using non-outlier loci (K = 4); C: AWclust analysis using all loci (K = 6); D: AWclust analysis using non-outlier loci (K = 3). For STRUCTURE analyses each line represents an individual and individuals can belong to multiple clusters. For AWclust analyses individuals are assigned to a single cluster, the width of the bars indicates the proportion of each cluster allocated to that population.

3.3.3 Estimates of contemporary and historical migration

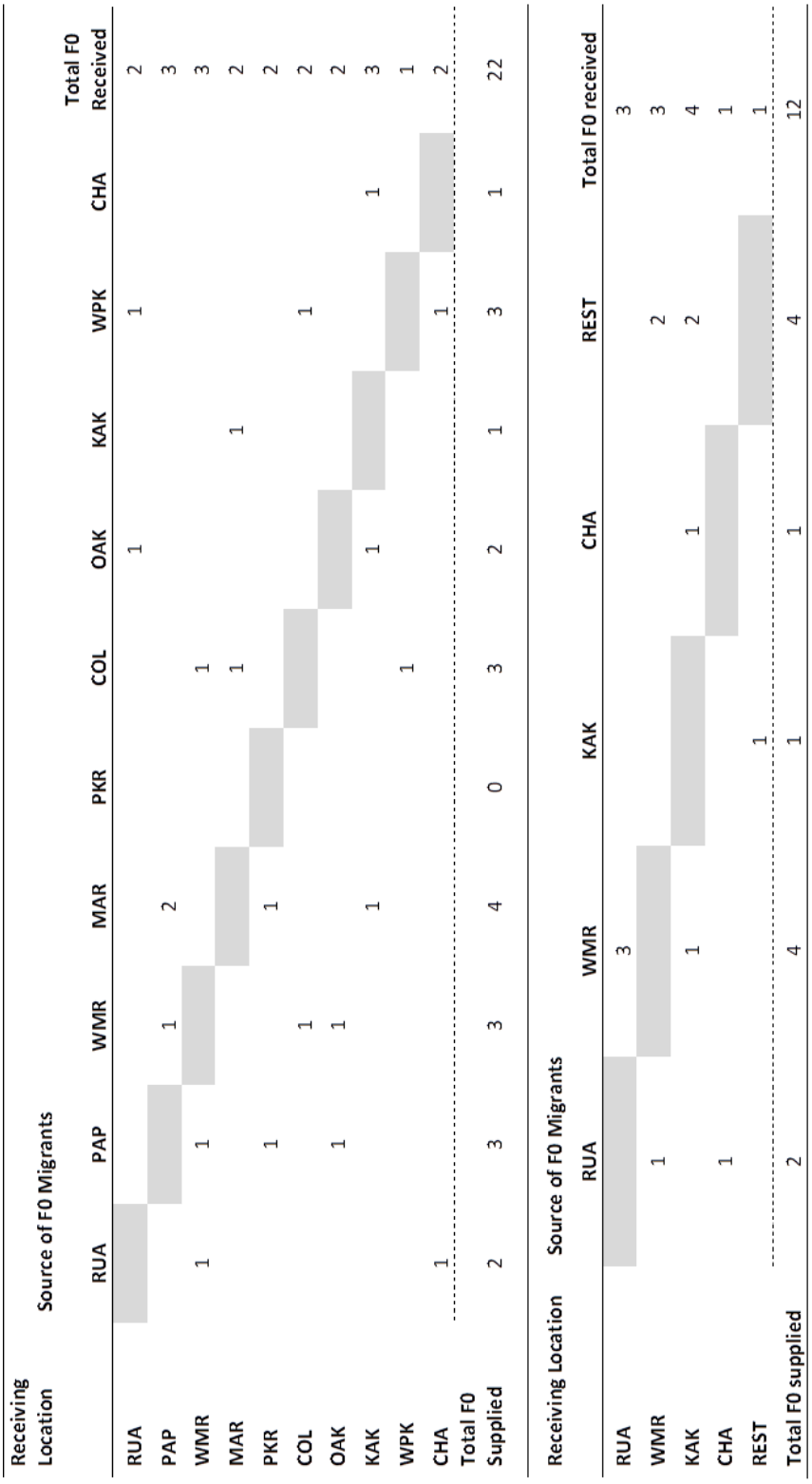
GENECLASS2 detected low numbers of first generation migrants (F0). The number of F0 migrants with a probability less than 0.01 for all *P. subtriangulata* locations was 22 (4% of individuals sampled; Table 3.9A) and the number of F0 migrants detected at each location ranged from one to three. The low number of migrants was probably an underestimation of the true number of migrants as low levels of differentiation and linkage disequilibrium among most populations mean that migrant genotypes will differ little from 'home' genotypes. This is supported by log-likelihood plots which suggested that Ruakaka and Chatham Island were the only sufficiently differentiated populations with enough power to detect migrants (data not shown). By repeating the analysis using the five population groupings shown to be significantly differentiated by exact G and AMOVA tests the power to detect F0 migrants was increased. Twelve F0 migrants were detected among the five groups (2% of individuals samples; Table 3.9B). The number of F0 migrants in each population ranged from one to four. The majority of migrants were exchanged among mainland New Zealand locations; only one F0 migrant was detected at Chatham Island and one migrant originated from this location.

BAYESASS detected modest rates of recent migration among most locations, although a limited number of source populations were found. Migration rates ranged from 4-19% and self-recruitment rates ranged from 67-95% (Figure 3.11A). Three main sources of migrants were detected (Waimarama, Paekakariki and Collingwood). The six locations that were identified as having no significant differentiation (PAP, MAR, PKR, COL, OAK, WPK) were well connected and tended to show higher migration rates. Despite being significantly differentiated from all other locations Waimarama supplied moderate numbers of migrants to Kakamatua (13%) and Waipapakauri (10%) and smaller numbers of migrants to Marfell Beach (4%). Ruakaka also received a small number of migrants from Collingwood (7%) despite being significantly differentiated from this location. No recent migration among Chatham Island and mainland New Zealand was detected. Due to the limitations of the BAYESASS program (i.e., assumes that migration is less than 1/3; Faubet *et al.* 2007) and low F_{ST} values between some locations, the power to detect true migration rates is decreased, and therefore these are likely to be underestimates of the true rates. As a consequence it is possible that such locations could experience relatively unrestricted connectivity. However

by repeating this analysis among the five genetically differentiated groups the power to accurately estimate migration rates was increased. This analysis showed that the grouping of PAP, MAR, PKR, COL, OAK and WPK was the main source of migrants (Figure 3.11B). There were high levels of migration to Waimarama (29%) and Kakamatua (31%), and lower levels of migration to Ruakaka (8%). No migration to Chatham Island was detected (Figure 3.11B).

Long-term migration rates among all locations estimated by MIGRATE displayed a large range of values but showed that migration has occurred among all locations when averaged over time. Migration rates ranged from 1.59-53.03 migrants per generation and the population size parameter θ ranged from 0.06-0.23 (Table 3.10A). Using a mutation rate of 1×10^{-4} this equates to an effective population size range of 108.33 to 608.33 individuals. On average, northern locations (Ruakaka, Waipapakauri) had the largest population sizes and supplied the largest number of migrants. Long-term migration rates estimated among the five genetically differentiated *P. subtriangulata* groups ranged from 1.83-49.50 migrants per generation (Table 3.10B). θ ranged from 0.05-0.18, resulting in effective population sizes that ranged from 125.00 to 441.68 individuals. Again, this combined analysis identified Ruakaka as a large source of migrants, followed by the grouping of PAP, MAR, PKR, COL, OAK and WPK (referred to as 'REST' in results tables). A moderate level of migration from mainland New Zealand to Chatham Island was detected, but few migrants were sourced from Chatham Island.

Table 3.9: Results of an assignment test to detect first generation (F0) migrants at each *Paphies subtriangulata* location, conducted in GENECLASS2. Migrants were detected using the *Lh/Lhmax* statistic with a probability < 0.01. A: Migrants exchanged among all locations. B: Migrants exchanged among genetically differentiated population groups as determined by exact G and AMOVA tests. 'REST' referrs to the grouping of PAP, MAR, PKR, COL, OAK and WPK.



A

Receiving Location	Source of F0 Migrants												Total F0 Received
	RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA			
RUA							1		1				2
PAP			1	2									3
WMR	1	1				1							3
MAR						1		1					2
PKR		1		1									2
COL			1						1				2
OAK		1	1										2
KAK				1			1			1			3
WPK						1							1
CHA	1								1				2
Total F0 Supplied	2	3	3	4	0	3	2	1	3	1			22

B

Receiving Location	Source of F0 Migrants					Total F0 received
	RUA	WMR	KAK	CHA	REST	
RUA		3				3
WMR	1				2	3
KAK		1		1	2	4
CHA	1					1
REST			1			1
Total F0 supplied	2	4	1	1	4	12

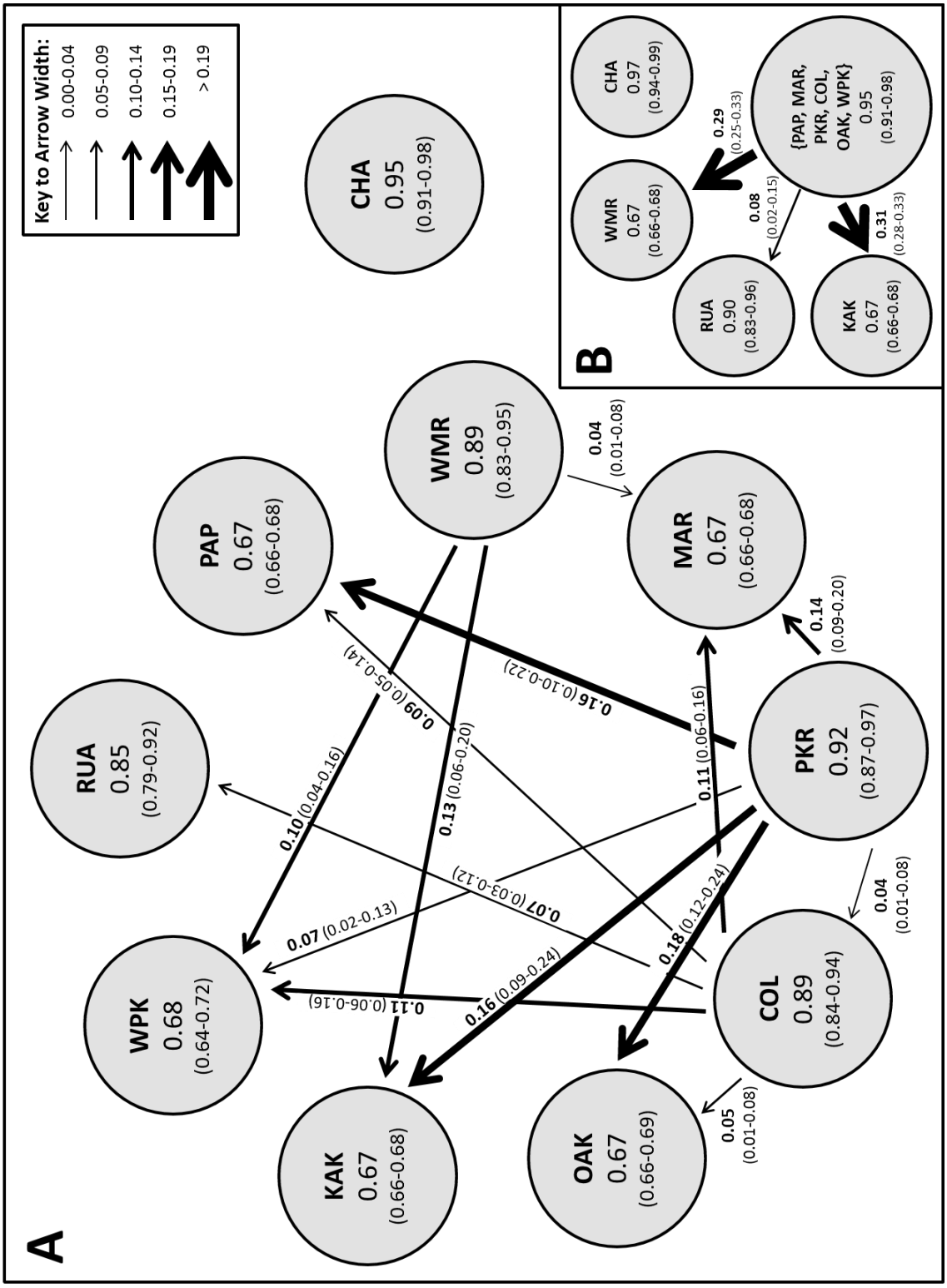


Figure 3.11: Recent migration rates for *Paphies subtriangulata* as determined by BAYESASS. Values in grey circles represent self-recruitment rates for each location (95% confidence intervals in brackets). Arrows represent migration among locations with the proportion of migrants indicated in bold (95% confidence interval in brackets). Arrow thickness indicates the relative contribution of migrants from each putative source location. A: Migration among all locations; B: Migration among genetically differentiated population groups as determined by exact G and AMOVA tests.

Table 3.10: Estimates of the number of migrants per generation (N_em), the population size parameter theta (θ), their credible intervals and effective population size (N_e, using the mutation rate 1 x 10^{-4}) for *Paphies subtriangulata*. A: Migrants exchanged among all locations. B: Migrants exchanged among genetically differentiated population groups as determined by exact G and AMOVA tests. 'REST' referrs to the grouping of PAP, MAR, PKR, COL, OAK and WPK.

A

Receiving Location	Source Location										θ	N_e
	RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA		
RUA		5.01	4.03	1.59	3.79	5.01	2.08	2.81	8.92	4.52	0.23	575.00
		0-78.67	0-53.3	0-41.33	0-53.33	0-58.67	0-46.67	0-46.67	0-98.67	0-56.00	0.05-0.40	
PAP	19.17		5.50	2.74	6.54	4.10	2.83	3.32	13.79	6.99	0.14	341.68
	0-120.00		0-56.89	0-46.85	0-60.45	0-56.71	0-49.78	0-48.22	0-108.53	0-61.60	0-0.39	
WMR	23.77	9.57		3.32	7.92	6.83	2.83	4.34	10.54	5.34	0.11	275.00
	0-104.00	0-91.78		0-50.98	0-65.78	0-62.58	0-49.78	0-52.89	0-105.24	0-57.87	0-0.27	
MAR	20.70	6.83	9.90		47.88	10.48	6.23	5.88	21.90	8.63	0.04	108.33
	0-101.33	0-86.53	0-67.56		0-176.00	0-70.40	0-66.89	0-57.56	0-144.71	0-65.33	0-0.21	
PKR	16.10	10.48	8.43	3.90		5.92	2.08	2.81	21.90	53.03	0.10	258.33
	0-93.33	0-94.40	0-65.78	0-50.98		0-60.62	0-46.67	0-45.11	0-128.27	0-171.73	0-0.27	
COL	22.23	14.12	12.10	4.19	5.86		3.97	5.37	23.52	6.99	0.14	341.68
	0-101.33	0-104.89	0-72.89	0-53.73	0-56.89		0-54.44	0-54.44	0-157.87	0-61.60	0-0.29	
OAK	11.50	10.48	4.77	3.32	7.92	10.48		6.90	17.03	6.99	0.06	141.68
	0-85.33	0-97.02	0-55.11	0-49.60	0-64.00	0-72.36		0-62.22	0-115.11	0-61.60	0-0.23	
KAK	13.03	6.83	6.97	3.32	5.86	8.66	2.83		17.03	5.34	0.08	191.68
	0-88.00	0-83.91	0-60.45	0-49.60	0-60.45	0-66.49	0-51.33		0-115.11	0-57.87	0-0.23	
WPK	8.43	5.01	4.77	1.59	4.48	6.83	2.08	3.83		4.52	0.24	608.33
	0-80.00	0-78.67	0-55.11	0-41.33	0-55.11	0-62.58	0-46.67	0-49.78		0-56.00	0-0.49	
CHA	11.50	5.92	4.03	2.46	3.79	5.92	2.08	3.32	12.17		0.12	308.33
	0-85.33	0-81.29	0-53.33	0-45.47	0-53.33	0-60.62	0-45.11	0-48.22	0-105.24		0-0.28	

B

Receiving Location	Source Location					θ	N_e
	RUA	WMR	CHA	KAK	REST		
RUA		4.03	2.32	1.83	4.90	0.18	441.68
		0-53.33	0-45.33	0-41.24	0-56.00	0-0.35	
WMR	10.01		2.32	2.17	5.83	0.11	275.00
	0-76.27		0-45.33	0-44.09	0-57.56	0-0.27	
CHA	13.55	4.77		1.83	4.90	0.06	158.33
	0-83.20	0-55.11		0-42.67	0-54.44	0-0.23	
KAK	11.19	49.50	4.86		7.23	0.05	125.00
	0-78.58	0-177.78	0-54.40		0-63.78	0-0.21	
REST	11.19	4.77	2.32	3.17		0.07	175.00
	0-85.51	0-55.11	0-45.33	0-49.78		0-0.23	

3.4 Discussion

Relatively continuous open coast marine areas are often viewed as a habitat that should promote high connectivity among locations with few barriers to dispersal and few opportunities for population differentiation (Hedgecock *et al.* 2007). Despite this perception, there is growing evidence that coastal habitats can be characterised by complex and stochastic physical processes that form barriers to connectivity and result in the genetic differentiation of populations (An *et al.* 2012; Kong *et al.*2007; Li *et al.* 2013; St-Onge *et al.* 2013). In this chapter, the influence of these processes on connectivity rates and the resulting patterns of genetic population structure were assessed for ten populations of the surf clam, *Paphies subtriangulata*. Many of the sampled locations appeared to be well connected with high levels of migration detected, and little evidence for the genetic differentiation of populations was observed. However, in other parts of the sampled range connectivity processes appear less straight-forward, as barriers to dispersal appeared to be present in some locations that have restricted larval dispersal and promoted localised genetic differentiation of populations. These patterns could be related to both geographic distance among locations and complexity in oceanographic processes.

3.4.1 Genetic diversity

Levels of genetic diversity in *P. subtriangulata* were similar to what has been reported for other clam species (e.g., Cassista & Hart 2007; Ni *et al.* 2011; Kang *et al.* 2012; Su *et al.* 2012). The high F_{IS} values, homozygote excess and large scale departures from HWE observed in *P. subtriangulata* are also commonly reported in other bivalve species (e.g., Cassista & Hart 2007; Kong *et al.* 2007; Varela *et al.* 2009; An *et al.* 2012). A likely cause of HWE departures are null alleles, which for unknown reasons are particularly common in bivalves (e.g., Launey & Hedgecock 2001; Varela *et al.* 2009; Zhan *et al.* 2009). Estimates of null allele frequency were high for some loci; when a locus had a null allele frequency greater than 20% no populations were in HWE. Although null alleles are commonly observed in microsatellite studies little is known about how they might affect estimates of population differentiation (Chapuis & Estoup 2007). In general it is thought that high frequencies of null alleles will reduce population genetic diversity, inflating F_{ST} and estimates of the level of population differentiation (Paetkau *et al.* 1997; Slatkin 1995). Simulations suggest that if effective population size (N_e) and gene flow are high, and frequency of null alleles is

moderate (5-20%) then F_{ST} estimates should be unbiased by null alleles (Chapuis & Estoup 2007). Correcting F_{ST} for null alleles was not found to have a consistent effect across loci in the present study. While null alleles are not predicted to have a large influence on estimates of population differentiation in this study, results suggest that, if anything, null alleles could be masking population structure rather than inflating it, as evidenced by the increased slope of the of the regression line in IBD analyses where null allele corrected F_{ST} was used.

At a population level, all of the sampled locations failed to meet HWE expectations, although some individual loci were in HWE within populations. The observed patterns of HWE deviation suggest that biological factors other than null alleles were responsible. Undetected spatial or temporal patterns of genetic structure (i.e., Wahlund effect) could explain some of these observations. Clustering analyses did not provide evidence for a spatial Wahlund effect, but a temporal effect cannot be ruled out. Alternatively, there is known to be a strong correlation between marine invertebrate life histories and deviation from HWE, whereby species with free-spawned planktonic sperm have significantly higher F_{IS} (Addison & Hart 2005). The mechanisms behind this correlation are unknown but could be caused by higher mutation rates and variability in reproductive success associated with high reproductive output (Addison & Hart 2005; Launey & Hedgecock 2001).

3.4.2 Genetic population structure

Fitting the pattern of genetic population structure observed for *P. subtriangulata* into the categories described for other New Zealand coastal marine species by Gardner *et al.* (2010) was not straight forward as the type of structure was dependent on the spatial scale of the analysis. When all of the sampled locations are considered, a strong IBD pattern was observed. However this pattern was driven by the strong genetic differentiation of the Chatham Island population, and is likely to reflect hierarchical structure rather than true IBD. Among mainland locations it was difficult to reject the null hypothesis of panmixia as many of the sampled locations formed a genetically undifferentiated group. However, in other parts of the sampled range divergence among populations was also observed. In general, pairwise F_{ST} was low, as is usual for marine invertebrates (Hedgecock *et al.* 2007), although there was still sufficient power to detect significant population differentiation. The patterns of genetic population differentiation and connectivity seen in *P. subtriangulata* serve as a reminder that there is spatial variation in the processes that shape populations of

marine organisms. The result is that, even for species where panmictic population structure is predicted, spatially complex patterns of genetic population structure can form and could change over time, a phenomenon referred to as 'chaotic' genetic patchiness (Hellberg *et al.* 2002; Johnson & Black 1982).

The most striking feature of genetic population structure in *P. subtriangulata* was the differentiation of the Chatham Island population from other mainland populations. Differentiation of the Chatham Island population was mainly due to differences in the frequency of alleles shared with mainland populations but some private alleles were also found at high frequency, which would contribute to differentiation. F_{ST} and ϕ'_{ST} between Chatham Island and other mainland locations was very high, even when compared to what has been reported for other marine bivalves that inhabit large geographic ranges (e.g., Benzie & Smith-Keune 2006; Lind *et al.* 2007; Ni *et al.* 2011; St-Onge *et al.* 2013; Xiao *et al.* 2010). It is similar to the levels of differentiation that have been reported for populations of the cockle *Cerastoderma glaucum* across the Mediterranean region (Tarnowska *et al.* 2010) and across the Atlantic Ocean for populations of the softshell clam *Mya arenaria* (Strasser & Barber 2008). In these studies genetic differentiation of this magnitude was observed over ranges of 3,000 to 5,000 km, rather than the 700 km between Chatham Island and mainland New Zealand. The Chatham Island population of *P. subtriangulata* was previously shown to be differentiated from mainland populations based on allozyme markers (Smith *et al.* 1989) and morphological analyses provide support that this population may represent a separate species of tuatua (Beu & de Rooij-Schuiling 1982), although species status has never been formally proposed or verified. Genetic differentiation of Chatham Island populations has been previously reported for other coastal marine species (e.g., Goldstien *et al.* 2009; Hickey *et al.* 2009; Stevens & Hogg 2004; Will *et al.* 2011) but for others there appears to be little to no restrictions on gene flow to mainland New Zealand (e.g., Buchanan & Zuccarello 2012; Ross *et al.* 2009).

Among mainland *P. subtriangulata*, most locations were found to be genetically undifferentiated from each other. Although weak IBD was detected the effect of null alleles and deviation from HWE on genetic population structure make it difficult to determine the significance of this pattern. Several lines of evidence suggest that Papamoa, Marfell Beach, Paekakariki, Collingwood, Oakura and Waipapakauri can be considered as one population

unit, suggesting few restrictions to connectivity among these locations. Ruakaka, Waimarama and Kakamatua are differentiated from this group to varying degrees. Of all the mainland populations Ruakaka was the most genetically distinct from other locations but interestingly, the Ruakaka population also showed the highest degree of genetic similarity to the Chatham Island population. Why this pattern exists is unclear but could represent the remnants of historic gene flow, be an artefact of the markers used or be related to other stochastic effects independent of gene flow. There was also evidence for genetic similarities between Waimarama and Chatham Island populations and this is the mainland population with the closest proximity to Chatham Island.

The pattern of genetic population differentiation observed in *P. subtriangulata* differs somewhat from that reported by Smith *et al.* (1989). These authors reported a 'northern' group and a 'central' group with evidence that Stewart Island samples may form a third 'southern' group. Smith *et al.* (1989) did sample populations in close proximity to Ruakaka and Waimarama (but not to Kakamatua). In contrast to the present study, these populations were not found to be differentiated from other 'northern' populations, but there was evidence that they were differentiated from populations at the top of the South Island and bottom of the North Island (i.e., the 'central' group). Differences in the population structure inferred by different marker types is commonly observed, e.g., the greenshell mussel *Perna canaliculus* (Apte & Gardner 2001; Star *et al.* 2003) and the cockle *Austrovenus stutchburyi* (Lidgard 2001; Ross *et al.* 2011). Alternatively, the differences observed between the two studies might mean that genetic population structure in *P. subtriangulata* has changed over time, which could be the result of temporal variability in connectivity.

The pattern of population differentiation seen among mainland *P. subtriangulata* locations is unusual among New Zealand coastal marine species, although not unexpected. Large areas of panmictic genetic population structure were predicted based on the life history characteristics of the species and results from previous genetic analyses. Isolated cases of genetic population differentiation are not unexpected for coastal marine species and the areas of restricted dispersal identified for *P. subtriangulata* in this study have also been reported for other species. Differentiation of populations from the northern part of the North Island has been reported in the seaweed *Carpophyllum maschalocarpum*

(Buchanan & Zuccarello 2012) and the cockle *Austrovenus stutchburyi* (Ross *et al.* 2011). A genetic break at the East Cape on the eastern side of the North Island has also been commonly reported (e.g., amphipods *Paracorophium lucasi* and *P. excavatum*, Stevens & Hogg 2004; black-foot abalone *Haliotis iris*, Will *et al.* 2011), similar to that seen in the Waimarama population. There is little evidence from other marine organisms for differentiation on the west coast of the North Island but the Kakamatua population is unusual as it is the only *P. subtriangulata* population in this study that is located inside a harbour.

There was little evidence for the north-south differentiation pattern that has commonly been observed in New Zealand coastal marine species. The two populations from the top of the South Island (Marfell Beach and Collingwood) were undifferentiated from most North Island populations, although similar genetic population structure has been seen in the greenshell mussel *Perna canaliculus*. Star *et al.* (2003) reported a genetic break for *P. canaliculus* at -42° S. Evidence from Smith *et al.* (1989) suggests that a similar pattern of differentiation could exist for *P. subtriangulata* (i.e., there was evidence that the southernmost sample from Stewart Island was differentiated from other populations). Because samples from mainland populations south of this latitude could not be obtained it was not possible to test for this pattern.

3.4.3 Patterns of migration

Patterns and rates of migration generally support the population structure that was observed for *P. subtriangulata*. It appears that little to no migration has occurred between Chatham Island and mainland New Zealand in the recent past. The vast majority of the Chatham Island population has originated from self-recruitment (95-97%), resulting in significant differentiation of this population due to restricted gene flow. Some genetic similarities to mainland populations remain, but it is unclear whether these similarities represent contemporary gene flow or are maintained due to slow rates of genetic drift. The two F0 migrants exchanged among the mainland and Chatham Island populations suggest that rare migration events could occasionally occur. Long-term migration rates provide support for past gene flow, although migration rates to and from Chatham Island were low compared to migration rates among mainland locations.

Among mainland locations both recent and historic migration rates were high, particularly among locations that were not genetically differentiated. Recent migration among locations was found to be unidirectional (i.e., gene flow between locations was detected in one direction only) and not detected at all among some locations, but migration among all locations in both directions was detected in the long-term. All analyses suggested that the group consisting of the six undifferentiated locations (PAP, MAR, COL, PKR, OAK, WPK) were well connected via gene flow. Waimarama and Kakamatua appear to be well connected to this group in both the short and long-term, despite their significant genetic differentiation. In contrast, Ruakaka appears to be more heavily reliant on self-recruitment (85-90%), which is consistent with the significant genetic differentiation detected at this location. However, long-term analyses indicate that the Ruakaka population may occasionally be connected to other populations.

It seems unusual that genetic differentiation could develop among locations when migration estimates are high; however this finding could also be a consequence of temporally patchy larval dispersal and recruitment (Hedgecock *et al.* 2007). Temporal variation in reproductive output and/or physical oceanographic conditions means that fluctuations in the supply of larvae from different source populations can occur (Hedgecock & Pudovkin 2011). Populations may be well connected in some years and rely on large amounts of self-recruitment at other times, allowing genetic differentiation to occur via genetic drift, mutation or selection. The contrasting patterns of short versus long-term migration reported here for *P. subtriangulata* provide evidence for this phenomenon and studies reporting temporal variability in population connectivity are common for marine species (e.g., Planes & Lenfant 2002; Selkoe *et al.* 2006; Varela *et al.* 2009). Furthermore, given that the distance among some of the locations sampled was greater than the estimated dispersal distance for *P. subtriangulata* larvae, it seems highly likely that some locations will only be connected by rare long-distance dispersal events. These results provide evidence that consistent connectivity occurs among many genetically undifferentiated locations, but in some places it could be more intermittent, allowing genetic differentiation to arise between populations. Furthermore, these results highlight how larval dispersal for marine species can be temporally variable.

Temporal variability in population connectivity is expected to have a more pronounced effect on the genetic structure of populations when effective population size is low (Lowe & Allendorf 2010). Effective population size is hypothesised to be low among species that are characterised by high fecundity and high mortality in early life stages (Hedrick 2005; Palstra & Ruzzante 2008) and long-term N_e for *P. subtriangulata* was found to be low compared to estimates for other marine invertebrate populations (e.g., Cassista & Hart 2007; Jolly *et al.* 2009; Sui *et al.* 2009). Low N_e is often attributed to 'sweepstakes reproductive success' (Hedgecock 1994), whereby chance events mean that a small number of breeders may be responsible for producing the majority of offspring in each generation. This can result in a very low N_e/N ratio, where N is the census size of the population. Frankham (1995) calculated that the average N_e/N ratio from published studies was 0.11 but ratios as low as 10^{-5} to 10^{-6} have been observed in Pacific oysters (Hedgecock 1994). This means that despite low estimates of N_e, actual population size for *P. subtriangulata* may be very large, possibly in the range of 10^6 or 10^7 individuals per population. Low N_e will also have implications for how opposing gene flow, genetic drift and mutational forces will structure populations. Theoretically, gene flow will override the effects of random genetic drift when $m > 1/4N_e$ (where m is the migration rate; Wright 1931). Despite low N_e, $1/4N_e$ is still much higher than m for the majority of *P. subtriangulata* populations so it is likely that genetic drift is a dominant force in structuring populations for this species. Coupled with the high mutation rates expected for invertebrate microsatellite loci (Gow *et al.* 2005) and low effective population size it might be possible for populations to become significantly differentiated over short time periods despite on-going long-term gene flow (Lowe & Allendorf 2010).

3.4.4 Possible explanations for observed genetic population structure and migration

The genetic population structure observed in *P. subtriangulata* is consistent with what is expected for a free-spawning marine bivalve with a three week pelagic larval duration. From a simplistic point of view, this period should be sufficient for short scale dispersal among populations over hundreds of kilometres (Shanks 2009; Shanks *et al.* 2003) and over larger scales via the stepping-stone model of connectivity thought to characterise many coastal marine species (Hellberg 2009). The relatively continuous nature of the open coast habitat of *P. subtriangulata* means that larvae should quickly disperse into coastal

currents and suitable habitat should be readily available for settlement, resulting in well-connected, undifferentiated populations. Certainly this is the case for the majority of mainland locations, which show little genetic differentiation and consistent levels of migration at all timescales investigated. However, the genetic differentiation observed among some populations provides evidence that this simplistic model is not applicable across the whole of the range sampled in this study. Instead, restrictions to gene flow are occurring in some parts of New Zealand that are allowing more complex patterns of genetic population structure to form.

For the Chatham Island population the obvious explanation for genetic differentiation from the mainland is that distance is limiting larval dispersal. The distance between Chatham Island and mainland New Zealand seems to pose no limitations to dispersal for some species but does limit gene flow to varying degrees for other species. The offshore flow of the East Cape and Southland currents across the Chatham Rise is hypothesised to transport larvae away from the coast into unsuitable open ocean habitat (Hadfield *et al.* 2007; Heath 1985). However, some authors have proposed this as a mechanism of connectivity from mainland populations to Chatham Island for some organisms with pelagic larval duration of sufficient length to survive the crossing (e.g., Ross *et al.* 2011).

The inconsistent pattern of genetic population differentiation seen at Chatham Island across multiple species is probably due to the unique reproductive and life history characteristics of each species. For example, it is not unexpected that the seaweed *Carpophyllum maschalocarpum* shows little differentiation at Chatham Island as seaweed is known for dispersing large distances via rafting (Buchanan & Zuccarello 2012). However it is unclear why *P. subtriangulata* populations would be genetically differentiated when the bivalve *Austrovenus stutchburyi*, with similar reproductive characteristics and larval duration, shows no differentiation at Chatham Island (Ross *et al.* 2011). A possible explanation is that *A. stutchburyi* is likely to have larger, higher density populations compared to *P. subtriangulata* (personal observation), which could result in higher reproductive output and provide more opportunity for long distance dispersal events. Genetic similarities and detection of F0 migrants between the Chatham Island population and some mainland locations provides evidence for *P. subtriangulata* that infrequent long-

distance migration events may occur to Chatham Island. Alternatively, if the Chatham Island population represents a separate, recently diverged species of tuatua, as suggested by previous authors (Beu & de Rooij-Schuiling 1982; Smith *et al.* 1989), time since divergence may make it possible for mainland and Chatham Island populations to still share alleles but other biological mechanisms, such as reproductive isolation, may be driving differentiation.

The majority of mainland *P. subtriangulata* locations were found to be genetically undifferentiated, suggesting that larval dispersal between these locations has been unrestricted in both the short and long-term. Where genetic differentiation was observed among mainland populations there are several mechanisms that could have contributed to these patterns forming. It is possible that the distance between locations may be responsible for limiting gene flow, although not to the same extent as observed for the Chatham Island population. This would be consistent with the weak IBD pattern that was found among mainland locations. Finding a small degree of IBD was not surprising because the sampling sites in this study were further apart than what larvae should be able to travel in a single generation. Where populations were genetically differentiated (e.g., Ruakaka and Waimarama) it is possible that eddies may be entraining larvae to promote self-recruitment and prevent larval exchange with other populations.

The North Cape Eddy (Stanton *et al.* 1997) and Wairarapa Eddy (Chiswell 2003; Chiswell & Roemmich 1998), located in the Northland and East Coast regions respectively, may be the mechanism behind the genetic differentiation observed in *P. subtriangulata* populations from these regions (see Figure 1.1). High levels of self-recruitment estimated for these populations provide support for local larval retention, but evidence for temporal variability in migration rates implies that this population structure could be transient and change over time. Differences between the genetic population structure reported in the present study and that reported by Smith *et al.* (1989) provide evidence for such temporal instability in *P. subtriangulata* genetic population structure, despite apparently high levels of genetic connectivity. This is supported by low estimates of N_e, suggesting that genetic drift should be a powerful force in shaping *P. subtriangulata* genetic population structure by promoting the differentiation of populations.

Differentiation of the Kakamatua population could be attributed to its location inside the Manukau Harbour. Tidal circulation inside harbours is known to retain larvae, limit connectivity and promote genetic population differentiation (Bilton 2002). Furthermore, the characteristics of a harbour habitat are very different from those of the open coast beach habitat where *P. subtriangulata* more commonly occurs. As a consequence the Kakamatua population may be locally adapted to this different habitat type, which could be driving population differentiation.

Life history characteristics (e.g., habitat, pelagic larval duration) and ocean circulation can be invoked to explain much of the genetic population structure and migration patterns that were observed in *P. subtriangulata*. However, these processes are often not able to adequately explain why genetic differentiation has occurred in some locations (e.g., Kakamatua). Chapter 5 will further investigate the environmental factors that may be responsible for influencing the patterns of genetic population structure and connectivity that were observed in *P. subtriangulata*.

3.4.5 Conclusions

The types of genetic population structure seen in *P. subtriangulata* showed spatial variation across the sampled range. Overall, genetic population structure was found to be characterised by an isolation by distance pattern, although this pattern is likely to represent hierarchical structure driven by the high level of genetic differentiation between the Chatham Island population and the remaining mainland New Zealand populations. This structure is likely to be a result of the distance between these locations, which poses a barrier for the dispersal of larvae. Among mainland *P. subtriangulata*, most locations were undifferentiated and well-connected, suggesting that to some extent the paradigm of unrestricted linear gene flow along a relatively unstructured coast is applicable to *P. subtriangulata*. However, there was also evidence that genetic divergence has occurred among some populations.

While most locations were found to be genetically well connected over historic time scales, recent migration rates suggest that gene flow may be variable on a temporal scale. This is likely to be a consequence of oceanographic features that restrict larval dispersal and promote self-recruitment for some populations. As a consequence, isolated occurrences of

significant population differentiation among mainland *P. subtriangulata* might not be temporally stable, i.e., genetic population structure could change over time. Although barriers to connectivity appear to exist in some locations for *P. subtriangulata* these barriers might not be absolute. Instead they may be semi-permeable, allowing for gene flow among locations on occasion. The results of this study have highlighted how despite reproductive characteristics and habitat preferences that are suggestive of well-connected and 'open' genetic population structure, species inhabiting open coastal marine environments can have unpredicted population structure consistent with restricted connectivity; these patterns can show variation over time and over the geographic range of a species.

4 Interaction between habitat availability and oceanographic processes structure populations of an estuarine species: population genetics and connectivity in *Paphies australis*

4.1 Introduction

In chapter 3 I highlighted that while the oceans are seemingly continuous and free of restrictions to dispersal, it is still common for genetic diversity within a species to become geographically structured into distinct population units. It is important for population geneticists to define the number and geographic location of these units, as well as the rates and pattern of gene flow among them, to assist with management and/or conservation of marine species. While the biological characteristics of a species are important drivers of population connectivity, the complex physical nature of the coastal marine environment can have unpredictable and variable effects on how populations are genetically structured over time and space (Cowen *et al.* 2007; Pineda *et al.* 2007). The way in which environmental complexity influences genetic population structure is particularly relevant for organisms living in estuaries and harbours, such as *Paphies australis* (pipi), as the geographical features of these habitats are likely to create barriers that will restrict larval dispersal to varying degrees. Estuaries are the type of habitat where connectivity is likely to become 'crinkled' (Ovenden 2013); gene flow may be restricted to the point where genetic connectivity can be maintained in the long-term, but short-term demographic connectivity may be limited and variable, allowing populations to become genetically differentiated.

Compared to open coastal habitats, estuaries are typically enclosed, discrete environments, separated from other such habitats by considerable geographical distance (Bilton 2002). The organisms that inhabit estuaries cope with variable and extreme

environmental gradients (e.g., temperature, salinity, sedimentation), and complex circulation regimes (Watts & Johnson 2004). Locally adapted and genetically distinct estuarine populations are common for many organisms (e.g., Caudill & Bucklin 2004; McCairns & Bernatchez 2008; McCraney *et al.* 2010; Tarnowska *et al.* 2010). These habitats contain both specialised estuarine organisms as well as estuarine ecotypes of typically marine species (Bilton 2002). Furthermore, the geologically recent origin of many estuaries means that sufficient time might not have passed for populations to reach migration-drift equilibrium and population genetic signals may be influenced by historic events (Williams *et al.* 2008). The characteristics of estuarine habitats along with the high spatial and temporal variation in the physical processes operating within estuaries can cause complex and unpredictable genetic signals to arise in estuarine organisms (Pelc *et al.* 2009).

Understanding how populations of estuarine organisms are genetically structured and connected is important; estuaries are on the interface of the terrestrial/marine environment and are heavily impacted by human activity (Cole *et al.* 2000). They are highly diverse environments, providing habitat for important fishery species and providing spawning grounds for coastal fishery species (Watts & Johnson 2004). Studies of genetic population structure in New Zealand estuarine organisms are few (Gardner *et al.* 2010) and have been limited to sea grass (Jones *et al.* 2008), amphipods (Knox *et al.* 2011; Stevens & Hogg 2004), cockles (Ross *et al.* 2011) and triplefin fish (Hickey *et al.* 2009). All of these studies found high levels of genetic structuring, with divergence of northern and southern populations around the location of Cook Strait being commonly reported. In addition, most of these studies reported further population subdivision with genetic breaks observed in locations additional to Cook Strait (Knox *et al.* 2011; Ross *et al.* 2011; Stevens & Hogg 2004). Marine organisms inhabiting fiords are expected to experience similar restrictions to dispersal and the divergence seen among Fiordland populations of black coral (Miller 1997), red coral (Miller *et al.* 2004) and starfish (Perrin *et al.* 2004) provide support for this.

Studies of genetic structure in estuarine species, such as those mentioned above, have a strong emphasis on habitat as a controlling factor in the genetic structuring of populations. The expectation is that, because coastal marine habitats are often viewed as a ‘linear’ environment, and estuaries are discontinuous ‘islands’ distributed along the coast, a simple one-dimensional isolation by distance (IBD) pattern will form (Kimura & Weiss 1964;

Slatkin 1993). However, *P. australis* is not strictly an estuarine species and can also be found in sheltered harbour and coastal areas (Morton & Miller 1968). Coastal currents may play a more significant role in structuring and connecting *P. australis* populations and it is possible that more complex patterns could be observed than those previously reported for other estuarine species. For example, Pampoulie *et al.* (2004) found a complex pattern of significant population differentiation coupled with high gene flow for a species of goby that is distributed through estuarine, coastal and marine habitats.

In chapter 3 *P. subtriangulata* populations were found to be well connected by gene flow with isolated cases of genetic differentiation that were likely to be the result of geographic distance or localised current circulation. In this chapter, the more complex nature of *P. australis* habitat presents an opportunity to further investigate the effect of habitat on determining genetic population structure. Genetic population structure and connectivity in *P. australis* is likely to be driven by two main processes. Firstly, suitable estuarine/harbour habitat must be available for settlement within the distance that larvae can disperse during their pelagic larval duration (PLD). For estuarine species this distance may be shorter than that of open coast species, as larvae originating in estuaries must first overcome any barriers associated with exiting the estuary and entering coastal currents (Bilton 2002). If suitable habitat can be reached within the PLD then connectivity is likely to be high and stepping-stone IBD patterns could be observed. If suitable habitat is not available then this may present a barrier to dispersal and should manifest as a sudden genetic discontinuity. Secondly, despite the availability of suitable habitat, dispersal could be restricted or promoted by local oceanographic conditions. Circulation of coastal currents may act to restrict the dispersal of larvae, resulting in the formation of genetically differentiated populations over small spatial scales. Alternatively, fast flowing linear currents could disperse larvae over long distances, allowing for high genetic connectivity over long stretches of coastline, even where large gaps between suitable habitat exist.

In this study of genetic population structure and connectivity in *P. australis*, I explored the relationship between habitat availability and coastal current patterns to investigate how their interaction can shape the genetic population structure of estuarine species. This study examined the relevance of the simple one-dimensional IBD model for estuarine species and whether distance between suitable habitats could be used as a proxy

for levels of genetic population differentiation. Alternatively, habitat availability could combine with oceanographic processes to create a more complex genetic structure where levels of connectivity vary spatially and temporally to create regions of high gene flow and barriers in other locations. As a clearer understanding of the genetic population structure and connectivity patterns found in estuary/harbour species develops, it will have implications for how these species are managed.

In this chapter genetic population diversity, differentiation and gene flow were estimated using the panel of microsatellite markers described for *P. australis* in chapter 2. Specifically, this chapter aimed to test a null hypothesis of panmictic genetic population structure by:

1. Quantifying levels of genetic diversity in *P. australis* sampled from thirteen locations;
2. Investigating levels of genetic differentiation among the sampled localities to determine which of the five categories from Gardner *et al.* (2010) best describes the genetic population structure of *P. australis*;
3. Estimating levels of migration among *P. australis* locations over both recent and historic time scales.

4.2 Methods

4.2.1 Sample collection and laboratory methods

Samples of *Paphies australis* were collected between June 2010 and February 2012 from 13 locations (Table 4.1; Figure 4.1). The biogeographic classification scheme of Shears *et al.* (2008) and the New Zealand fishery quota management areas were used as a basis for selecting sampling sites. The aim was to sample at least one site within each biogeographic and fishery area, dependent on being able to locate suitable populations within those areas. Samples were gathered by hand at low tide, photographed, shell length was measured to the nearest 0.5 cm and a sub-sample of foot tissue was taken and stored in 80% ethanol at 4°C. DNA extraction and quantification followed the protocols described in chapter 3.

Fourteen *P. australis* microsatellite loci were amplified in multiplex PCR reactions following the protocols described in chapter 2. PCR products were visualised and alleles scored following the protocols described in chapter 3. Each plate of 96 samples that was genotyped contained four previously genotyped samples as a positive control and to estimate levels of genotyping error. The mean error rate per allele was calculated following Pompanon *et al.* (2005).

4.2.2 Genetic diversity, differentiation and migration analyses

Data quality checks were carried out following the methods described in chapter 3. Briefly, genotype data was checked for scoring errors, large allele drop out, linkage disequilibrium and Hardy-Weinberg equilibrium (HWE). Allele discovery curves were calculated and simulations were run to check for F_{ST} outlier loci. The number of alleles, allelic range, proportion of null alleles and allele frequencies were calculated for each locus. The diversity statistics mean number of alleles (N_a), allelic richness (R_a), private alleles, observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding co-efficient (F_{IS}) were calculated for each location.

All population differentiation analyses were first performed using all loci, then were repeated excluding loci identified as outliers by LOSITAN (Antao *et al.* 2008) to determine any effects of outlier loci on genetic population structure. Global F_{ST}, pairwise F_{ST} and ϕ'_{ST} were calculated and tested for significance following the methods described in chapter 3. Tests were conducted to determine if the patterns of genetic population structure observed for other New Zealand coastal marine species (Gardner *et al.* 2010), were present for *P. australis* populations (i.e., no structure, IBD, divergence within and/or among populations, north-south divergence, east-west divergence). Mantel tests, principal component analyses (PCA) and analyses of molecular variance (AMOVA) were used to test for genetic population structure, as described in chapter 3.

Paphies australis individuals were clustered into genetically similar groups using the Bayesian method implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000) and the non-parametric method implemented in AWclust (Gao & Starmer 2008). Analyses were conducted using the methods and parameters described in chapter 3 except that the range of *K* values tested for *P. australis* in STRUCTURE was 1 to 13. Results from both clustering

analyses included all *P. australis* populations sampled and were conducted both including and excluding outlier loci.

Recent migration rates among *P. australis* locations were estimated using GENECLASS2 (Piry *et al.* 2004) and BAYESASS 3.0.3 (Wilson & Rannala 2003) using the same parameters as described in chapter 3. Long-term migration rates and effective population size (N_e) were estimated using MIGRATE 3.5.1 (Beerli 2006; Beerli & Felsenstein 2001) using the same parameters as described in chapter 3. Based on results of population differentiation analyses, all three migration estimation methods were carried out excluding outlier loci. Analyses were first performed to estimate migration among all sampling locations, then were repeated to estimate migration among population groupings identified as significantly differentiated by exact G (Goudet *et al.* 1996) and AMOVA (Excoffier *et al.* 1992) tests.

Table 4.1: Location, number of samples and geographical co-ordinates for *Paphies australis* samples used in this study.

Location	Abbreviation	Sample Size	Geographical Co-ordinates	
Raglan	RAG	55	-37.82° S	174.83° E
Huia	HUI	54	-37.01° S	174.57° E
Tapotupotu Bay	TAP	55	-34.43° S	172.71° E
Waiwera	WAI	54	-36.54° S	174.71° E
Tauranga	TAU	53	-37.66° S	176.13° E
Napier	NAP	50	-39.48° S	176.89° E
Petone	PET	56	-41.23° S	174.86° E
Hakahaka Bay	HAK	55	-41.31° S	174.11° E
Lyttelton	LYT	52	-43.64° S	172.75° E
Bluff	BLU	51	-46.57° S	168.49° E
Doubtful Sound	DBT	53	-45.28° S	166.91° E
Okuru	OKU	35	-43.89° S	168.92° E
Karamea	KAR	51	-41.26° S	172.11° E

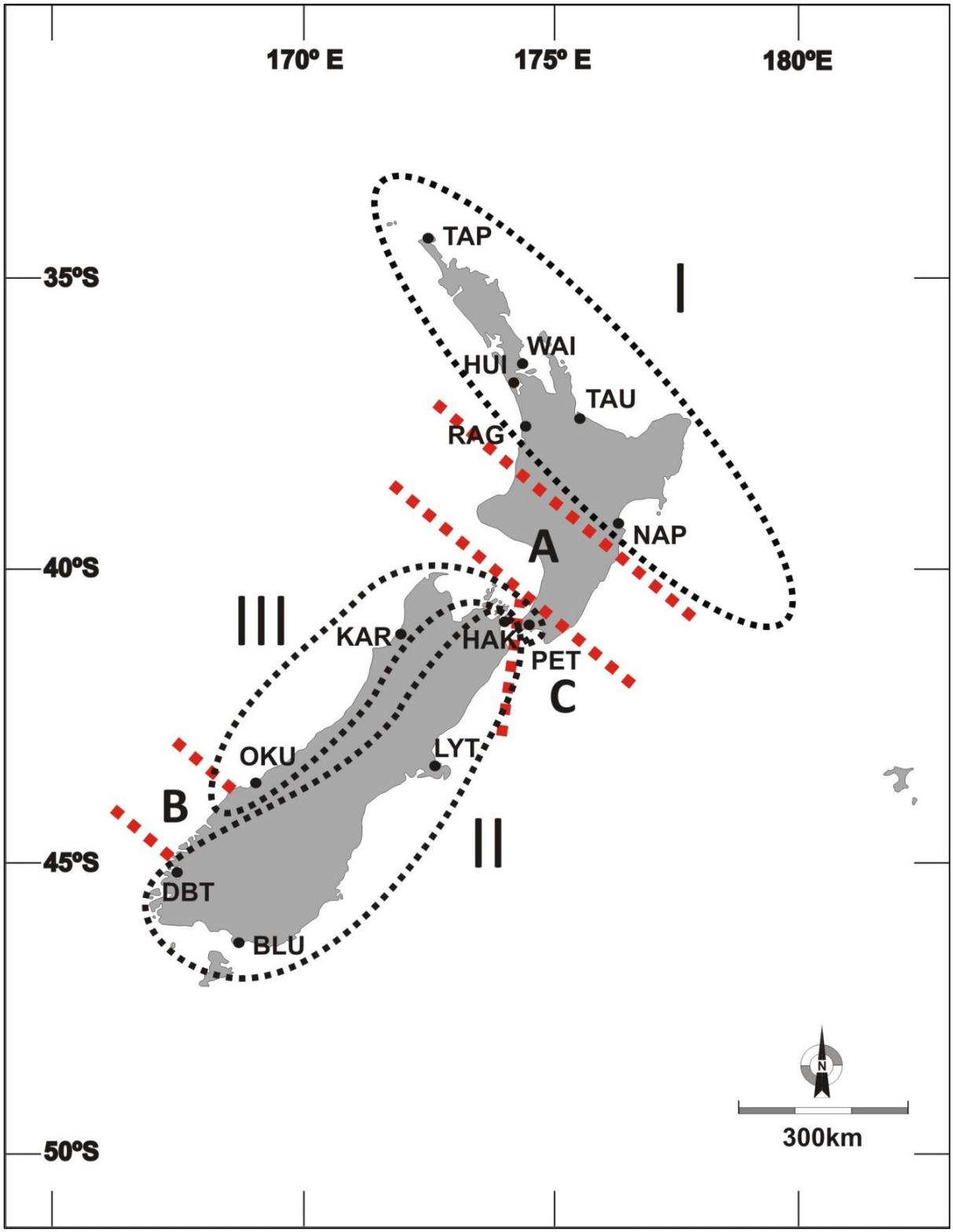


Figure 4.1: Collection locations for *Paphies australis.* Location abbreviations refer to those listed in Table 4.1. A, B and C denote putative genetic breaks detected in the present study, although the exact location of these breaks could not be pinpointed. The dashed red lines either side of the letters represent the possible geographic range where these genetic breaks may occur. I, II and III denote three significantly differentiated groups detected by AMOVA analyses. Dashed black lines indicate the geographic areas that encompass these grouping.

4.3 Results

4.3.1 Data quality checks and summary statistics

A total of 674 *P. australis* individuals from 13 locations were genotyped (Table 4.1; Figure 4.1). *Paphies australis* was found to be common and widespread around the main islands of New Zealand; seven North Island and six South Island locations were sampled for this study. Forty-five *P. australis* samples (6.7% of total samples) were repeat genotyped and the observed mean error rate was calculated at 3.7%. The majority of error (76%) was associated with alleles that had dropped out, rather than mistyping of alleles. One locus (*Paus_6*) was unable to be consistently genotyped and had a large number of null alleles so was excluded from all analyses.

MICROCHECKER found no evidence for genotype scoring errors or large allele drop out and there was no significant linkage disequilibrium among loci. One locus was in HWE for the total sample (*Paus_5*) and seven loci were not in HWE for any of the locations tested (*Paus_2, Paus_3, Paus_4, Paus_8, Paus_9, Paus_11, Paus_14*; Table 4.2). LOSITAN detected one outlier locus (*Paus_9*; Figure 4.2). Allele frequency graphs and discovery curves can be seen in appendices 2 and 3. For all loci allele discovery curves seem to have reached or are approaching an asymptote, suggesting that apart from a few rare alleles, most of the allelic diversity has been sampled for this species. The number of alleles per locus ranged from 11-48, the frequency of null alleles ranged from 0.011-0.295 and F_{IS} ranged from -0.078-0.664 (Table 4.3). For most loci observed heterozygosity was lower than expected except for one locus (*Paus_5*) where it was higher than expected (Range H_O: 0.240-0.660; H_E: 0.411-0.871; Table 4.3). As well as having similar levels of observed and expected heterozygosity, *Paus_5* was also the locus with the lowest proportion of null alleles and F_{IS} values closest to zero.

The mean number of alleles at each location ranged from 8.231-11.231, allelic richness ranged from 7.072-8.259 and the number of private alleles ranged from 1-10 (Table 4.4). In general the frequency of these private alleles was low (<2%; Appendix 1), suggesting that they represent 'rare alleles' that have not contributed significantly to genetic population variation. Observed heterozygosity was lower than expected for all locations (Range H_O: 0.378-0.497; H_E: 0.635-0.719) and no locations were in HWE (Table 4.4). F_{IS} was above zero for all locations due to an excess of homozygotes (0.284-0.456; Table 4.4).

Table 4.2: Hardy-Weinberg equilibrium (HWE) *p*-values for each locus and location for *Paphies australis*. Values in bold text represent significant departures from HWE expectation after false discovery rate (FDR) correction for multiple tests ($p < 0.05$). Location abbreviations as per Table 4.1.

Locus	Location													
	RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	Total
Paus_1	0.682	**<0.001**	0.259	**0.007**	**<0.001**	0.170	**<0.001**	**0.015**	0.345	0.051	0.317	**<0.001**	**0.001**	**<0.001**
Paus_2	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Paus_3	**<0.001**	**<0.001**	**<0.001**	**0.002**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.006**	**<0.001**	**<0.001**	**<0.001**
Paus_4	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Paus_5	0.886	0.712	0.586	0.851	0.774	0.650	0.753	0.499	0.650	0.975	0.835	0.945	0.814	1.000
Paus_7	0.164	**<0.001**	**<0.001**	**0.026**	**<0.001**	**0.006**	0.528	0.141	0.055	**0.009**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Paus_8	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Paus_9	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Paus_10	0.568	0.055	**0.001**	**<0.001**	**0.017**	**0.002**	**<0.001**	0.105	**<0.001**	**<0.001**	0.164	**0.001**	**0.001**	**<0.001**
Paus_11	**<0.001**	**<0.001**	**<0.001**	**0.023**	**<0.001**	**<0.001**	**0.019**	**0.014**	**<0.001**	**<0.001**	**0.002**	**<0.001**	**<0.001**	**<0.001**
Paus_12	0.047	**0.003**	**<0.001**	**0.008**	**0.008**	0.338	0.550	**<0.001**	**0.026**	0.073	0.259	**0.014**	**<0.001**	**<0.001**
Paus_13	0.845	**0.004**	**0.025**	0.058	**0.001**	0.085	0.079	0.061	0.318	0.697	1.000	**0.032**	1.000	**<0.001**
Paus_14	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.003**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

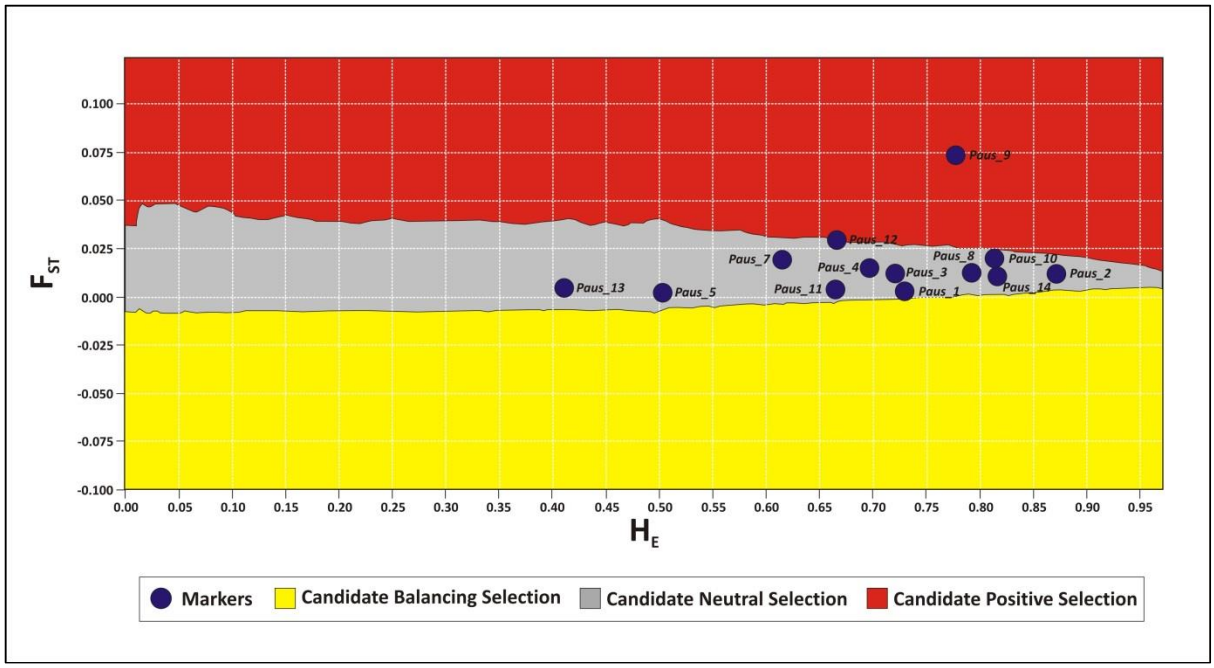


Figure 4.2: Output of LOSITAN analyses for outlier loci for *Paphies australis*. Blue dots represent the relationship between F_{ST} and H_E for each locus. Loci falling within the grey area represent 'neutral' loci, loci falling within the red or yellow areas represent F_{ST} outliers as determined by a 95% confidence interval.

Table 4.3: Diversity statistics for *Paphies australis* loci used in this study (N_a: number of alleles, H_O: observed heterozygosity, H_E: expected heterozygosity, N_F: estimated frequency of null alleles, F_{IS}: inbreeding co-efficient).

Locus	Statistic					
	N_a	Allele Size Range (bp)	H_O	H_E	N_F	F_{IS}
Paus_1	14	153-217	0.570	0.729	0.094	0.218
Paus_2	48	227-303	0.409	0.871	0.245	0.527
Paus_3	22	131-209	0.411	0.720	0.187	0.425
Paus_4	18	258-300	0.240	0.693	0.274	0.653
Paus_5	12	186-230	0.540	0.500	0.011	-0.078
Paus_7	11	86-110	0.431	0.612	0.118	0.288
Paus_8	22	210-276	0.265	0.791	0.294	0.664
Paus_9	16	103-169	0.246	0.758	0.295	0.661
Paus_10	19	164-216	0.606	0.813	0.110	0.242
Paus_11	11	218-263	0.395	0.662	0.159	0.405
Paus_12	13	173-209	0.660	0.662	0.047	-0.021
Paus_13	14	100-136	0.354	0.411	0.057	0.136
Paus_14	34	138-257	0.405	0.815	0.225	0.500

Table 4.4: Genetic diversity statistics for each location and total for *Paphies australis* (N_a: mean number of alleles, R_a: allelic richness, PA: private alleles, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{IS}: inbreeding co-efficient). F_{IS} values in bold indicate significant departure from HWE expectations after FDR correction ($p < 0.05$). Location abbreviations as per Table 4.1.

Location	Statistic					
	N_a	R_a	PA	H_O	H_E	F_{IS}
RAG	9.077	7.210	4	0.449	0.635	**0.289**
HUI	10.308	8.253	5	0.380	0.691	**0.439**
TAP	10.000	7.912	3	0.426	0.689	**0.376**
WAI	9.615	7.360	5	0.453	0.646	**0.284**
TAU	9.538	7.582	2	0.390	0.704	**0.436**
NAP	9.308	7.714	3	0.389	0.680	**0.416**
PET	9.462	7.659	4	0.400	0.699	**0.415**
HAK	11.231	8.259	10	0.463	0.672	**0.303**
LYT	9.769	7.789	6	0.482	0.699	**0.298**
BLU	10.692	8.270	10	0.497	0.700	**0.285**
DBT	9.000	7.072	3	0.448	0.669	**0.315**
OKU	8.231	7.498	3	0.380	0.719	**0.456**
KAR	9.154	7.700	1	0.378	0.709	**0.450**
Total	19.538	8.202	59	0.462	0.688	**0.372**

4.3.2 Genetic population differentiation

Estimates of global F_{ST} showed that correcting for null alleles seemed to have little effect on F_{ST} values, except for the outlier locus *Paus_9*, where corrected F_{ST} was much lower than uncorrected F_{ST}. Global F_{ST} ranged from 0-0.660 for *P. australis* loci and 0.002-0.016 when corrected for null alleles (Figure 4.3). Pairwise F_{ST} among *P. australis* locations was low, ranging from 0-0.030 when calculated using all loci and 0-0.024 when the outlier locus *Paus_9* was excluded (Table 4.5). Exact G tests indicated that 67 out of 78 pairwise F_{ST} comparisons showed significant differentiation when all loci were used and 62 out of 78 pairwise comparisons were significantly differentiated when *Paus_9* was excluded. Pairwise ϕ'_{ST} was higher than F_{ST} but showed a similar pattern of differentiation. ϕ'_{ST} ranged from 0-0.082 when calculated using all loci and from 0-0.060 when the outlier locus *Paus_9* was excluded (Table 4.6). AMOVA ϕ_{ST} tests for population differentiation indicated that 63 of the 78 pairwise ϕ'_{ST} comparisons were significantly differentiated when all loci were used and 51 of the 78 pairwise comparisons were significant if *Paus_9* was excluded. In general F_{ST}

and ϕ'_{ST} values were lower when the outlier locus was excluded from analysis, suggesting that this locus was inflating estimates of differentiation. In general F_{ST} and ϕ'_{ST} values were highest when comparing North Island locations to South Island locations. F_{ST} and ϕ'_{ST} were also higher among North Island locations than they were among South Island locations.

Consistent with the population differentiation results, Mantel tests showed a significant IBD pattern using both measures of F_{ST} when the outlier locus *Paus_9* was included, indicating that the further apart locations were, the more differentiated they became (*p*-value range 0.0002-0.0004; Figure 4.4A, B). When *Paus_9* was excluded the IBD pattern was still significant for F_{ST} when corrected for null alleles (p = 0.0224; Figure 4.4D) but was not significant for uncorrected F_{ST} (p = 0.1035; Figure 4.4C). A PCA using all loci showed that the X axis explained 29.92% of variation (p = 0.001) and the Y axis 19.17% of variation (p = 0.001; Figure 4.5A). Locations from the South Island grouped together and locations from the North Island grouped together, with samples from the Petone (at the bottom of the North Island) between the two groups. When the analysis was repeated without the *Paus_9* outlier locus the X axis explained 16.46% of variation (p = 0.036) and the Y axis 25.57% of variation (p = 0.001; Figure 4.5B). Similar patterns of genetic population structure were seen but Petone was grouped with South Island locations.

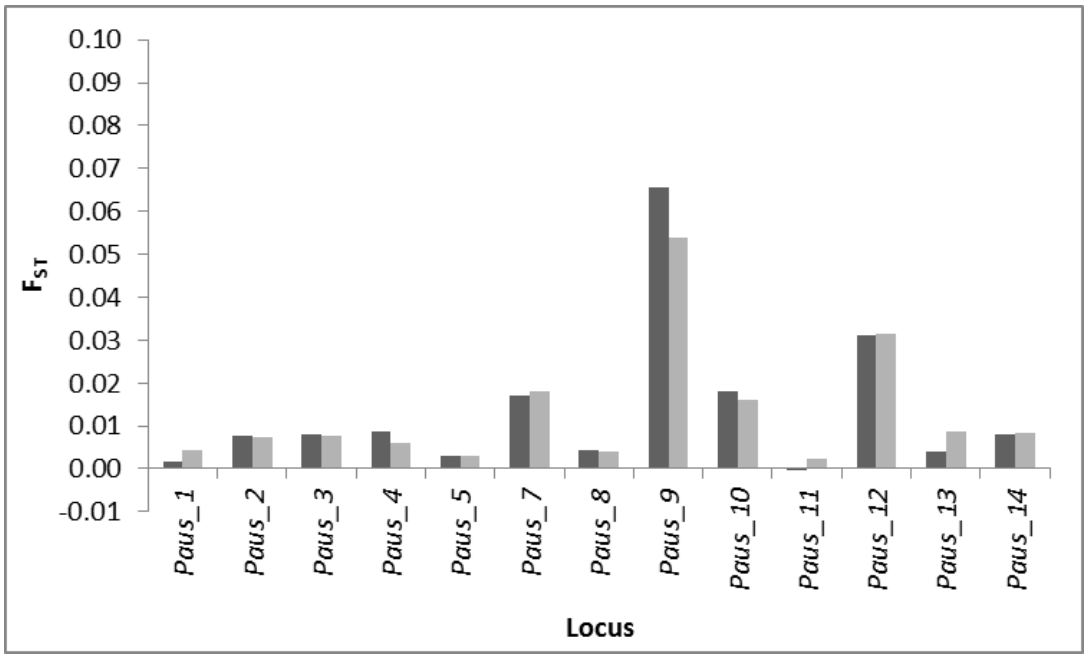


Figure 4.3: Global F_{ST} for each locus for *Paphies australis*. Dark grey bars represent F_{ST} uncorrected for null alleles, light grey bars represent F_{ST} corrected for null alleles using the algorithm described by Dempster *et al.* (1977).

Table 4.5: Pairwise F_{ST} values among *Paphies australis* locations using all loci (lower diagonal) and excluding the outlier locus *Paus_9* (upper diagonal). Values in bold represent significant genetic population differentiation after FDR correction for multiple tests ($p < 0.05$) as assessed by an exact G test (Goudet *et al.* 1996). Location abbreviations as per Table 4.1.

	RAG	**HUI**	**TAP**	**WAI**	**TAU**	**NAP**	**PET**	**HAK**	**LYT**	**BLU**	**DBT**	**OKU**	**KAR**
RAG	-	**0.017**	**0.009**	**0.003**	**0.023**	**0.005**	**0.012**	**0.011**	**0.006**	**0.006**	**0.006**	**0.011**	**0.019**
HUI	**0.016**	-	**0.013**	**0.020**	**0.007**	0.000	0.001	**0.021**	**0.010**	**0.015**	**0.015**	0.001	**0.008**
TAP	**0.009**	**0.012**	-	0.005	**0.008**	**0.004**	**0.009**	**0.013**	**0.006**	0.008	**0.010**	**0.006**	**0.012**
WAI	**0.004**	**0.020**	0.005	-	**0.022**	**0.011**	**0.017**	**0.010**	**0.006**	**0.012**	**0.013**	**0.016**	**0.023**
TAU	**0.022**	**0.007**	**0.007**	**0.019**	-	0.003	**0.011**	**0.024**	**0.018**	**0.022**	**0.022**	**0.006**	**0.013**
NAP	**0.005**	0.000	**0.004**	**0.010**	0.002	-	0.001	**0.011**	0.006	**0.009**	**0.011**	0.000	0.006
PET	**0.014**	0.002	**0.009**	**0.016**	**0.010**	0.001	-	**0.014**	0.001	**0.005**	**0.007**	0.000	0.000
HAK	**0.020**	**0.030**	**0.022**	**0.013**	**0.028**	**0.017**	**0.015**	-	**0.008**	**0.007**	**0.007**	**0.008**	**0.016**
LYT	**0.016**	**0.019**	**0.015**	**0.011**	**0.022**	**0.012**	0.003	**0.008**	-	**0.002**	**0.004**	0.000	0.003
BLU	**0.022**	**0.029**	**0.022**	**0.018**	**0.030**	**0.020**	**0.009**	**0.006**	**0.003**	-	**0.004**	**0.002**	**0.012**
DBT	**0.014**	**0.022**	**0.016**	**0.014**	**0.024**	**0.016**	**0.009**	**0.008**	**0.005**	**0.005**	-	**0.005**	**0.007**
OKU	**0.029**	**0.017**	**0.022**	**0.025**	**0.017**	**0.010**	**0.004**	**0.011**	0.003	**0.003**	**0.009**	**-**	0.001
KAR	**0.028**	**0.017**	**0.019**	**0.024**	**0.017**	0.012	0.002	**0.013**	0.002	**0.011**	**0.006**	0.003	-

Table 4.6: Pairwise ϕ'_{ST} values among *Paphies australis* locations using all loci (lower diagonal) and excluding the outlier locus *Paus_9* (upper diagonal). Values in bold represent significant genetic population differentiation after FDR correction for multiple tests ($p < 0.05$), as assessed by an AMOVA ϕ_{ST} test for population differentiation. Location abbreviations as per Table 4.1.

	RAG	**HUI**	**TAP**	**WAI**	**TAU**	**NAP**	**PET**	**HAK**	**LYT**	**BLU**	**DBT**	**OKU**	**KAR**
RAG	-	**0.043**	**0.022**	0.004	**0.060**	0.005	**0.030**	**0.022**	**0.013**	**0.012**	**0.014**	**0.022**	**0.048**
HUI	**0.037**	-	**0.025**	**0.048**	0.001	0.000	0.000	**0.044**	**0.019**	**0.032**	**0.037**	0.000	**0.005**
TAP	**0.018**	**0.020**	-	0.008	**0.007**	0.000	**0.015**	**0.026**	0.010	**0.015**	**0.023**	0.000	**0.020**
WAI	0.004	**0.044**	**0.006**	-	**0.051**	**0.018**	**0.043**	**0.019**	**0.012**	**0.026**	**0.032**	**0.034**	**0.055**
TAU	**0.052**	**0.000**	**0.002**	**0.040**	-	0.000	**0.020**	**0.051**	**0.046**	**0.056**	**0.058**	0.000	**0.019**
NAP	**0.004**	0.000	0.000	**0.015**	0.000	-	0.000	**0.014**	**0.006**	**0.014**	**0.026**	0.000	0.000
PET	**0.030**	0.000	**0.015**	**0.032**	**0.012**	0.000	-	**0.030**	0.000	0.007	**0.016**	0.000	0.000
HAK	**0.051**	**0.071**	**0.052**	**0.026**	**0.063**	**0.033**	**0.032**	-	**0.014**	**0.009**	**0.015**	**0.004**	**0.030**
LYT	**0.046**	**0.051**	**0.040**	**0.024**	**0.057**	**0.024**	0.000	**0.011**	-	0.000	0.009	0.000	0.000
BLU	**0.062**	**0.079**	**0.059**	**0.045**	**0.082**	**0.048**	**0.019**	**0.006**	0.000	-	0.007	0.000	**0.024**
DBT	**0.039**	**0.059**	**0.042**	**0.029**	**0.060**	**0.039**	**0.014**	**0.012**	**0.004**	0.006	-	0.009	**0.014**
OKU	**0.081**	**0.033**	**0.051**	**0.060**	**0.029**	**0.009**	0.000	**0.011**	0.000	0.000	**0.013**	-	0.000
KAR	**0.075**	**0.035**	**0.045**	**0.057**	**0.030**	**0.014**	0.000	**0.020**	0.000	**0.018**	**0.004**	0.000	-

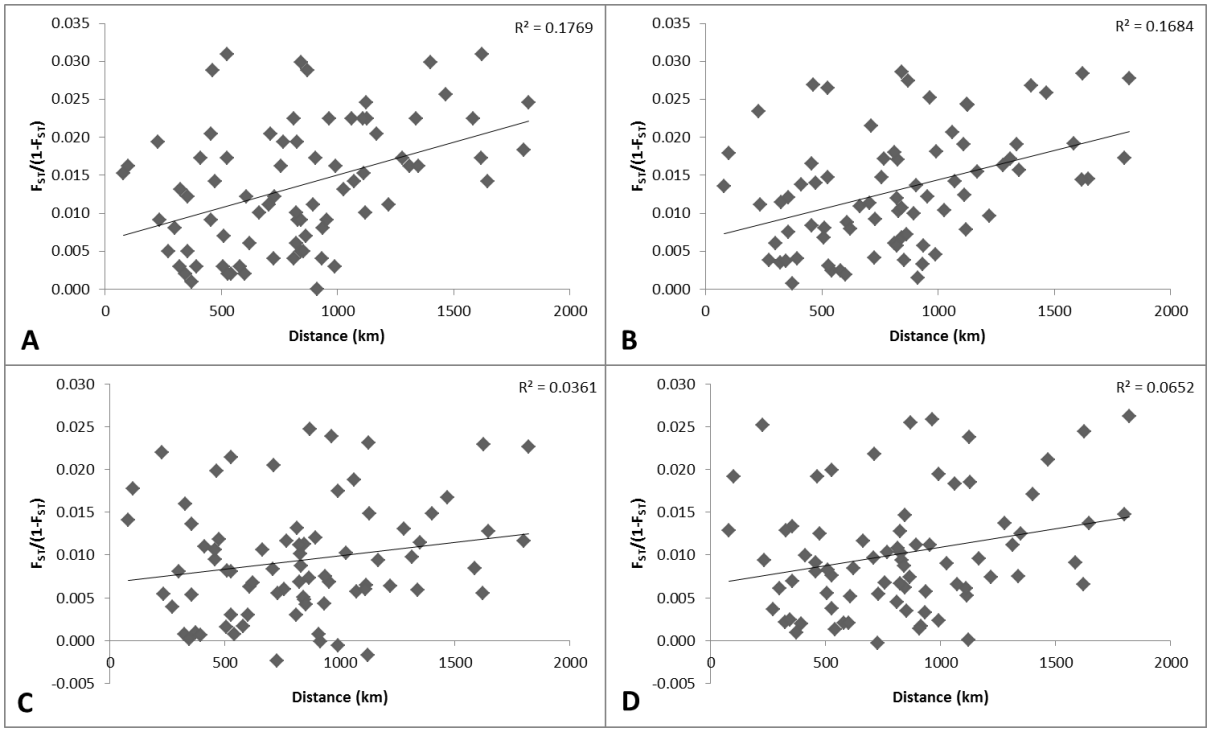


Figure 4.4: Results of Mantel tests for IBD for *Paphies australis* (significant if $p < 0.05$). A: based on F_{ST} for all loci ($p = 0.0004$); B: based on F_{ST} corrected for null alleles using the ENA method described in Chapuis and Estoup (2007) for all loci ($p = 0.0002$); C: based on F_{ST} for non-outlier loci ($p = 0.1035$); D: based on ENA corrected F_{ST} for non-outlier loci ($p = 0.0224$).

An AMOVA (Table 4.7) of all locations using all loci showed that 61% of variation was partitioned among individuals and 37% of variation among individuals within populations. Exclusion of the outlier locus *Paus_9* had little effect, with 64% of variation partitioned among individuals and 35% of variation among individuals within populations. There was significant support for a north-south grouping (All loci: {RAG, HUI, TAP, WAI, TAU, NAP, PET} vs {HAK, LYT, BLU, DBT, OKU, KAR}, $p < 0.001$; non-outlier loci: {RAG, HUI, TAP, WAI, TAU, NAP} vs {PET, HAK, LYT, BLU, DBT, OKU, KAR}, $p = 0.007$) but not for an east-west grouping (All loci: {TAP, HUI, RAG, PET, OKU, KAR} vs {WAI, TAU, NAP, HAK, LYT, BLU, DBT}, $p = 0.46$; non-outlier loci: {HUI, RAG, PET, KAR, OKU, DBT} vs {TAP, WAI, TAU, NAP, HAK, LYT, BLU}, $p = 0.218$). There was also strong support for further division of the southern locations; the arrangement of locations that explained the most variation among groups was {RAG, HUI, TAP, WAI, TAU, NAP} {HAK, LYT, BLU, DBT} {PET, KAR, OKU} ($p = 0.001$ both including and excluding *Paus_9*).

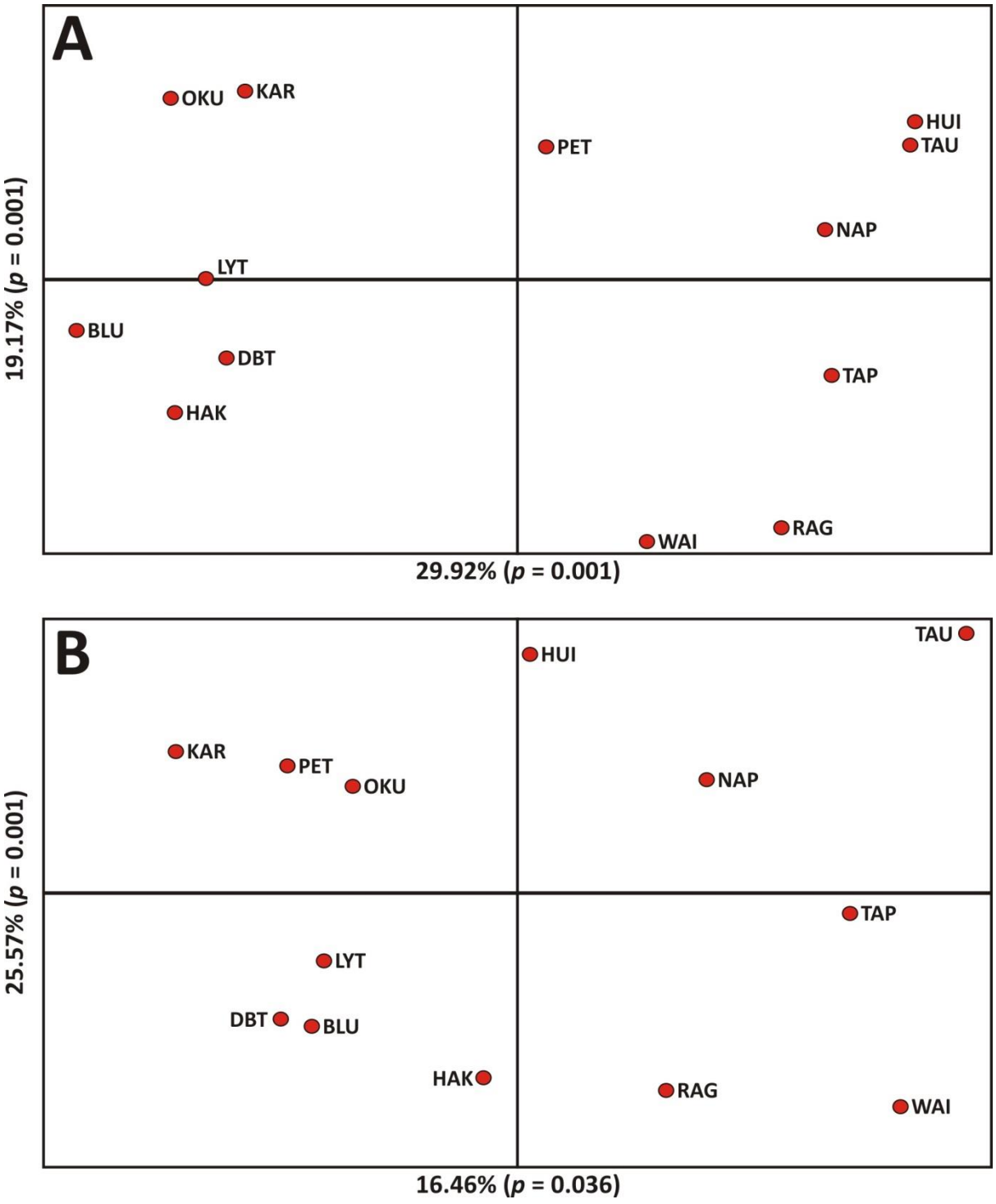


Figure 4.5: Principal component analysis (PCA) for *Paphies australis* showing patterns of genetic differentiation among locations A: using all loci; B: excluding outlier loci. The percentage of inertia explained by each axis and significance of the axis is displayed. An axis was considered significant if $p < 0.05$. Location abbreviations as per Table 4.1.

Table 4.7: AMOVA analyses for *Paphies australis*. Genetic population structure was considered significant if $p < 0.05$ after FDR correction for multiple tests.

Groupings Tested	**Group Configurations**	**Source of Variance** Among groups			Among populations within groups			Among individuals within populations			Among individuals		
		% Var.	F_{CT}	*p*-value	% Var.	F_{SC}	*p*-value	% Var.	F_{IS}	*p*-value	% Var.	F_{IT}	*p*-value
All loci													
1 group		-	-	-	1.416	0.014	**0.001**	37.196	0.377	**0.001**	61.388	0.386	**0.001**
North-South Division	{RAG, HUI, TAP, WAI, TAU, NAP, PET}{HAK, LYT, BLU, DBT, OKU, KAR}	1.149	0.011	**0.001**	0.800	0.008	**0.001**	36.995	0.377	**0.001**	61.056	0.389	**0.001**
East-West Division	{TAP, HUI, RAG, PET, OKU, KAR}{WAI, TAU, NAP, HAK, LYT, BLU, DBT}	-0.022	0.011	0.460	1.428	0.014	**0.001**	37.200	0.377	**0.001**	61.394	0.386	**0.001**
3 groups	{RAG, HUI, TAP, WAI, TAU, NAP}{HAK, LYT, BLU, DBT}{PET, KAR, OKU}	1.031	0.010	**0.001**	0.720	0.007	**0.001**	37.070	0.377	**0.001**	61.179	0.388	**0.001**
Non-outlier loci													
1 group		-	-	-	0.940	0.006	**0.005**	34.963	0.313	**0.001**	64.098	0.318	**0.001**
North-South Division	{RAG, HUI, TAP, WAI, TAU, NAP}{PET, HAK, LYT, BLU, DBT, OKU, KAR}	0.325	0.003	**0.007**	0.762	0.008	**0.001**	34.911	0.353	**0.001**	64.002	0.356	**0.001**
East-West Division	{HUI, RAG, PET, KAR, OKU, DBT}{TAP, WAI, TAU, NAP, HAK, LYT, BLU}	0.075	0.001	0.218	0.899	0.009	**0.001**	34.951	0.353	**0.001**	64.075	0.359	**0.001**
3 groups	{RAG, HUI, TAP, WAI, TAU, NAP}{HAK, LYT, BLU, DBT}{PET, KAR, OKU}	0.387	0.004	**0.001**	0.674	0.007	**0.001**	34.920	0.353	**0.001**	64.019	0.360	**0.001**

STRUCTURE analyses revealed seven genetic clusters for *P. australis* when all loci were used and only two clusters when *Paus_9* was excluded (Figure 4.6A, B). Each location showed a high degree of cluster admixture and the proportion of each cluster varied among locations (Table 4.8; Figure 4.7A, B). There was no obvious geographic pattern to the distribution of clusters, although for the analysis using all loci cluster 5 had a significant north-south cline (Spearman's *r* -0.84; *p* = 0.0004) while clusters 1 and 2 showed a significant east-west cline (Spearman's *r* -0.91, 0.73; *p* = 0.00002-0.0005). AWclust analyses produced similar results to SRUCTURE. Gap statistics revealed six genetic clusters for *P. australis* when all loci were used and only two clusters when *Paus_9* was excluded (Figure 4.6C, D). AWclust also showed high levels of cluster admixture among locations (Table 4.8; Figure 4.7C, D). When the analysis was conducted using all loci, clusters 2, 3 and 4 showed a significant north-south cline (Spearman's *r* -0.81-0.86; *p* =0.0001-0.002).

Figure 4.1 summarises the findings of analyses for genetic population structure in *P. australis* and illustrates the location of putative population groupings and genetic breaks. In summary, *P. australis* in New Zealand can be divided into three significantly different genetic groups with three putative barriers to dispersal (A – C) between these groups:

I. Northern group consisting of Raglan, Huia, Tapotupotu Bay, Waiwera, Tauranga and Napier
II. South Eastern group consisting of Hakahaka Bay, Lyttelton, Bluff and Doubtful Sound
III. South Western group consisting of Petone, Karamea and Okuru.

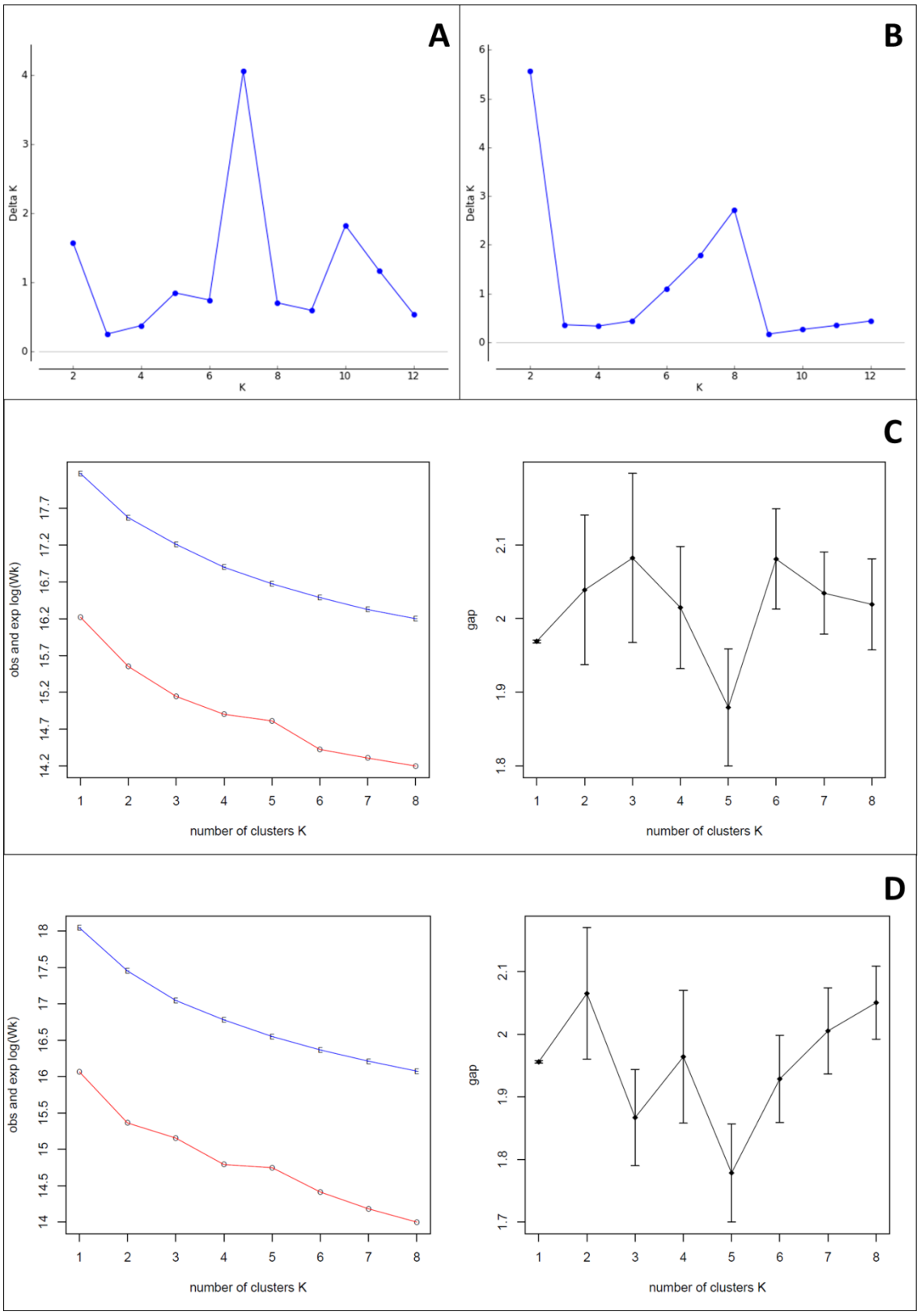


Figure 4.6: Plots to determine optimum *K* values for STRUCTURE and AWclust analyses. Delta*K* values were derived from the Evanno *et al.* (2005) method and generated in STRUCTURE HARVESTER (Earl & vonHoldt 2011) to determine optimal *K* values for *Paphies australis* A: using all loci (*K* = 7) and B: using non-outlier loci (*K* = 2). Gap statistics generated from AWclust (Gao & Starmer 2008) to determine optimal *K* values for *P. australis* C: using all loci (*K* = 6) and D: using non-outlier loci (*K* = 2).

Table 4.8: Proportion of clusters found in *Paphies australis* locations from STRUCTURE and AWclust analyses. Cluster colours relate to those used in Figure 4.7.

Analysis	Cluster	Location														
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	Total
STRUCTURE: *Paphies australis* - all loci															
	1	0.080	0.077	0.098	0.078	0.080	0.058	0.078	0.095	0.090	0.117	0.223	0.136	0.114	0.102
	2	0.126	0.215	0.203	0.118	0.351	0.204	0.190	0.098	0.094	0.076	0.081	0.097	0.140	0.153
	3	0.090	0.150	0.195	0.137	0.153	0.103	0.100	0.253	0.118	0.254	0.067	0.123	0.086	0.141
	4	0.026	0.221	0.048	0.034	0.102	0.163	0.197	0.044	0.103	0.028	0.061	0.252	0.135	0.109
	5	0.054	0.052	0.034	0.054	0.040	0.089	0.121	0.100	0.191	0.143	0.161	0.113	0.236	0.107
	6	0.190	0.072	0.116	0.229	0.136	0.086	0.147	0.323	0.213	0.246	0.184	0.227	0.144	0.178
	7	0.434	0.213	0.307	0.350	0.138	0.297	0.167	0.088	0.191	0.137	0.222	0.052	0.145	0.211
STRUCTURE: *Paphies australis* - non-outlier loci															
	1	0.238	0.641	0.408	0.229	0.672	0.514	0.543	0.246	0.363	0.282	0.262	0.527	0.500	0.417
	2	0.762	0.359	0.592	0.771	0.328	0.486	0.457	0.754	0.637	0.718	0.738	0.474	0.500	0.583
AWclust: *Paphies australis* - all loci															
	1	0.309	0.093	0.164	0.222	0.094	0.160	0.143	0.200	0.231	0.216	0.321	0.257	0.255	0.205
	2	0.236	0.278	0.273	0.222	0.132	0.220	0.232	0.182	0.212	0.118	0.151	0.114	0.078	0.188
	3	0.273	0.259	0.255	0.259	0.264	0.300	0.179	0.091	0.250	0.078	0.208	0.057	0.118	0.199
	4	0.109	0.130	0.109	0.167	0.132	0.140	0.179	0.346	0.154	0.353	0.245	0.229	0.196	0.191
	5	0.055	0.111	0.091	0.111	0.245	0.100	0.179	0.146	0.096	0.196	0.057	0.200	0.216	0.139
	6	0.018	0.130	0.109	0.019	0.132	0.080	0.089	0.036	0.058	0.039	0.019	0.143	0.137	0.078
AWclust: *Paphies australis* - non-outlier loci															
	1	0.727	0.482	0.527	0.537	0.377	0.520	0.625	0.546	0.673	0.608	0.717	0.400	0.471	0.555
	2	0.273	0.519	0.473	0.463	0.623	0.480	0.375	0.455	0.327	0.392	0.283	0.600	0.529	0.445

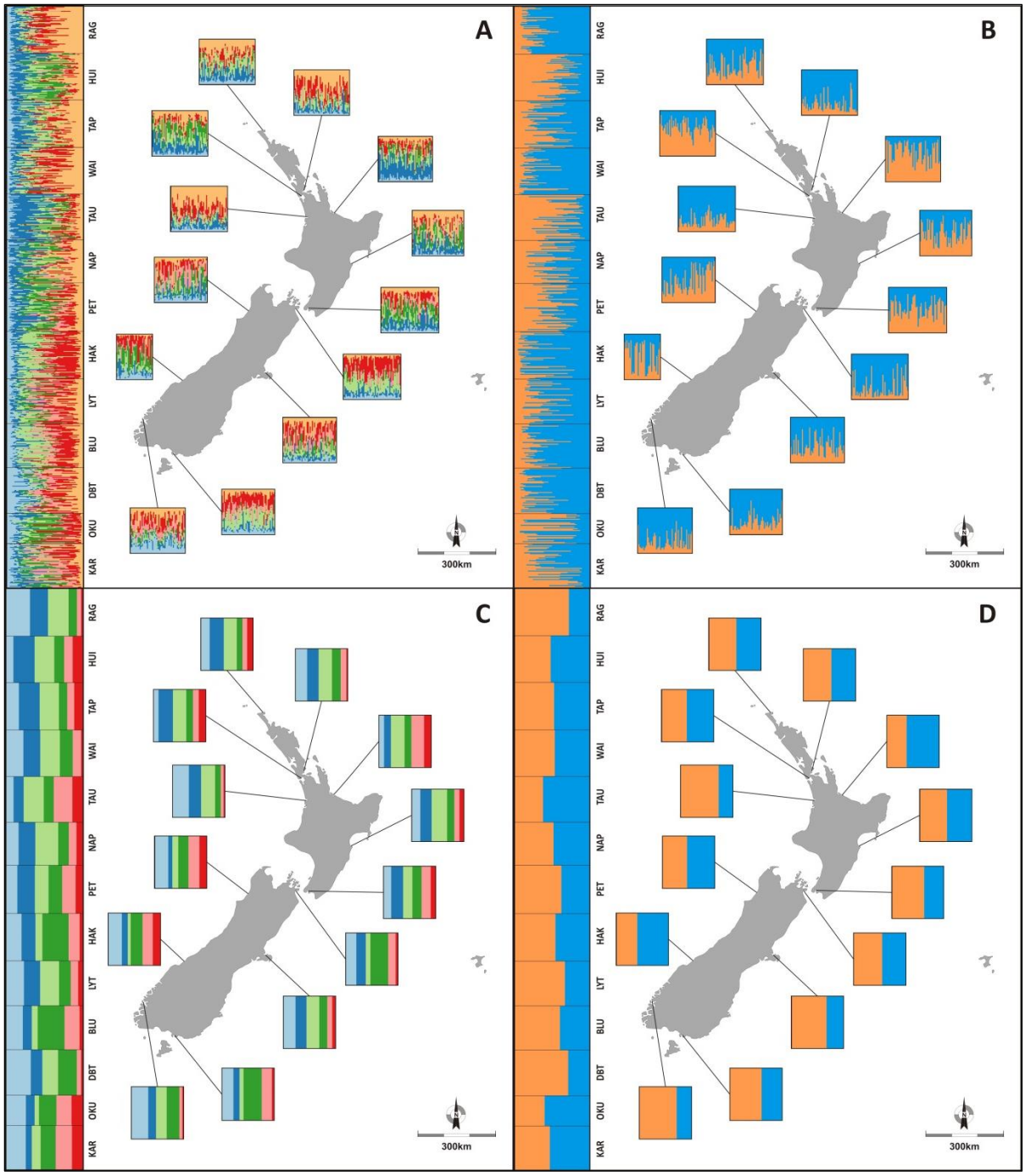


Figure 4.7: Output from cluster analyses for *Paphies australis* showing the proportion of each cluster assigned to each location. Each colour denotes a different cluster as specified in Table 4.8. A: STRUCTURE analysis using all loci (*K* = 7); B: STRUCTURE analysis using non-outlier loci (*K* = 2); C: AWclust analysis using all loci (*K* = 6); D: AWclust analysis using non-outlier loci (*K* = 2). For STRUCTURE analyses each line represents an individual individuals can belong to multiple clusters. For AWclust analyses individuals are assigned to a single cluster, the width of the bars indicates the proportion of each cluster allocated to that location.

4.3.3 Estimates of contemporary and historical migration

GENECLASS2 detected low levels of contemporary migration; the number of first generation migrants (F0) with a probability less than 0.01 for all *P. australis* locations was 52 (8% of individuals; Table 4.9A). The number of F0 migrants detected at each location ranged from one to seven. Group I locations were the largest source of migrants (33/52); group II and III locations supplied modest numbers of F0 migrants. As discussed in chapter 3, the true number of F0 migrants could be higher, as this test may not have sufficient power to detect all F0 migrants. Repeating this analysis for the three differentiated population groups identified 23 F0 migrants (3% of individuals sampled; Table 4.9B). The number of migrants exchanged was low but similar among the three groups, ranging from three to four.

BAYESASS identified limited source populations for migrants and migration rates showed high spatial variation, ranging from 3-26% (Figure 4.8A). Self-recruitment rates ranged from 67-93%. Raglan and Huia were the main source locations, but Petone and Hakahaka Bay supplied lower numbers of migrants to some locations. For reasons described in chapter 3, some migration rates could be underestimates due to lack of power and the restrictive assumptions of the analysis. Migration estimates among the three differentiated groups showed high levels of migration from group I to III (27%; Figure 4.8B). Groups I and II exchanged similar amounts of migrants (5-7%) and 6% of group III individuals originated from group II. Self-recruitment was high for groups I and II (93-95%) but lower for group III (67%; Figure 4.8B).

A large range of long-term migration rates were estimated by MIGRATE but migration was detected among all locations. Migration rates ranged from 2.50-38.61 migrants per generation (Table 4.10A). The population size parameter θ ranged from 0.07-0.46. Using a mutation rate of 1×10^{-4} this equated to an effective population size range of 175.00 to 1,158.33 individuals. Long-term migration rates estimated among the three genetically differentiated groups ranged from 24.83-72.41 migrants per generation (Table 4.10B). Group I was the largest source of migrants, supplying 72.41 and 57.63 migrants per generation to groups II and III, respectively. Large numbers of migrants were also exchanged among the two southern groups (II→III: 42.17; III→II: 34.77) but there was lower migration from the two southern groups to the north (II→I: 29.39, III→I: 24.83). θ ranged from 0.38-0.50, resulting in effective population sizes that ranged from 958.33 to 1,241.68 individuals.

Table 4.9: Results of an assignment test to detect first generation (F0) migrants at each *Paphies australis* location, conducted in GENECLASS2. Migrants were detected using the *Lh/Lhmax* statistic with a probability < 0.01. A: Migrants exchanged among all locations. B: Migrants exchanged among genetically differentiated population groups as determined by exact G and AMOVA tests.

A

Receiving Location	Source of F0 Migrants														Total F0 Received
	RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR		
RAG			1			1									2
HUI				1		1				1	1				4
TAP		1			1								1		3
WAI	1		1			1	2								5
TAU		1				1			1			1			4
NAP	1		1						1						3
PET	1	1		1									3		6
HAK		1			1	1				1		1	1		6
LYT		1					1				1				3
BLU				1											1
DBT		3		1	1	2									7
OKU			1					1	1	1					4
KAR		1	2	1											4
Total F0 Supplied	3	9	6	5	3	7	3	1	3	3	2	2	5		52

B

Receiving Location	Source of F0 Migrants			Total F0 Received
	I: Northern	II: South East	III: South West	
I: Northern		4	5	9
II: South East	3		4	7
III: South West	4	3		7
Total F0 Supplied	7	7	9	23

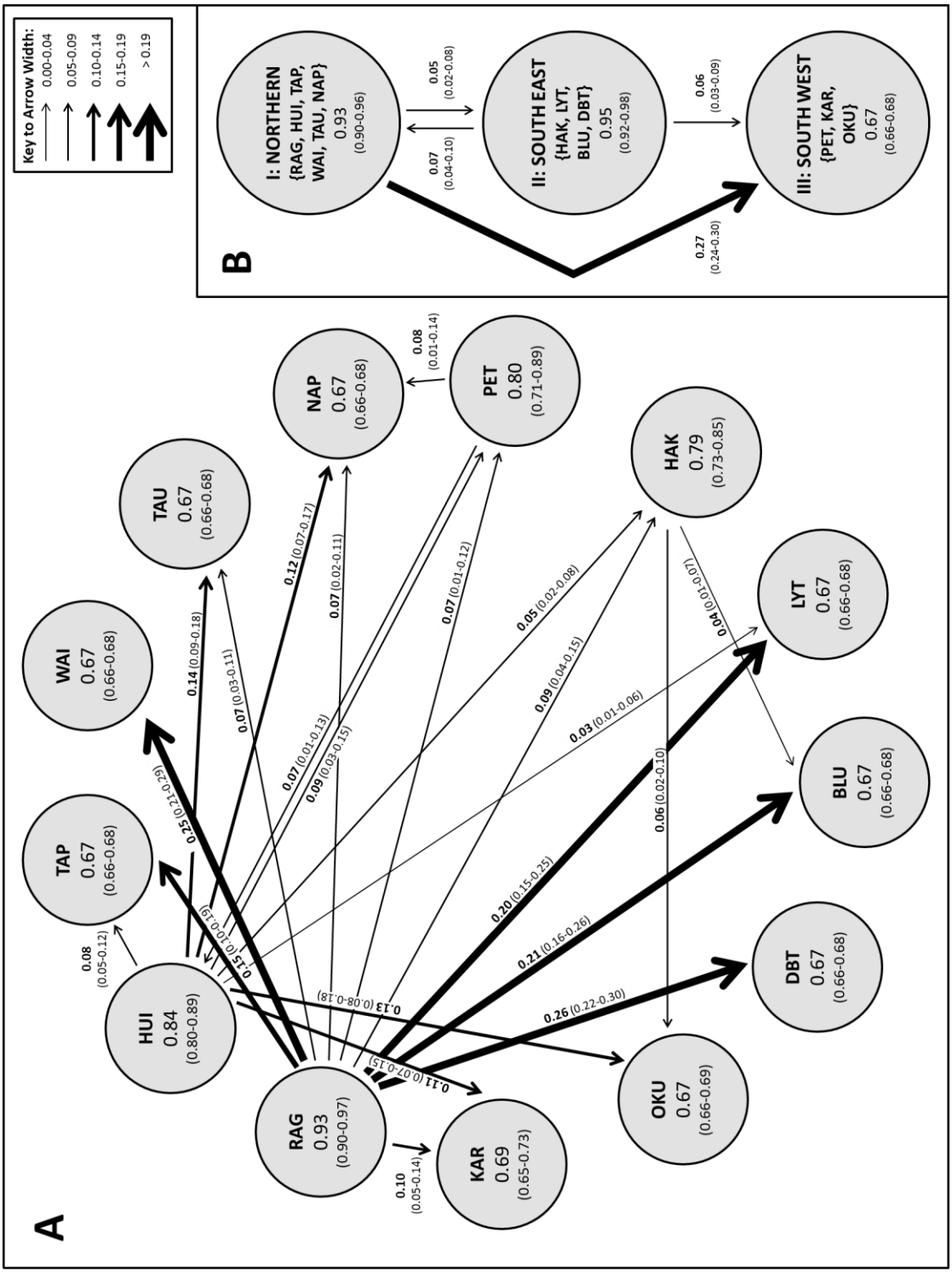


Figure 4.8: Recent migration rates for *Paphies australis* as determined by BAYESASS. Values in grey circles represent self-recruitment rates for each location (95% confidence intervals in brackets). Arrows represent migration among populations with the proportion of migrants indicated in bold (95% confidence interval in brackets). Arrow thickness indicates the relative contribution of migrants from each putative source location. A: Migration among all locations; B: Migration among genetically differentiated population groups as determined by exact G and AMOVA tests.

Table 4.10: Estimates of the number of migrants per generation (N_em), the population size parameter theta (θ), their credible intervals and effective population size (N_e, using the mutation rate 1 x 10^{-4}) for *Paphies australis*. A: Migrants exchanged among all locations. B: Migrants exchanged among genetically differentiated population groups as determined by exact G and AMOVA tests.

A

Receiving Location	Source Location												θ	N_e
	RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
RAG		15.83	8.50	2.50	8.81	21.21	11.50	9.97	4.03	10.81	2.57	8.52	19.50	0.21
		0-119.46	0-86.13	0-47.69	0-77.16	0-104.00	0-73.61	0-68.80	0-53.33	0-120.71	0-43.82	0-75.78	0-118.04	0.04-0.41
HUI	10.50		6.50	4.17	8.81	11.99	6.50	5.63	4.77	16.99	3.97	9.83	19.50	0.32
	0-86.75		0-78.53	0-52.62	0-77.16	0-89.56	0-61.33	0-59.25	0-55.11	0-129.33	0-49.87	0-78.22	0-118.04	0.09-0.56
TAP	11.90	13.72		3.61	10.17	13.83	8.50	5.63	4.77	23.17	3.03	8.52	22.10	0.15
	0-89.47	0-115.74		0-50.98	0-79.64	0-92.44	0-67.47	0-59.25	0-55.11	0-137.95	0-46.85	0-75.78	0-121.73	0-0.38
WAI	13.30	20.05	7.50		7.46	19.37	12.50	7.37	4.77	23.17	3.50	11.15	11.70	0.08
	0-92.18	0-126.94	0-83.60		0-72.18	0-101.11	0-75.65	0-63.07	0-55.11	0-137.95	0-48.36	0-80.67	0-106.98	0-0.25
TAU	13.30	13.72	8.50	5.83		13.83	9.50	8.23	4.77	35.52	3.97	13.77	19.50	0.20
	0-94.89	0-115.74	0-83.60	0-57.56		0-92.44	0-69.51	0-64.98	0-55.11	0-155.20	0-48.36	0-88.00	0-118.04	0.01-0.37
NAP	10.50	11.61	9.50	4.17	10.17		8.50	7.37	5.50	26.26	3.97	9.83	22.10	0.28
	0-86.75	0-112.00	0-86.13	0-52.62	0-79.64		0-65.42	0-64.98	0-56.89	0-142.27	0-49.87	0-78.22	0-121.73	0-0.43
PET	17.50	30.61	12.50	4.17	12.88	17.52		8.23	9.17	26.26	4.43	9.83	11.70	0.15
	0-100.31	0-141.86	0-93.73	0-52.62	0-84.62	0-98.22		0-64.98	0-65.78	0-142.27	0-52.89	0-78.22	0-106.98	0-0.31
HAK	13.30	24.28	8.50	5.28	8.81	11.99	15.50		4.77	38.61	5.37	15.08	16.90	0.13
	0-92.18	0-134.40	0-83.60	0-55.91	0-77.16	0-89.56	0-79.73		0-56.89	0-159.51	0-55.91	0-88.00	0-114.36	0-0.29
LYT	16.10	32.72	5.50	6.39	8.81	17.52	8.50	12.57		29.34	4.43	12.46	14.30	0.11
	0-100.31	0-149.34	0-73.47	0-59.20	0-77.16	0-101.11	0-67.47	0-74.53		0-146.58	0-51.38	0-83.11	0-110.67	0-0.27
BLU	9.10	11.61	6.50	3.61	7.46	11.99	8.50	5.63	4.03		4.90	9.83	16.90	0.46
	0-84.05	0-112.00	0-78.53	0-49.33	0-74.67	0-89.56	0-67.47	0-59.25	0-51.56		0-52.89	0-78.22	0-114.36	0.27-0.65
DBT	17.50	17.95	8.50	6.39	12.88	21.21	10.50	10.83	8.43	35.52		13.77	16.90	0.07
	0-103.02	0-126.94	0-83.60	0-59.20	0-84.62	0-104.00	0-73.60	0-70.71	0-65.78	0-155.20		0-85.56	0-114.36	0-0.23
OKU	11.90	17.95	11.50	5.83	12.88	17.52	9.50	11.70	5.50	13.90	4.43		35.10	0.20
	0-92.18	0-126.94	0-91.20	0-59.20	0-87.11	0-101.11	0-69.51	0-72.62	0-56.89	0-120.71	0-52.89		0-143.87	0.01-0.37
KAR	9.10	13.72	7.50	3.06	7.46	11.99	6.50	4.77	4.77	16.99	3.03	11.15		0.39
	0-84.05	0-119.46	0-81.07	0-47.69	0-74.67	0-89.56	0-61.33	0-57.33	0-55.11	0-129.33	0-46.85	0-83.11		0.01-0.55

B

Receiving Location	Source Location			θ	N_e
	I: Northern	II: South East	III: South West		
I: Northern		29.39	24.83	0.44	1108.33
		0-139.20	0-150.75	0.27-.061	
II: South East	72.41		34.77	0.38	958.33
	0-198.18		0-164.88	0.17-0.58	
III: South West	57.63	42.17		0.50	1241.68
	0-186.05	0-162.4		0.26-0.71	

4.4 Discussion

Increasing evidence for estuarine species suggests that their habitat characteristics can restrict dispersal opportunities, leading to genetic differentiation among populations inhabiting different estuaries (Watts & Johnson 2004). This chapter assessed this feature for populations of a New Zealand estuarine bivalve, *Paphies australis*, by estimating levels of genetic diversity, population differentiation and migration for thirteen populations encompassing both the North and South Islands of New Zealand. The results of these analyses (i.e., the three population groups that were detected and significant F_{ST} comparisons over small spatial scales) provide evidence for restricted dispersal among some locations, disproving the null hypothesis of panmictic genetic structure. However, inconsistent patterns of genetic differentiation and evidence for long-range migration among some locations suggest that factors other than simple geographic separation of habitat (e.g., oceanographic processes) may be influencing patterns of genetic population structure in this species.

4.4.1 Genetic diversity

High F_{IS} and homozygote excess was observed in *P. australis*, indicating widespread departures from HWE for all locations tested, similar to the patterns seen in *P. subtriangulata*. As with *P. subtriangulata*, the high frequency of null alleles was probably the main reason for the observed HWE departures, but undetected spatial or temporal population differentiation, and other reproductive characteristics of the species could also have contributed (as discussed in chapter 3, but see also Addison & Hart 2005). Null alleles could be affecting the genetic population structure reported here for *P. australis* as IBD analyses showed increased slope of regression lines when F_{ST} was corrected for null alleles. Despite this, there was still sufficient power to detect significant patterns of population differentiation in *P. australis*. One locus (*Paus_9*) had very high F_{ST} even when corrected for null alleles and was identified as an outlier locus. This locus could be subject to positive selection, which is inflating estimates of genetic population differentiation and structure as a consequence. For this reason the most conservative approach is to base conclusions on analyses that exclude this locus.

Levels of genetic diversity in *P. australis* were lower than what has been reported for other estuarine bivalve species (e.g., Martínez *et al.* 2009; Melo *et al.* 2012; Tarnowska *et al.*

2010) but higher than what was reported in chapter 3 for *P. subtriangulata* in terms of N_a, R_a and H_E. It is difficult to determine if the higher genetic diversity observed in *P. australis* (compared to *P. subtriangulata*) is a true indicator that populations of this species are more diverse. Alternatively it could be an artefact of the different markers used, differences in population size or the higher number of locations/individuals sampled for *P. australis* over a larger geographic range (Selkoe & Toonen 2006).

4.4.2 Genetic population structure

In relation to the patterns described for other New Zealand coastal marine species by Gardner *et al.* (2010), the overall genetic population structure in *P. australis* can be described as an isolation by distance pattern. This result suggests that at a large scale, the traditional estuarine species paradigm that habitat availability controls population structure is applicable to *P. australis*. However, it is difficult to assess the importance of this relationship as Mantel tests for IBD appeared to be influenced by null alleles, and PCA and AMOVA analyses suggested that hierarchical structure could also be present. The IBD pattern was further confounded by significant F_{ST} values over small geographic distances. Tests for IBD are known to be confounded by hierarchical structure, and vice versa, making it challenging to tease apart true population genetic structure in such instances (Meirmans 2012).

A stepping-stone model of connectivity, whereby populations in close geographic proximity exchange more migrants (Kimura & Weiss 1964), is thought to characterise many populations of coastal marine species, especially those inhabiting discrete areas of habitat such as estuaries. In terms of genetic population structure, stepping-stone connectivity should manifest as an IBD pattern because distance will increasingly act as a barrier to larval dispersal as populations become more remote. The exact relationship between geographic distance and genetic differentiation will depend on the configuration of the stepping-stones, migration rate, mutation rate and whether populations have reached migration-drift equilibrium (Hellberg *et al.* 2002). The basic stepping-stone model assumes that populations are continuously distributed over their range so that there is a consistent relationship between distance and dispersal (Slatkin 1993). This basic model can be modified to incorporate populations that occupy discontinuous habitat (i.e., harbour and estuarine habitat) if the distance between habitat is less than the ‘saturation distance’ that a species

can potentially dispersal (Rousset 1997). However, the relationship between genetic and geographic distance can become decoupled when distance between habitat is large, migration-drift equilibrium has not been reached or other controlling factors, such as currents, become complex (Selkoe & Toonen 2011). Decoupling of this relationship appears to have occurred among the *P. australis* locations sampled for this study.

Isolation by distance patterns have been commonly reported among New Zealand estuarine organisms (e.g., Hickey *et al.* 2009; Perrin *et al.* 2004; Stevens & Hogg 2004). However, many of these studies report enhanced population differentiation on smaller geographic scales and several lines of evidence suggest that this is also the case for *P. australis*. Principal component analysis and AMOVA provide evidence for medium scale hierarchical structure (i.e., the three genetically differentiated population groups). Results from these analyses were consistent with the genetic discontinuity observed among many New Zealand coastal marine species around the location of the Cook Strait. However, reports on the location of this discontinuity differ (Gardner *et al.* 2010; Ross *et al.* 2009). A division at approximately 39° S on the east coast of the North Island has been reported for some species, e.g., amphipods (Stevens & Hogg 2004), snapper (Bernal-Ramírez & Adcock 2003) and cockles (Ross *et al.* 2011). However, for other species a division at approximately 42° S (location of Cape Campbell on the east coast of the South Island) has been reported, e.g., cushion stars (Ayers & Waters 2005; Waters & Roy 2004), limpets (Goldstien *et al.* 2006) and greenshell mussels (Apte *et al.* 2003; Star *et al.* 2003). The inclusion of Petone with other South Island locations suggests that for *P. australis* this division occurs somewhere between 39.5° S and 41° S. However, Cook Strait may also act as a barrier to dispersal as AMOVA analyses indicated that Petone and Hakahaka Bay are in different groups and significantly differentiated, despite their close proximity to each other.

Interestingly, analyses using the outlier locus *Paus_9* suggested that Petone was more closely aligned to other North Island populations; further confounding attempts to pinpoint the exact location of this north-south division. This locus strongly affected genetic population structure and is possibly under selection, suggesting that this barrier around the Cook Strait area is placing selective pressure on certain parts of the *P. australis* genome. A similar effect has been documented in the greenshell mussel *Perna canaliculus* (Wei *et al.* 2013a) where a locus that is apparently under selection shows strong north-south

differentiation of populations either side of the Cook Strait region. Further analysis of *P. australis* populations from these latitudes and more detailed investigation of this potential outlier locus would be required to pinpoint the exact location of this barrier and gain insight to the selective pressure it has on marine species.

The genetic break on the south-western coast of the South Island between Okuru and Doubtful Sound is similar to what was reported for the cockle *Austrovenus stutchburyi* (Ross *et al.* 2011) and the estuarine triplefin *Grahamina nigripenne* (Hickey *et al.* 2009). Fiordland populations are known to be genetically distinct for some species (e.g., sea urchins, Mladenov *et al.* 1997; sea stars, Perrin *et al.* 2004) but it is more common for genetic breaks to be reported on the east coast of the South Island (e.g., Hickey *et al.* 2009; Ross *et al.* 2009; Will *et al.* 2011). This west coast genetic break that is seen in *P. australis* appears to be unusual among New Zealand coastal marine species.

Significant pairwise F_{ST} and ϕ'_{ST} comparisons observed among locations within the three *P. australis* groups suggest that further small scale patterns of 'chaotic' genetic patchiness are present among some locations. For example, significant genetic differentiation among some North Island locations shows that differentiation can occur over small spatial scales of 100-200 km. Both statistics were in agreement that a larger amount of divergence exists between North and South Island locations, as well as among North Island locations, and divergence among South Island locations was lower by comparison. This observation implies more restricted dispersal among North Island locations and higher dispersal among South Island locations.

Overall there is evidence that the stepping stone model of gene flow that leads to IBD genetic population structure seems to be operating among many of the *P. australis* locations that were sampled in this study. However it appears that barriers to dispersal exist that lead to the breakdown of stepping stone migration and IBD patterns, and drive the hierarchical structure detected by PCA and AMOVA analysis. Furthermore, the Cook Strait and Fiordland regions do not seem to represent absolute barriers to gene flow among the three groups, as some locations showed low levels of genetic differentiation despite belonging to different population groups. Despite showing some evidence for a north-south cline, clustering analyses appeared to lack the sensitivity required to detect the patterns of

genetic population structure uncovered by other analyses. However clustering algorithms such as STRUCTURE are known to perform poorly when IBD structure is present (Schwartz & McKelvey 2008). The admixture of clusters observed among all locations suggests that all locations are connected to some degree. These findings are in agreement with historical rates of genetic connectivity estimated by MIGRATE, which showed that all locations were connected by migration when averaged over long periods of time. The hierarchical structure and small scale 'chaotic' patterns of genetic structure provide evidence that, while habitat availability might be important for driving overall IBD genetic structure, variability in oceanographic processes at smaller scales can restrict dispersal and create complex patterns of genetic population structure.

Interestingly, the only other study of an estuarine bivalve in New Zealand (*Austrovenus stutchburyi,* Ross *et al.* 2011) reported similar genetic population structure to *P. australis*. Although *A. stutchburyi* populations were divided into seven genetically differentiated groups, evidence for IBD and some connectivity among groups was reported. Similar genetic patterns have been observed in estuarine bivalves from outside the New Zealand region (e.g., Tarnowska *et al.* 2010; Xiao *et al.* 2010).

4.4.3 Patterns of migration

Estimates of migration rates and geographic patterns of connectivity support the patterns of genetic population structure seen in *P. australis*. Migration rates were found to be highly variable in the short-term, but all populations were well connected in the long-term. The variability in short-term migration rates is likely to be contributing to the small scale patterns of genetic differentiation seen, and implies that these patterns of genetic differentiation might not be temporally stable. In contrast, the regular exchange of migrants observed when gene flow was averaged over long time periods is probably maintaining large scale IBD genetic structure and the admixture seen in cluster analyses.

Short-term gene flow estimates highlighted how migration can be an important source of recruits for some locations but self-recruitment can be high for others. GENECLASS2 indicated that first generation migration is low, a result that is supported by increasing evidence for limited larval dispersal and high self-recruitment in coastal marine invertebrates (e.g., Broekhuizen *et al.* 2011; Stephens *et al.* 2006). BAYESASS estimated

migration averaged over the number of cohorts that are present in the population (approximately five to ten years) and showed how over that period, some locations received large proportions of migrants but for other locations migration was low and self-recruitment more important. The detection of high gene flow rates among some locations is at odds with the significant genetic differentiation observed among many locations as measured by F_{ST}, ϕ'_{ST} and AMOVA. Similar to *P. subtriangulata*, recent migration among *P. australis* locations is patchy; limited source locations were identified, migration was mostly unidirectional and was not detected among some locations. For the reasons discussed in chapter 3, temporal variation in the exchange of larvae could allow significant genetic differentiation to develop over short time scales if effective population size is low and rates of genetic drift and mutation are high (Hedgecock 1994; Hedgecock *et al.* 2007). These processes would quickly break down the types of gametic equilibrium that short-term Bayesian analyses such as GENECLASS2 and BAYESASS are designed to detect. Again, this result suggests temporal variability of recruitment processes in estuarine species.

Both methods used to estimate recent migration rates were consistent with an asymmetric pattern of gene flow among populations. The main source of migrants is from the northern group. These source locations are well connected to other locations within the northern group, as well as to populations in the two southern groups. In contrast, very few migrants disperse in the opposite direction from locations in the southern groups to northern locations. This pattern is primarily driven by high levels of connectivity along the west coast of the North and South Islands (i.e., from Huia and Raglan to Karamea, Okuru, Doubtful Sound and Bluff). In comparison, a low proportion of individuals in group I were sourced from the southern groups II and III, a pattern that persists for historic gene flow estimates. However, the recent high levels of connectivity that were detected along the west coast seem to be less important in the long-term. Importantly, despite the variability in recent estimates of connectivity, when averaged over long time periods all locations were found to be genetically connected. This finding provides evidence that despite the presence of barriers to dispersal, rare gene flow events can overcome these barriers to connect populations and/or stepping-stone migration could maintain genetic similarities among distant populations over long periods of time.

Long-term effective population size for *P. australis* was also lower than what has been reported for other bivalve species (see references in chapter 3) but was higher than that reported for *P. subtriangulata*. The ratio of N_e to census population size is also likely to be very low in *P. australis* (Frankham 1995; Palstra & Ruzzante 2008) meaning that actual population size could be five to six orders of magnitude larger than the effective population sizes reported here. Similar to the trend reported by Palstra & Ruzzante (2008), migration was higher into *P. australis* populations that had lower effective population sizes. This is expected to be a function of density (Pulliam 1988) and the effects of gene flow will be more pronounced over genetic drift for populations that are small in size. This could explain why differentiation is lower among South Island populations; low N_e and net flow of migrants from north to south would homogenise allele frequencies and counteract the effects of genetic drift occurring in each population. In general, low N_e means that changes to the genetic population structure of *P. australis* could occur quickly as the migration-drift equilibrium changes in response to variable oceanographic conditions and intermittent influxes of external recruitment, a finding that is consistent with the small scale patterns of genetic differentiation that were observed.

Few studies have attempted to estimate migration among populations of New Zealand marine invertebrates using the Bayesian methods described here, so it is difficult to know if the patterns observed for *P. australis* are typical. Wei *et al.* (2013a) estimated recent migration for the greenshell mussel *Perna canaliculus*, which also shows a pattern of north-south differentiation. Although few F0 migration events were observed between the two groups, higher migration from south to north was observed. However, it is difficult to make comparisons due to the small number of F0 migrants detected and because their study included more populations from the South Island than North Island. Other studies of marine invertebrates that have used Bayesian methods to estimate migration rates on both recent and historic time scales have often reported asymmetric migration rates and high spatial variation in migration, within a range similar to that seen in *P. australis* (e.g., Crandall *et al.* 2012; Einfeldt & Addison 2013; Richards *et al.* 2007; Weetman *et al.* 2006).

4.4.4 Possible explanations for observed genetic population structure and migration

In general, the genetic population structure and patterns of migration observed for *P. australis* are in agreement with what might be expected for a free-spawning marine

bivalve with a three week pelagic larval duration, occupying a discontinuous habitat type. This period should be sufficient for short scale dispersal among populations over approximately 100-200 km each generation (Shanks 2009; Shanks *et al.* 2003) and over longer distances via indirect migration. However, small scale population differentiation as well as long distance migration events were observed, suggesting that IBD for this species may not be a simple, constant relationship between genetic and geographic distance. Instead, other environmental factors are likely to be influencing dispersal by creating barriers in some locations (as evidenced by the three significantly differentiated groups) and by promoting regular long-distance dispersal in other locations (e.g., along the west coast).

Where enhanced genetic population differentiation and IBD have been reported among other New Zealand estuarine species (e.g., Hickey *et al.* 2009; Perrin *et al.* 2004; Stevens & Hogg 2004), it has often been attributed to restricted access to coastal currents, dispersal ability and habitat availability (Watts & Johnson 2004). The variable and stochastic interaction of these processes means that there is great potential for small scale genetic differentiation of populations to arise and change over time, similar to the patterns observed in *P. australis*. For gene flow to occur, larvae that are spawned in estuaries must first overcome tidal flows and complex freshwater/estuarine circulation patterns to reach open water and enter coastal currents, then must re-enter an estuary to find suitable settlement habitat. The discrete nature of the estuarine environment and distance between suitable estuarine habitat is expected to limit the dispersal potential of larvae for estuarine organisms (Bilton 2002). Therefore the potential dispersal distance reported by Shanks (2009) for coastal marine organisms with similar larval duration might not equate to the same dispersal distance for an estuarine species like *P. australis*. Simulations of propagule dispersal from New Zealand harbours by Chiswell and Rickard (2011) show how the distance between harbour/estuary habitat and oceanographic processes interact to modify patterns of larval dispersal. Their study found that dispersal rate was highly dependent on both the velocity and direction of currents, meaning that the distance larvae could disperse within a set time period was highly variable for different parts of the country. Dispersal between some harbours (e.g., Bay of Islands and Whangarei) was achieved within a matter of days. But the dispersal time between most of the harbours that were modelled is much longer than *P. australis* larval duration (Chiswell & Rickard 2011). This study helps to explain why

significant genetic differentiation was observed among many of the *P. australis* populations that were sampled in the present study.

The combined effect of habitat availability and oceanographic processes on genetic population structure is further illustrated by the contrasting patterns of population subdivision observed among North Island versus South Island locations. Suitable *P. australis* habitat is more available and continuous in the northern regions of New Zealand (Hume *et al.* 2007) so high connectivity and low genetic differentiation could be expected among group I populations. However within-group genetic differentiation is actually highest among group I populations; in this case differentiation is more likely to be driven by small scale differences in local physical/oceanographic features in the north of New Zealand, rather than by absence of habitat. Similarly, the west of the North and South Islands contain large sections of coastline that are unsuitable habitat for *P. australis* yet short-term migration rates show that west coast populations can have high levels of connectivity. Lower connectivity for west coast populations in the long-term and the genetic demarcation detected on the South Island's west coast further demonstrate how distance between suitable habitat can be overcome on occasion and suggest temporal variation in rates of connectivity along this coast. The differences observed between recent and historic migration rates also suggest that there is an element of temporal variability to the genetic population structure of this species. Where small scale patterns of genetic variation have been observed in other marine species it has been suggested that this structure may change over time (Hedgecock & Pudovkin 2011).

Current direction and velocity are likely to explain much of the migration patterns and rates that were observed in *P. australis* (see Figure 1.1). On a large scale, the predominant direction of current flow around New Zealand is west to east. On the west coast of the North Island the southward flowing West Auckland Current (WAUC) would transport larvae from northern populations to locations on the west coast of the South Island and to the east coast via the D'Urville Current (DC) through Cook Strait, explaining the asymmetrical migration rates from north to south. At a regional scale, oceanographic processes are likely to have led to the formation of three areas of significant genetic discontinuity. The explanations that have been proposed for the genetic discontinuities observed in other New Zealand coastal marine organisms include coastal current circulation

(Sponer & Roy 2002; Star *et al.* 2003), upwelling (Ayers & Waters 2005; Waters & Roy 2004) and historical events such as glaciation (Ross *et al.* 2011; Stevens & Hogg 2004).

The north-south division observed in many New Zealand marine species coincides with the mixing and upwelling of the southward flowing East Cape Current (ECC) and the northward flowing Southland Current (SC; Laing & Chiswell 2003). Additionally, the Westland Current (WC) and WAUC mix on the west coast of New Zealand and are channelled into the Cook Strait, creating fast flowing currents and turbulent mixing through the Strait (Stanton *et al.* 2001). This speed of this current could act as a barrier to dispersal across the Strait and might explain why differentiation was observed between Petone and Hakahaka Bay despite the narrowness of this body of water. The combined flow of Cook Strait, ECC and SC interact to create complex eddies on the east coast of New Zealand, and forces current flow offshore across the Chatham Rise (Laing & Chiswell 2003). Mixing of larvae between northern and southern current systems is potentially restricted and larvae will be transported offshore to unsuitable open ocean habitat. This area of current mixing is also associated with a known division between biogeographic provinces (Shears *et al.* 2008) where the warmer northern water mass meets the cooler southern water mass. The difference in water temperature regimes may also form a barrier, preventing locally adapted populations from becoming established outside their biogeographic province. This could provide a mechanism for the selective pressure that is potentially being placed on the *Paus_9* outlier locus.

Temporal variation in migration along the west coast of New Zealand is supported by the variation that has been observed in current patterns for this region. Both the WAUC and the northerly flowing WC on the west coast of the South Island are weak, variable and influenced by wind driven events (Chiswell & Rickard 2011; Sutton & Bowen 2011). Furthermore, the weakness of these currents and their susceptibility to wind driven events could explain why some long-term migration rates estimated for west coast locations are contrary to what might be expected given the predominant current patterns (e.g., high levels of long-term migration from Karamea to North Island locations). Southwards of Fiordland (in a similar location to where a genetic demarcation was observed in *P. australis*) the predominant current direction changes to a southerly flow, joining the SC which flows along the bottom of the South Island then northwards up the east coast (Laing & Chiswell

2003). This current is one of the least variable and most predictable of the coastal currents (Chiswell & Rickard 2011) so it is not surprising that long-term migration estimates identified Bluff as one of the largest sources of migrants, (particularly to other east coast locations) and this higher level of connectivity could help explain the lower levels of genetic differentiation among South Island locations.

The east coast of the North Island is characterised by a more complex situation. The East Auckland Current (EAUC) and ECC create a series of complex coastal eddies, which act to retain larvae and limit dispersal (Chiswell & Roemmich 1998; Stanton *et al.* 1997). Consequently, long-term migration rates were much lower among most North Island locations and this restricted connectivity could explain the higher levels of differentiation observed among North Island locations in spite of greater habitat availability. Despite what we know about the predominant flow of New Zealand's coastal currents, they are all modified to an extent by tides, winds, upwelling and changes in velocity (Laing & Chiswell 2003). This variability will in turn influence rates and direction of larval dispersal, resulting in estimates of both short and long-term migration rates that are counter to what might be expected based on typical current patterns.

Long-term genetic signals from range restrictions and fragmentation of populations associated with the last glacial maximum (LGM) are thought to still influence genetic structure of some coastal marine organisms, particularly in the South Island (Wallis & Trewick 2009). Barriers that formed during the fragmentation of South Island *P. australis* populations during the LGM would have promoted population differentiation. The genetic disjunction between Doubtful Sound and Okuru could be the result of this historic signal. Glaciation has been invoked as an explanation for decreased diversity in South Island populations of triplefin fish and as an explanation for a genetic demarcation on the west coast of the South Island for *Grahamina nigripenne* (Hickey *et al.* 2009). This same genetic demarcation has been attributed to glaciation for New Zealand terrestrial invertebrates (Trewick & Wallis 2001). Glaciation has also been invoked as an explanation for the genetic structure observed in Northern Hemisphere and Pacific bivalve populations (Arnaud-Haond *et al.* 2003; Ilves *et al.* 2010; Ni *et al.* 2012). Glaciation could explain the lower levels of differentiation seen among South Island *P. australis* populations as the number of source populations for post-glacial range expansion would have been restricted. As populations re-

colonised suitable habitat they would have maintained a similar genetic signal to their source population, potentially resulting in genetically similar individuals inhabiting a large range.

A further mechanism that could explain the genetic population structure and connectivity observed in *P. australis* is human-mediated dispersal. Patterns of population genetic structure are often explained in light of natural dispersive processes. Human-mediated dispersal is often overlooked, but can result in patterns of genetic structure that are independent of natural routes and barriers to dispersal (Carlton 2003). Species living in harbours and near ports are known to be affected by human-mediated dispersal (Marins *et al.* 2010; Preda *et al.* 2012; Torkkola *et al*. 2013), although the focus of these studies is often on invasion of exotic species. The impact of human-mediated dispersal on the genetic population structure of native species is often overlooked and difficult to tease apart from natural dispersal processes due to the confounding nature of past dispersal processes within the natural range of the species. However it has been suggested as an explanation for unusual or inconsistent patterns of genetic population structure seen in other New Zealand coastal marine invertebrates (Apte & Gardner 2001; Ross *et al*. 2011). Human mediated dispersal could explain the long distance estimates of connectivity that were observed in this study, especially where connectivity occurred over distances that exceeded the probable dispersal distances for larvae, as simulated by Chiswell and Rickard (2011). It could also explain instances of geographically unstructured genetic similarity among populations, e.g., Raglan and Napier are genetically undifferentiated despite being significantly different from geographically proximate populations, suggesting recent connectivity between these two locations but not to locations in between.

While habitat availability, current patterns and historical glaciation can be invoked to explain much of the population structure and patterns of migration observed in *P. australis*, it is likely that other environmental characteristics also play a role. As mentioned previously, estuaries are typically discrete environments characterised by variable physical and environmental features (Bilton 2002). It is possible that physical/environmental differences in the estuaries sampled for this study could be driving local adaptation, and hence genetic differentiation, in the populations that inhabit them; a possibility that will be further investigated in chapter 5.

4.4.5 Conclusions

Similar to other estuarine and harbour species, *P. australis* was characterised by isolation by distance genetic population structure, suggesting that habitat availability plays a role in shaping *P. australis* populations. However, significant genetic differences were observed among many populations, even those in close proximity, and three areas of genetic demarcation were observed where barriers to dispersal are likely to exist. These findings suggest that the relationship between genetic and geographic distance may not be a simple linear one for *P. australis* and that distance between populations by itself may not be a suitable proxy for levels of genetic differentiation. Instead, oceanographic processes act in unison with habitat availability to drive genetic population structure and connectivity in this species. Furthermore, the relative importance of habitat availability versus oceanographic processes in shaping the genetic population structure of *P. australis* appeared to vary across the range sampled for this study. Around the North Island of New Zealand, complex circulation patterns are likely to retain larvae and restrict dispersal, despite large areas of suitable habitat, whereas further South, high levels of connectivity are possible despite large gaps between areas of suitable habitat.

In addition to spatial variation in the processes driving *P. australis* genetic population structure there is also likely to be an element of temporal variation to this structure. Overall, genetic population structure and connectivity in *P. australis* is a good example of 'crinkled connectivity' (Ovenden 2013), where barriers are somewhat permeable and dispersal is restricted rather than absolute. Crinkled connectivity can result in the type of migration patterns observed in *P. australis*: spatial variation in short-term migration but more consistent levels of connectivity among all locations when averaged over time. Temporal variability in larval dispersal is a strong possibility for *P. australis* and may lead to a situation where genetic population differentiation will build up and break down over short time scales but would result in a signal of high connectivity and population admixture when averaged over long time scales. This study has provided further evidence that the unique habitat characteristics of estuaries can influence population genetic patterns in the species that inhabit these environments, and that this structure can be modified by oceanographic and other processes. The result is complex patterns of genetic population structure that change across spatial scales, ranging from hundreds to thousands of kilometres.

5 Seascape genetic analysis of *Paphies subtriangulata* and *Paphies australis*: correlation between environmental and genetic variation and evidence for local adaptation in New Zealand's coastal marine species

5.1 Introduction

Populations of marine species often show significant genetic differentiation but it can be challenging to identify clear geographic patterns in how this genetic variation is structured (Selkoe *et al.* 2010). Results presented in chapter 3 and 4 certainly suggest that this is the case for *Paphies subtriangulata* (tuatua) and *P. australis* (pipi). While both species displayed genetic structure that was somewhat consistent with larval duration, coastal current patterns and habitat availability, unusual patterns of small scale genetic differentiation and connectivity were observed between some populations that cannot be explained by these factors alone. These types of patterns are often dismissed as 'chaotic' and not ecologically meaningful (Hedgecock 1994; Selkoe *et al.* 2006, 2010; Siegel *et al.* 2008). In this situation the emerging field of 'seascape genetics' can be applied to investigate how spatially variable environmental features can influence patterns of genetic variation in marine species (Riginos & Liggins 2013).

Results presented for *P. subtriangulata* and *P. australis* in previous chapters highlight the problems associated with population genetic analyses in marine species. Processes in the marine environment are spatially and temporally heterogeneous (Kinlan & Gaines 2003; Levin 2006), which is generally unaccounted for in simple panmictic, island or stepping-stone models of genetic population structure. This variation, coupled with large population

sizes that respond weakly (if at all) to genetic drift and are slow to reach equilibrium conditions, means that statistical power to detect spatial genetic patterns can be low (Selkoe *et al.* 2008). By using a multi-disciplinary approach such as seascape genetics, traditional population genetics techniques can be complemented with ecologically meaningful information. The power to detect subtle patterns of spatial genetic variation is increased, allowing for better understanding of the processes that might be driving these patterns. Despite the relatively recent origin of the field, seascape genetics has already revealed fascinating insights to how and why marine species are spatially structured (Banks *et al.* 2007; Galindo *et al.* 2006; González-Wangüemert *et al.* 2009), particularly for commercially important species (Coscia *et al.* 2012; Jørgensen *et al.* 2005; Kenchington *et al.* 2006; Wei *et al.* 2013b). These studies have provided evidence for strong associations between genetic variation and environmental gradients, such as sea surface temperature (SST) and salinity. They have shown the value of seascape genetics methods for identifying locations where oceanographic features are limiting or promoting gene flow and how, coupled with the variable life history characteristics of the organism, this can lead to the fine-scale patchy genetic structure that characterises many marine species.

Seascape genetics methods are yet to be widely applied to New Zealand coastal marine species (but see Wei *et al.* (2013b) for one example). The scale and complexity of the New Zealand marine environment means that marine species will experience a large degree of temporal and spatial environmental variability (Laing & Chiswell 2003; Shears *et al.* 2008). These conditions are ideal for testing correlations between environmental and genetic variation to understand the processes that drive population structure in the New Zealand marine environment. Furthermore, comparative studies of multiple species, such as the present study, can help us understand where and why common patterns of genetic population variation exist, and conversely how variation in environmental features can create different patterns of genetic population structure. This is particularly important for commercially and recreationally valuable species, such as *P. subtriangulata* and *P. australis*, as the insights gained from these analyses may be directly applicable to the management of the fisheries of these two species, and also to other fishery species.

In this chapter several different but complementary seascape genetic analyses were used to investigate the relationship between genetic variation among populations and

variability in environmental and geospatial features, for both *P. subtriangulata* and *P. australis*. Population differentiation metrics, based on microsatellite markers, were derived from the genetic data described in chapters 3 and 4. The recommendations of Bird *et al.* (2011) and Meirmans & Hedrick (2010) were followed: both a fixation index (F_{ST}) and measure of genetic distance (ϕ'_{ST}) were used, as they represent different aspects of genetic population differentiation. In addition, raw allelic frequency data were also used to summarise further aspects of population genetic variation. Environmental variables were derived from readily available public data sets. Following Wei *et al.* (2013b), two statistical approaches were implemented: (1) a generalised linear model (GLZ) that uses multiple regression to analyse the relationship between a dependent variable and a number of predictor variables, and (2) a biological environmental stepwise (BEST) model that tests for associations among resemblance matrices of dependent and predictor variables. Agreement between different methods and genetic distance measures can provide confidence that variables identified as significant in resulting models are truly associated with population genetic variation. Such methodology can help overcome the problem of low statistical significance commonly associated with genetic analyses of marine species. However, it is important to keep in mind when interpreting these analyses that correlation among genetic and environmental variation does not necessarily imply causation; for this, demonstration of a mechanistic link is required (Faurby & Barber 2012), something that is beyond the scope of this project.

Specifically, this chapter aimed to further examine patterns of spatial genetic variability among *P. subtriangulata* and *P. australis* populations by:

1. Using the population genetic data from chapters 3 and 4 to test for associations between population genetic variability and geospatial/environmental variability, testing the null hypothesis that there is no correlation between genetic and geospatial/environmental variables.
2. Revisiting patterns of population genetic structure reported in chapters 3 and 4 to verify if the geospatial/environmental variables identified as significant are consistent with genetic patterns previously reported for the study species, and to establish further testable hypotheses on how these population genetic patterns might have formed.

3. Comparing the results of analyses for both species to look for common geospatial/environmental variables that could be responsible for driving genetic population structure of New Zealand coastal marine species, and to understand species-specific responses to environmental variation.

5.2 Methods

5.2.1 Sample collection and population genetic data

Seascape genetic analyses for *P. subtriangulata* were based on 517 samples from ten locations, genotyped with ten non-outlier microsatellite loci, as described in chapter 3. Seascape genetic analyses for *P. australis* were based on 674 samples from 13 locations, genotyped with 12 non-outlier microsatellite loci, as described in chapter 4. Allele frequencies and pairwise Weir and Cockerham (1984) F_{ST} values were calculated using GENEPOP 4.2 (Rousset 2008). Standardised pairwise ϕ'_{ST} values (Meirmans 2006) were calculated between locations using GENODIVE 2.0b23 (Meirmans & van Tienderen 2004). From these genetic data two dependent variables, representing different aspects of population genetic variation, were calculated: (1) a mean multilocus F_{ST} or ϕ'_{ST} value for each location, derived from nine pairwise F_{ST} or ϕ'_{ST} values for *P. subtriangulata* and 12 pairwise F_{ST} or ϕ'_{ST} values for *P. australis*; and (2) the frequency of allele variants observed at each location.

5.2.2 Spatial genetic analyses

For both species, site-specific environmental data were obtained from the New Zealand Marine Environment Classification (MEC) scheme (New Zealand Ministry for the Environment 2005) for seven variables (Table 5.1A). These variables have a spatial resolution of 1 km and are based on long-term averages. For *P. australis* additional site-specific data were obtained from the New Zealand Estuarine Environment Classification (EEC) system (Hume *et al.* 2007) as this information is of direct relevance to the habitat of this species. Variables in the EEC are derived from a variety of sources but primarily from the New Zealand Digital Topographic Database. Eleven variables were obtained for each location (Table 5.1B). In addition to environmental variables, three geospatial variables were obtained for both species for each site: (1) latitude, (2) longitude and (3) a distance index

Table 5.1: Environmental variables, definitions, abbreviations and units from A: the Marine Environment Classification (MEC) scheme (New Zealand Ministry for the Environment 2005) and B: the Estuarine Environment Classification (EEC) scheme (Hume *et al.* 2007) that were used in generalised linear models (GLZ) implemented in Statistica v.10 (StatSoft Inc.) and biological environmental stepwise (BEST) models implemented in Primer v.6.

Variable	Definition	Abbreviation	Units
A: Marine Environment Classification			
Annual mean solar radiation	Mean extra atmospheric solar radiation modified by mean annual cloud cover	rad_mean	W m^{-2}
Winter sea surface temperature	Mean of daily data from early September when SST is typically lowest	SSTwint	°C
Annual amplitude of sea surface temperature	Smoothed annual amplitude of SST	SSTanamp	°C
Spatial gradient annual mean sea surface temperature	Smoothed magnitude of the spatial gradient of annual mean SST	SSTgrad	°C km^{-1}
Mean orbital velocity	Orbital velocity at the bed for the mean significant wave height calculated from a 20-year wave hindcast	orb_v_mean	m s^{-1}
Tidal current	Depth averaged maximum tidal current	tidal	m s^{-1}
Seabed rate of change of slope	The rate of change of slope for each cell	bed_slope	0.01 m^{-1}
B: Estuarine Environment Classification			
Spring tidal range	The tidal range for a mean spring tide	spring_TR	m
Spring tidal prism	The volume of water entering an estuary on the flood or incoming spring tide	spring_TP	m
Width of mouth	Width of the seaward boundary or mouth of an estuary where the estuary water body meets the ocean	mouth_width	m
Catchment area	The area of the land catchment that drains into the estuary.	catch_area	km^2
Shoreline length	The length of the shoreline of the estuary	shore_length	m
Intertidal area	The area of estuary exposed at spring low tide	intertidal	% of high tide area
Mean depth	The total estuary volume at spring high water divided by the estuary area at high water spring tide	mean_depth	m
Estuary area at low tide	The total water area at spring low tide	area_LT	m^2
Estuary area at high tide	The total water area at spring high tide	area_HT	m^2
Total volume at spring high tide	The total volume of water at spring high tide	tot_vol_SHT	m^3
Mean annual river discharge	Mean annual river flow into the estuary	river_discharge	cumecs

(geo_dist, sum of all shortest coastal distances between locations in km). All three geospatial variables were drawn from Google Maps. Variables were tested for independence using a Pearson correlation test and a principal component analysis (PCA) as implemented in the statistical package STATISTICA v.10 (StatSoft Inc.). A subset of independent environmental variables were determined for each species and the three geospatial variables were assumed to be independent.

A generalised linear model (GLZ) was run using the GLZ routine in STATISTICA v.10, to test the effect of the independent variables on F_{ST} or ϕ'_{ST}. The following models were run using a mean calculated from all pairwise estimates of both F_{ST} and ϕ'_{ST} for each location as the dependent variable:

1. All *P. subtriangulata* locations using MEC and geospatial independent variables;
2. Mainland *P. subtriangulata* locations (i.e., excluding the Chatham Island population) using MEC and geospatial independent variables;
3. All *P. australis* locations using MEC and geospatial independent variables;
4. All *P. australis* locations using EEC and geospatial independent variables.
5. All *P. australis* locations used MEC, EEC and geospatial independent variables.

These analyses were then repeated using only geospatial variables or only environmental variables. This was to determine the relative importance of geospatial versus environmental variables in explaining genetic variation. For *P. australis*, MEC and EEC variables were first treated separately and then together. This was to make results comparable between species because the MEC was the only set of environmental variables used for *P. subtriangulata* analyses. Models were built using the 'best subsets' option and all models were ranked by the Akaike information criterion (AIC). The best fit model was the one with the lowest AIC score and was considered significant if *p*-values were less than 0.05. Results of tests for the contribution of all effects to the model were also presented.

The BEST routine was implemented in PRIMER v.6 (Clarke & Gorley 2006) to test for associations between site-specific allele frequencies (dependent variable) and environmental/geospatial variation (independent variables). The same five models employed in GLZ analyses were also run for BEST analyses and the same sets of independent variables were used. A Bray-Curtis resemblance matrix was calculated for allele frequencies

and a matrix of Euclidean distances was calculated for the environmental/geospatial variables. The BIOENV routine was used to test for correlation between the two matrices using the Spearman correlation coefficient (R_S). All possible models were tested and ranked by their correlation coefficient. Models with higher correlation coefficients contained environmental and geospatial variables that better explained the variation in allele frequencies among populations. Models were considered significant if *p*-values were less than 0.05 after 1,000 permutations.

5.3 Results

5.3.1 Correlation of variables

A total of 131 different alleles were observed across ten *P. subtriangulata* locations and 237 different alleles were observed across 13 *P. australis* locations (Appendix 1). The number of alleles at each locus ranged from 5-28 for *P. subtriangulata* and 11-48 for *P. australis*. Mean multilocus F_{ST} ranged from 0.021-0.121 for *P. subtriangulata* and 0.005-0.014 for *P. australis* (Appendix 4). Mean multilocus ϕ'_{ST} ranged from 0.029-0.230 in *P. subtriangulata* and 0.006-0.031 in *P. australis* (Appendix 4). The location specific values for the geospatial and environment variables that were used in both GLZ and BEST analyses are listed in appendix 4.

Pearson correlation tests and PCA for all ten *P. subtriangulata* locations showed that five MEC environmental variables (orb_v_mean, rad_mean, SSTgrad, tidal, bed_slope) were independent of all other variables. SSTwint was correlated with latitude and SSTanamp was correlated with rad_mean, therefore SSTwint and SSTanamp were not used in any analyses of all *P. subtriangulata* locations (Table 5.2A; Figure 5.1A). When the Chatham Island population was excluded, four of the environmental variables were independent (SSTanamp, SSTgrad, tidal, bed_slope). SSTwint and rad_mean were both correlated with latitude and geographic distance, and orb_v_mean was correlated with SSTgrad (Table 5.2B; Figure 5.1B).

Pearson correlation tests for *P. australis* locations also showed that five MEC environmental variables (orb_v_mean, SSTanamp, SSTgrad, tidal, bed_slope) were independent. SSTwint and rad_mean were correlated with latitude and longitude so were excluded from further analysis (Table 5.3A; Figure 5.2A). A Pearson correlation test and PCA

of the EEC environmental variables revealed that many of these variables were correlated, but eight were independent (spring_TR, mouth_width, shore_length, intertidal, mean_depth, area_LT, area_HT, tot_vol_SHT; Table 5.3B; Figure 5.2B). Catch_area and spring_TP were correlated with shore_length, and river_discharge was correlated with tot_vol_SHT. None of the EEC variables were correlated with any of the three geospatial variables. When MEC and EEC variables were combined for *P. australis* nine environmental variables were independent (orb_v_mean, SSTgrad, SSTwint, tidal, bed_slope, mouth_width, shore_length, intertidal, tot_vol_SHT; Table 5.3C, Figure 5.2C). The final set of independent geospatial, MEC and EEC variables that were used for each analysis are listed in Table 5.2 for *P. subtriangulata* and Table 5.3 for *P. australis*.

Table 5.2: Correlations between Marine Environment Classification variables for *Paphies subtriangulata* A: for all locations; B: for mainland locations only (excluding Chatham Island). Independent variables are listed along with variables that are significantly correlated as determined by a Pearson's Correlation test implemented in STATISTICA v.10 ($\alpha = 0.05$). These represent the set of independent variables that were used in GLZ and BEST analyses.

Independent Variables	Correlated Variables	Pearson Co-efficient	*p*-value
A: All locations			
latitude	SSTwint	0.953	0.001
longitude	-		
geo_dist	-		
orb_v_mean	-		
rad_mean	SSTanamp	0.848	0.002
SSTgrad	-		
tidal	-		
bed_slope	-		
B: Mainland locations			
latitude	rad_mean	0.939	0.001
	SSTwint	0.928	0.001
longitude	-		
geo_dist	rad_mean	0.821	0.007
	SSTwint	0.721	0.028
SSTanamp	-		
SSTgrad	orb_v_mean	0.781	0.022
tidal	-		
bed_slope	-		

Table 5.3: Correlations between A: Marine Environment Classification variables, B: Estuarine Environment Classification and C: Marine and Estuarine Environment Classification variables combined for *Paphies australis*. Independent variables are listed along with variables that are significantly correlated as determined by a Pearson's Correlation test implemented in STATISTICA v.10 ($\alpha = 0.05$). These represent the set of independent variables that were used in GLZ and BEST analyses.

Independent Variables	Correlated Variables	Pearson Co-efficient	*p*-value
A: Marine Environment Classification			
latitude	rad_mean	0.981	0.001
	SSTwint	0.927	0.001
longitude	rad_mean	0.799	0.001
geo_dist	-		
orb_v_mean	-		
SSTanamp	-		
SSTgrad	-		
tidal	-		
bed_slope	-		
B: Estuarine Environment Classification			
latitude	-		
longitude	-		
geo_dist	-		
spring_TR	-		
mouth_width	-		
shore_length	catch_area	0.653	0.015
	spring_TP	0.898	0.001
intertidal	-		
mean_depth	-		
area_LT	-		
area_HT	-		
tot_vol_SHT	river_discharge	0.698	0.008
C: Marine and Estuarine Environment Classification			
latitude	rad_mean	0.981	0.001
longitude	SSTanamp	0.644	0.017
geo_dist	-		
orb_v_mean	-		
SSTgrad	-		
SSTwint	-		
tidal	-		
bed_slope	-		
mouth_width	-		
shore_length	spring_TP	0.898	0.001
	catch_area	0.653	0.015
	area_LT	0.806	0.001
	area_HT	0.941	0.001
intertidal	spring_TR	0.686	0.049
tot_vol_SHT	mean_depth	0.993	0.001
	river_discharge	0.698	0.008

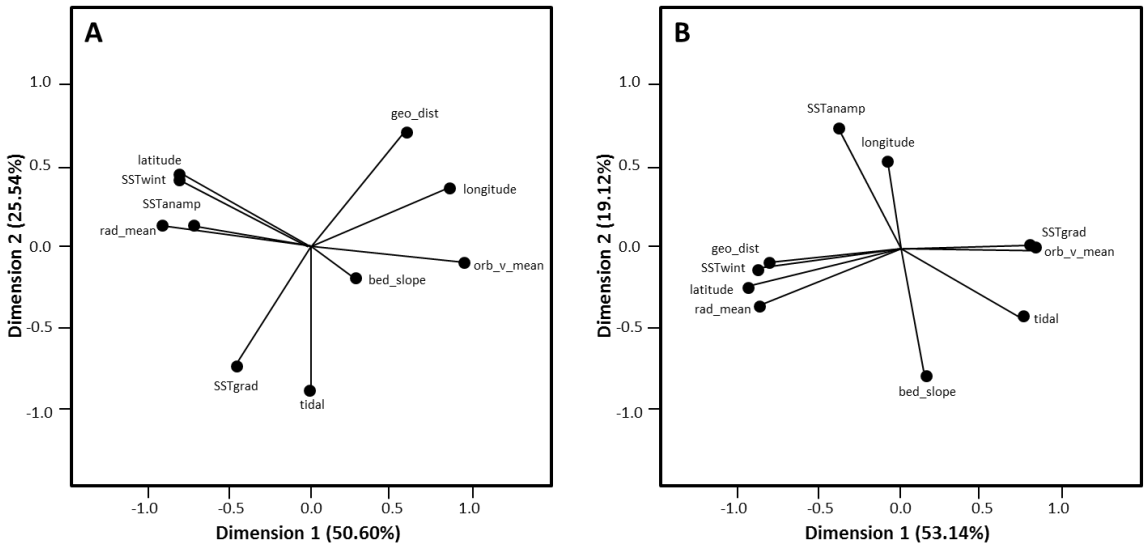


Figure 5.1: Principal component analysis for geospatial and Marine Environment Classification variables conducted in STATISTICA v.10. A: for all *Paphies subtriangulata* populations; B: for mainland *P. subtriangulata* populations, excluding the Chatham Islands.

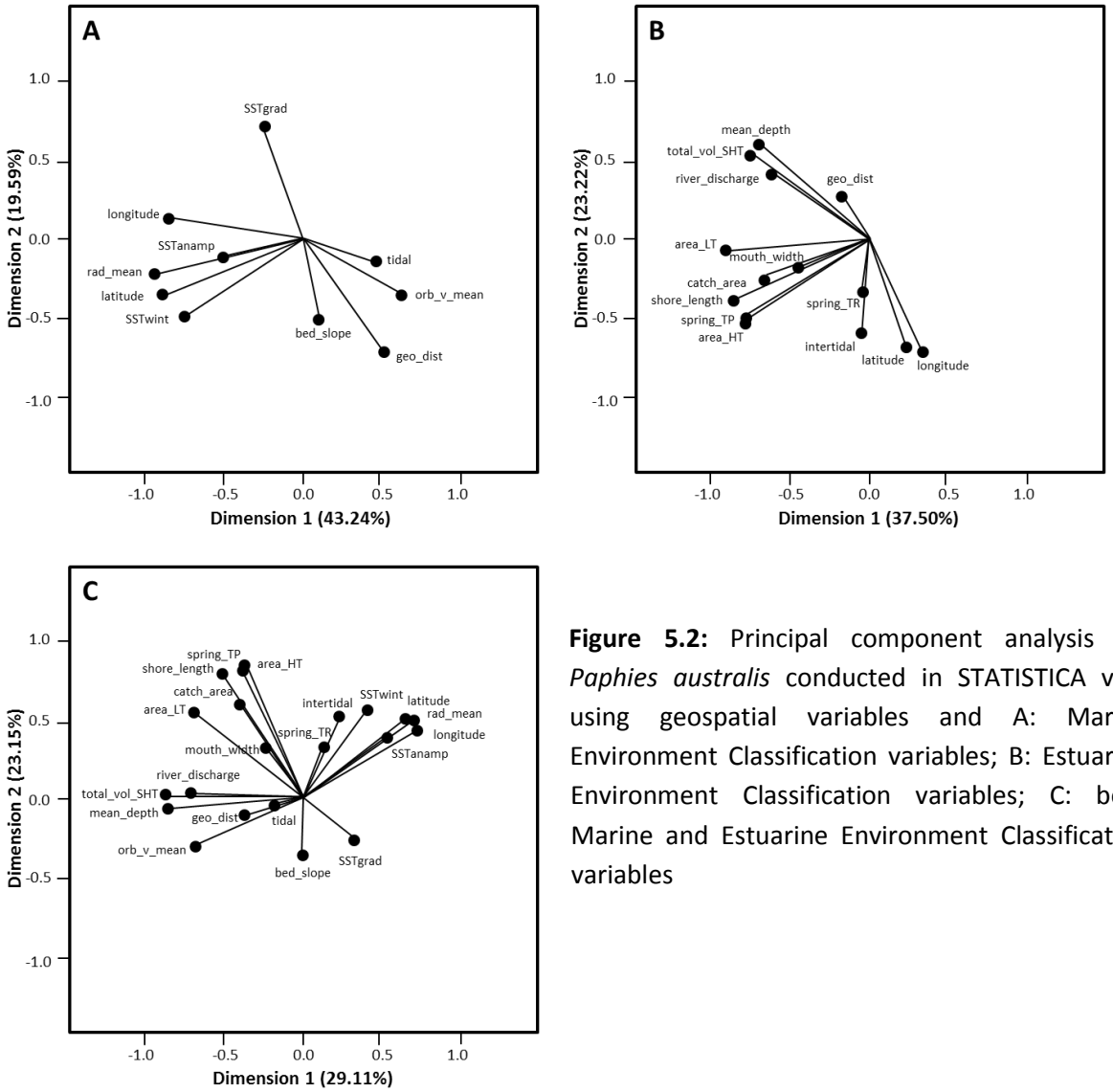


Figure 5.2: Principal component analysis for *Paphies australis* conducted in STATISTICA v.10 using geospatial variables and A: Marine Environment Classification variables; B: Estuarine Environment Classification variables; C: both Marine and Estuarine Environment Classification variables

5.3.2 Generalised linear modelling

Generalised linear models based on data from all *P. subtriangulata* locations showed that all possible models were statistically significant ($p < 0.001$), i.e., removing or adding variables to the model did not have a large overall effect on the model fit. This result is most likely attributed to the Chatham Island population, which was an outlier for many of the variables included in the models. The best fit model using F_{ST} included the variables geo_dist, orb_v_mean, SSTgrad and bed_slope ($p < 0.001$) and using ϕ'_{ST} included the variables geo_dist, rad_mean, tidal and bed_slope ($p < 0.001$; Table 5.4A). Both geo_dist and bed_slope were identified as explaining genetic distance when both F_{ST} and ϕ'_{ST} were included as the dependent variable. When the Chatham Island population was excluded from GLZ analyses *p*-values for models were much lower, although a number of models were still significant (31 significant models for F_{ST} and 24 for ϕ'_{ST}). The best fit model for F_{ST} included longitude, geo_dist, SSTanamp, SSTgrad, tidal and bed_slope ($p = 0.004$; Table 5.4B). For ϕ'_{ST} the best fit model included geo_dist, SSTanamp and bed_slope ($p = 0.010$; Table 5.4B).

When GLZ analyses were repeated using only geospatial or environmental variables the results showed that both groups of variables were important in explaining genetic variation among all *P. subtriangulata* locations. Best fit models for geospatial variables included latitude and geo_dist when both F_{ST} and ϕ'_{ST} were used ($p < 0.001$; Table 5.5A). SSTgrad and orb_v_mean were identified as important environmental variables when both F_{ST} and ϕ'_{ST} were used (Table 5.5B). Geospatial variables explained much less of the genetic variation for mainland locations (i.e., Chatham Island population excluded); best fit models for F_{ST} and ϕ'_{ST} included geo_dist only and were not significant (F_{ST}: $p = 0.086$; ϕ'_{ST}: $p = 0.131$; Table 5.5C). Environmental variables were able to better explain genetic variation among mainland populations; the best fit model for F_{ST} was SSTanamp, tidal and bed_slope ($p = 0.042$) and for ϕ'_{ST} the best fit model included bed_slope only ($p = 0.045$; Table 5.5D).

A GLZ for *P. australis* using geospatial and MEC variables identified 20 significant models using F_{ST} and only three significant models using ϕ'_{ST} (Table 5.6A). The best fit model for F_{ST} included latitude, longitude, geo_dist, SSTanamp, SSTgrad and bed_slope ($p = 0.010$). Using ϕ'_{ST}, the best fit model included geo_dist, orb_v_mean, SSTgrad, tidal and bed_slope ($p = 0.034$). There was a difference in the variables that best explained variation in F_{ST} versus

those that best explained variation in ϕ'_{ST}; the only variables that were common to both models were geo_dist, SSTgrad and bed_slope. A GLZ using geospatial and EEC variables resulted in a large number of highly significant models ($p < 0.001$; Table 5.6B) and while one best fit model was identified, many of the *p*-values for the individual variables were also highly significant, meaning that adding or removing different variables had little effect on the fit of subsequent models. The best fit model for F_{ST} included all variables except area_LT and the best fit model for ϕ'_{ST} included all variables except intertidal. Similar results were seen when analyses were run using a combination of geospatial, MEC and EEC variables. All possible models were highly significant ($p < 0.001$) and all variables were included in the best fit models for both F_{ST} and ϕ'_{ST} (Table 5.6C).

A GLZ using just geospatial variables showed that these variables contributed little to explaining the genetic variation seen among *P. australis* locations (Table 5.7A). Latitude was the geospatial variable that best explained F_{ST} but this model was not significant ($p = 0.089$). For ϕ'_{ST} latitude and geo_dist was the best fit model but this model was also not significant ($p = 0.169$). A GLZ using MEC variables showed that these variables were also unable to explain much genetic variation in *P. australis* (Table 5.7B). For F_{ST} the best fit model included only orb_v_mean and was not significant ($p = 0.238$). Similarly, the best fit model for ϕ'_{ST} was not significant and included only tidal ($p = 0.408$). In contrast, models using EEC variables were significant, suggesting that these variables were able to better explain the genetic variation seen in *P. australis* (Table 5.7C). Both F_{ST} and ϕ'_{ST} produced similar best fit models (F_{ST}: spring_TR, mouth_width, intertidal, mean_depth, area_HT, tot_vol_SHT, $p = 0.003$; ϕ'_{ST}: intertidal, mean_depth, area_HT, tot_vol_SHT, $p = 0.002$). When MEC and EEC variables were combined the models were highly significant. A five factor model best explained variation in F_{ST} (SSTgrad, SSTwint, tidal, intertidal, tot_vol_SHT; $p < 0.001$) and a seven factor model best explained variation in ϕ'_{ST} (SSTgrad, SSTwint, tidal, bed_slope, shore_length, intertidal, tot_vol_SHT; $p < 0.001$; Table 5.7D).

Table 5.4: Results of generalised linear model analyses for *Paphies subtriangulata* implemented in STATISTICA v.10. Analyses were run both using both geospatial and MEC variables including (all locations; A) and excluding (mainland locations; B) the Chatham Island population. The best fit model is shown, as determined by lowest AIC score, for the dependent and independent variables listed. The model was considered significant if $p < 0.05$.

Test	Dependent Variable	Independent Variables	*p*-value All effects	Best Fit Model	*p*-value Model
A: All locations	F_{ST}	latitude	0.133	geo_dist	<0.001
		longitude	0.244	orb_v_mean	
		geo_dist	0.053	SSTgrad	
		orb_v_mean	<0.001	bed_slope	
		rad_mean	0.708		
		SSTgrad	0.268		
		tidal	0.278		
		bed_slope	<0.001		
	ϕ'_{ST}	latitude	0.373	geo_dist	<0.001
		longitude	0.269	rad_mean	
		geo_dist	0.164	tidal	
		orb_v_mean	0.130	bed_slope	
		rad_mean	0.604		
		SSTgrad	0.420		
		tidal	0.134		
		bed_slope	0.006		
B: Mainland locations	F_{ST}	latitude	0.337	longitude	0.004
		longitude	0.003	geo_dist	
		geo_dist	<0.001	SSTanamp	
		SSTanamp	0.001	SSTgrad	
		SSTgrad	0.138	tidal	
		tidal	0.028	bed_slope	
		bed_slope	<0.001		
	ϕ'_{ST}	latitude	0.659	geo_dist	0.010
		longitude	0.879	SSTanamp	
		geo_dist	0.038	bed_slope	
		SSTanamp	0.189		
		SSTgrad	0.883		
		tidal	0.912		
		bed_slope	0.001		

Table 5.5: Results of generalised linear model analyses for *Paphies subtriangulata* implemented in STATISTICA v.10 where separate models were run for geospatial variables including all locations (A), MEC variables including all locations (B), geospatial variables for mainland locations only (C), and MEC variables for mainland locations only (D). The best fit model is shown, as determined by lowest AIC score, for the dependent and independent variables listed. The model was considered significant if $p < 0.05$.

Test	Dependent Variable	Independent Variables	*p*-value All Effects	Best Fit Model	*p*-value Model
A: All locations geospatial	F_{ST}	latitude	0.014	latitude	<0.001
		longitude	0.977	geo_dist	
		geo_dist	<0.001		
	ϕ'_{ST}	latitude	0.096	latitude	<0.001
		longitude	0.362	geo_dist	
		geo_dist	0.007		
B: All locations Marine Environment Classification	F_{ST}	orb_v_mean	<0.001	orb_v_mean	<0.001
		rad_mean	0.285	SSTgrad	
		SSTgrad	<0.001	tidal	
		tidal	0.074		
		bed_slope	0.288		
	ϕ'_{ST}	orb_v_mean	0.019	orb_v_mean	<0.001
		rad_mean	0.910	SSTgrad	
		SSTgrad	<0.001		
		tidal	0.278		
		bed_slope	0.430		
C: Mainland locations geospatial	F_{ST}	latitude	0.400	geo_dist	0.086
		longitude	0.746		
		geo_dist	0.076		
	ϕ'_{ST}	latitude	0.800	geo_dist	0.131
		longitude	0.459		
		geo_dist	0.302		
D: Mainland locations Marine Environment Classification	F_{ST}	SSTanamp	0.029	SSTanamp	0.042
		SSTgrad	0.727	tidal	
		tidal	0.030	bed_slope	
		bed_slope	0.007		
	ϕ'_{ST}	SSTanamp	0.050	bed_slope	0.045
		SSTgrad	0.865		
		tidal	0.089		
		bed_slope	0.005		

Table 5.6: Results of generalised linear model analyses for *Paphies australis* implemented in STATISTICA v.10. Analyses were run using Marine Environment Classification variables (A), Estuarine Environment Classification variables (B) and both Marine and Estuarine Environment Classification variables combined (C). The best fit model is shown, as determined by lowest AIC score, for the dependent and independent variables listed. The model was considered significant if $p < 0.05$.

Test	Dependent Variable	Independent Variables	*p*-value All Effects	Best Fit Model	*p*-value Model
A: Marine Environment Classification	F_{ST}	latitude	0.042	latitude	0.010
		longitude	0.145	longitude	
		geo_dist	0.025	geo_dist	
		orb_v_mean	0.381	SSTanamp	
		SSTanamp	0.736	SSTgrad	
		SSTgrad	0.001	bed_slope	
		tidal	0.235		
		bed_slope	0.001		
	ϕ'_{ST}	latitude	0.232	geo_dist	0.034
		longitude	0.327	orb_v_mean	
		geo_dist	0.017	SSTgrad	
		orb_v_mean	0.368	tidal	
		SSTanamp	0.958	bed_slope	
		SSTgrad	0.015		
		tidal	0.130		
		bed_slope	0.003		
B: Estuarine Environment Classification	F_{ST}	latitude	<0.001	latitude	<0.001
		longitude	<0.001	longitude	
		geo_dist	<0.001	geo_dist	
		spring_TR	<0.001	spring_TR	
		mouth_width	<0.001	mouth_width	
		shore_length	<0.001	shore_length	
		intertidal	<0.001	intertidal	
		mean_depth	<0.001	mean_depth	
		area_LT	<0.001	area_HT	
		area_HT	<0.001	tot_vol_SHT	
		tot_vol_SHT	<0.001		
	ϕ'_{ST}	latitude	0.021	latitude	<0.001
		longitude	<0.001	longitude	
		geo_dist	<0.001	geo_dist	
		spring_TR	<0.001	spring_TR	
		mouth_width	<0.001	mouth_width	
		shore_length	<0.001	shore_length	
		intertidal	0.199	mean_depth	
		mean_depth	<0.001	area_LT	
		area_LT	<0.001	area_HT	
		area_HT	<0.001	tot_vol_SHT	
		tot_vol_SHT	<0.001		

Table 5.6 continued.

Test	Dependent Variable	Independent Variables	*p*-value All Effects	Best Fit Model	*p*-value Model
C: Marine and Estuarine Environment Classification	F_{ST}	latitude	<0.001	latitude	<0.001
		longitude	<0.001	longitude	
		geo_dist	<0.001	geo_dist	
		orb_v_mean	<0.001	orb_v_mean	
		SSTgrad	<0.001	bed_slope	
		SSTwint	<0.001	SSTgrad	
		tidal	<0.001	tidal	
		bed_slope	<0.001	tot_vol_SHT	
		mouth_width	<0.001	mouth_width	
		shore_length	<0.001	shore_length	
		intertidal	<0.001	intertidal	
		tot_vol_SHT	<0.001	SSTwint	
	ϕ'_{ST}	latitude	<0.001	latitude	<0.001
		longitude	<0.001	longitude	
		geo_dist	<0.001	geo_dist	
		orb_v_mean	<0.001	orb_v_mean	
		SSTgrad	<0.001	bed_slope	
		SSTwint	<0.001	SSTgrad	
		tidal	<0.001	tidal	
		bed_slope	<0.001	tot_vol_SHT	
		mouth_width	<0.001	mouth_width	
		shore_length	<0.001	shore_length	
		intertidal	<0.001	intertidal	
		tot_vol_SHT	<0.001	SSTwint	

Table 5.7: Results of generalised linear model analyses for *Paphies australis* implemented in STATISTICA v.10 where separate models were run for geospatial variables only (A), MEC variables only (B), EEC variables only (C) and MEC and EEC variables combined (D). The best fit model is shown, as determined by lowest AIC score, for the dependent and independent variables listed. The model was considered significant if $p < 0.05$.

Test	Dependent Variable	Independent Variables	*p*-value All Effects	Best Fit Model	*p*-value Model
A: Geospatial Variables	F_{ST}	latitude	0.272	latitude	0.089
		longitude	0.540		
		geo_dist	0.138		
	ϕ'_{ST}	latitude	0.451	latitude	0.169
		longitude	0.555	geo_dist	
		geo_dist	0.062		
B: Marine Environment Classification	F_{ST}	orb_v_mean	0.216	orb_v_mean	0.238
		SSTanamp	0.068		
		SSTgrad	0.100		
		tidal	0.066		
		bed_slope	0.498		
	ϕ'_{ST}	orb_v_mean	0.066	tidal	0.408
		SSTanamp	0.008		
		SSTgrad	0.039		
		tidal	0.012		
		bed_slope	0.291		
C: Estuarine Environment Classification	F_{ST}	spring_TR	0.066	spring_TR	0.003
		mouth_width	0.211	mouth_width	
		shore_length	0.398	intertidal	
		intertidal	0.067	mean_depth	
		mean_depth	0.001	area_HT	
		area_LT	<0.001	tot_vol_SHT	
		area_HT	<0.001		
		tot_vol_SHT	<0.001		
	ϕ'_{ST}	spring_TR	0.189	intertidal	0.002
		mouth_width	0.093	mean_depth	
		shore_length	0.974	area_HT	
		intertidal	0.003	tot_vol_SHT	
		mean_depth	0.003		
		area_LT	<0.001		
		area_HT	<0.001		
		tot_vol_SHT	<0.001		

Table 5.7 continued.

Test	Dependent Variable	Independent Variables	*p*-value All Effects	Best Fit Model	*p*-value Model
D: Marine and Estuarine Environment Classification	F_{ST}	orb_v_mean	0.433	SSTgrad	<0.001
		SSTgrad	<0.001	SSTwint	
		SSTwint	<0.001	tidal	
		tidal	0.017	intertidal	
		bed_slope	0.349	tot_vol_SHT	
		mouth_width	0.766		
		shore_length	<0.001		
		intertidal	<0.001		
		tot_vol_SHT	<0.001		
	ϕ'_{ST}	orb_v_mean	0.225	SSTgrad	<0.001
		SSTgrad	<0.001	SSTwint	
		SSTwint	<0.001	tidal	
		tidal	0.001	bed_slope	
		bed_slope	0.013	shore_length	
		mouth_width	0.216	intertidal	
		shore_length	<0.001	tot_vol_SHT	
		intertidal	<0.001		
		tot_vol_SHT	<0.001		

5.3.3 BEST analyses

An analysis based on geospatial and MEC variables using all *P. subtriangulata* locations identified two best fit models that included combinations of rad_mean, bed_slope and SSTgrad (R_S: 0.643, p = 0.039; Table 5.8A). The remainder of the top 10 best models all had the same R_S value (0.634) and all included the variables geo_dist and orb_v_mean. Using just geospatial variables resulted in four best fit models that included all three geospatial variables, with geo_dist included in all four models (R_S: 0.613, p = 0.040; Table 5.8B). Using only MEC environmental variables two best fit models were identified that included combinations of rad_mean, bed_slope and SSTgrad (R_S: 0.643, p = 0.050; Table 5.8C). Other models also identified rad_mean and bed_slope as important variables in explaining allelic variation.

When the Chatham Island population was excluded from BEST analyses the ten best models using geospatial and MEC variables all had the same correlation coefficient and were not significant (R_S: 0.285, p = 0.590; Table 5.8D), suggesting that these variables poorly

explained allelic variation among mainland *P. subtriangulata*. The top ten models identified geospatial variables as important with geo_dist included in all ten models. An analysis using just geospatial variables identified three best fit models which also all included geo_dist (R_S: 0.285, p = 0.200; Table 5.8E). Using only MEC environmental variables resulted in two best fit models that included combinations of rad_mean, bed_slope and SSTgrad (R_S: 0.327, p = 0.470; Table 5.8F). Rad_mean and bed_slope also occurred in most of the ten best models.

All BEST analyses for *P. australis* had low correlation coefficients and models were not significant, suggesting that the variables used were not able to explain much of the allelic variation seen among *P. australis* locations. Using geospatial and MEC variables a best fit model of SSTgrad, tidal and bed_slope was obtained (R_S: 0.233, p = 0.460; Table 5.9A). Tidal current was included in eight of the ten best models and of the geospatial variables latitude was the only one included in any of the ten best models (8/10 models). Using geospatial and EEC variables the best fit model included latitude only (R_S: 0.159, p = 0.790; Table 5.9B). The remaining nine models all had the same correlation coefficient and all included area_HT in combination with one other geospatial or EEC variable (R_S: 0.151). Combining geospatial, MEC and EEC variables showed that MEC variables were able to explain more of the allelic variation seen in *P. australis* (Table 5.9C). The best fit model included SSTgrad, tidal and bed_slope (R_S: 0.233, p = 0.610) and tidal featured in all ten models.

Using only geospatial variables resulted in a best fit model that only included latitude (R_S: 0.159, p = 0.370; Table 5.9D). Using only MEC variables again resulted in a best fit model of SSTgrad, tidal and bed_slope, with tidal included in nine of the ten best models (R_S: 0.233, p = 0.420; Table 5.9E). Using only EEC variables resulted in the same correlation coefficient for all ten models, suggesting that EEC variables in particular were poor at explaining allelic variation. All models included the variable area_HT (R_S: 0.151, p = 0.450; Table 5.9F). When MEC and EEC variables were combined the best model was SSTgrad, tidal and bed_slope (R_S: 0.233, p = 0.660; Table 5.9G) but catch_area and river_discharge were also included in eight of the models.

To aid interpretation of all testing, the results of the GLZ and BEST analyses for both species are summarised in Table 5.10.

Table 5.8.

Test	No. Variables	R_S	Latitude	Longitude	geo _dist	orb_v _mean	rad _mean	SST anamp	SSTgrad	SSTwint	tidal	bed _slope	*p*-value
A: All locations	2	0.643					X					X	
all variables	3	0.643					X		X			X	
	2	0.634			X	X							
	3	0.634	X		X	X							
	3	0.634		X	X	X							
	3	0.634			X	X	X						
	3	0.634			X	X			X				
	3	0.634			X	X					X		
	3	0.634			X	X						X	
	4	0.634	X	X	X	X							
		Sum:	**2**	**2**	**8**	**8**	**3**		**2**		**1**	**3**	**0.039**
B: All locations	1	0.613			X								
geospatial variables	2	0.613	X		X								
	2	0.613		X	X								
	3	0.613	X	X	X								
	2	0.543	X	X									
	1	0.445		X									
	1	0.307	X										
													0.040
C: All locations	2	0.643					X					X	
environmental variables	3	0.643					X		X			X	
	1	0.626					X						
	3	0.610					X				X	X	
	4	0.610					X		X		X	X	
	3	0.605					X	X				X	
	4	0.605					X	X	X			X	
	4	0.597					X		X		X	X	
	5	0.597					X	X	X		X	X	
	2	0.591					X	X					
		Sum:				**0**	**10**	**4**	**5**	**0**	**4**	**8**	**0.050**

Table 5.8 continued.

Test	No. Variables	R_S	Latitude	Longitude	geo_dist	orb_v_mean	rad_mean	SST anamp	SSTgrad	SSTwint	tidal	bed_slope	*p*-value
D: Mainland locations all variables	2	0.285	X		X								
	2	0.285		X	X								
	3	0.285	X	X	X								
	3	0.285	X		X			X					
	3	0.285	X		X				X				
	3	0.285	X		X						X		
	3	0.285	X		X							X	
	3	0.285		X	X			X					
	3	0.285		X	X				X				
	3	0.285		X	X						X		
		Sum:	**6**	**5**	**10**			**2**	**2**	**0**	**2**	**1**	**0.590**
E: Mainland locations geospatial variables	2	0.285	X		X								
	2	0.285		X	X								
	3	0.285	X	X	X								
	1	0.284			X								
	2	0.122	X	X									
	1	0.062	X										
	1	-0.075		X									
													0.200
F: Mainland locations environmental variables	2	0.327					X					X	
	3	0.327					X		X			X	
	1	0.268					X						
	3	0.263					X				X	X	
	4	0.263					X		X		X	X	
	3	0.248					X	X				X	
	4	0.248					X	X	X			X	
	1	0.240										X	
	2	0.240							X			X	
	4	0.232					X	X			X	X	
		Sum:				**0**	**8**	**3**	**4**	**0**	**3**	**9**	**0.470**

Table 5.8 (previous pages): Results of BEST analyses for *Paphies subtriangulata* as implemented in Primer v.6. Models were run using all non-correlated variables for all locations (A), only geospatial variables for all locations (B), only MEC environmental variables for all locations (C), all non-correlated variables for mainland locations only (D), only geospatial variables for mainland locations (E), and only MEC environmental variables for mainland locations (F). The top 10 models, as determined by Spearman's rank co-efficient (R_S), are shown; X indicates the variables that were included in each model and the number of times each variable appears is summed at the bottom (note for geospatial variables all possible models are shown). The *p*-value for each model, as determined by 1,000 permutations is shown ($\alpha = 0.05$). Grey columns represent variables that were not included in the analysis.

Table 5.9 (over page): Results of BEST analyses for *Paphies australis* as implemented in Primer v.6. Models were run using geospatial and non-correlated MEC variables (A), geospatial and non-correlated EEC variables (B), geospatial and non-correlated MEC and EEC variables (C), only geospatial variables (D), only MEC environmental variables (E), only EEC environmental variables (F), and MEC and EEC environmental variables (G). Variables used in the models were latitude (1), longitude (2), geo_dist (3), orb_v_mean (4), rad_mean (5), SSTanamp (6), SSTgrad (7), SSTwint (8), tidal (9), bed_slope (10), spring_TR (11), spring_TP (12), mouth_width (13), catch_area (14), shore_length (15), intertidal (16), mean_depth (17), area_LT (18), area_HT (19), tot_vol_SHT (20) and river_discharge (21). The top 10 models, as determined by Spearman's rank co-efficient (R_S), are shown; X indicates the variables that were included in each model and the number of times each variable appears is summed at the bottom (note for geospatial variables all possible models are shown). The *p*-value for each model, as determined by 1,000 permutations is shown ($\alpha = 0.05$). Grey columns represent variables that were not included in the analysis.

Table 5.9.

Test	No. Variables	R_S	Variables																					*p*-value
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
A: MEC	3	0.233							X		X	X												
all variables	2	0.231							X		X													
	2	0.214									X	X												
	1	0.191									X													
	2	0.162	X								X													
	3	0.162	X						X		X													
	3	0.161	X								X	X												
	4	0.161	X						X		X	X												
	1	0.159	X																					
	2	0.159	X						X															
		Sum:	**6**	**0**	**0**	**0**		**0**	**5**		**8**	**4**												**0.460**
B: EEC	1	0.159	X																					
all variables	1	0.151																			X			
	2	0.151	X																		X			
	2	0.151		X																	X			
	2	0.151			X																X			
	2	0.151											X								X			
	2	0.151													X						X			
	2	0.151															X				X			
	2	0.151																X			X			
	2	0.151																	X		X			
		Sum:	**2**	**1**	**1**								**1**		**1**		**1**	**1**	**1**	**0**	**9**	**0**		**0.790**
C: MEC & EEC	3	0.233							X		X	X												
all variables	2	0.231							X		X													
	2	0.214									X	X												
	1	0.191									X													
	4	0.167							X	X	X	X												
	3	0.166								X	X	X												
	2	0.165								X	X													
	3	0.165							X	X	X													
	2	0.162	X								X													
	3	0.162	X						X		X													
		Sum:	**2**	**0**	**0**	**0**			**5**	**4**	**10**	**4**			**0**		**0**	**0**				**0**		**0.610**

Table 5.9 continued.

Test	No. Variables	R_S	Variables																					*p*-value
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
D: Geospatial variables	1	0.159	X																					
	2	0.093	X	X																				
	1	0.032		X																				
	1	-0.063			X																			
	2	-0.063	X		X																			
	2	-0.063		X	X																			
	3	-0.063	X	X	X																			
		Sum:																						**0.370**
E: MEC environmental variables	3	0.233							X		X	X												
	2	0.231							X		X													
	2	0.214									X	X												
	1	0.191									X													
	4	0.167							X	X	X	X												
	3	0.166								X	X	X												
	2	0.165								X	X													
	3	0.165							X	X	X													
	2	0.152							X	X														
	4	0.152					X		X		X	X												
		Sum:				**0**	**1**	**0**	**6**	**5**	**9**	**5**												**0.420**
F: EEC environmental variables	1	0.151																			X			
	2	0.151											X								X			
	2	0.151													X						X			
	2	0.151															X				X			
	2	0.151																X			X			
	2	0.151																	X		X			
	3	0.151											X		X						X			
	3	0.151											X				X				X			
	3	0.151											X					X			X			
	3	0.151											X						X		X			
		Sum:											**5**	**0**	**2**	**0**	**2**	**2**	**2**	**0**	**10**	**0**	**0**	**0.450**

Table 5.9 continued.

Test	No. Variables	R_S	Variables 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	*p*-value
G: MEC & EEC environmental variables	3	0.233				0	0	1	3	1	3	2	1	1	0	8	0	0	0	0	0	0	8	0.660
									X		X	X												
	2	0.231							X		X													
	2	0.230														X							X	
	3	0.230						X								X							X	
	3	0.230							X							X							X	
	3	0.230								X						X							X	
	3	0.230									X					X							X	
	3	0.230										X				X							X	
	3	0.230											X			X							X	
	3	0.230												X		X							X	
		Sum:				0	0	1	3	1	3	2	1	1	0	8	0	0	0	0	0	0	8	0.660

5.4 Discussion

In previous chapters it was shown that low but significant levels of genetic differentiation exist among many *Paphies subtriangulata* and among many *P. australis* locations. However, the geographic patterns of differentiation were often unclear or in conflict with what might be expected based on life history characteristics of the species and known oceanographic current patterns. In cases such as these seascape genetics can be a useful tool to search for associations among the patterns of genetic variation observed within a species and the environmental variation present in the marine environment (Riginos & Liggins 2013; Selkoe *et al.* 2008).

Three measures of genetic variation (F_{ST}, ϕ'_{ST} and allelic frequency) were used to gain more comprehensive insight into how environmental variation might be influencing different aspects of genetic variation. In general, many significant associations were found between the genetic and environmental variables, disproving the null hypothesis of no correlation between genetic and geospatial/environmental variation. However, results showed that some measures of genetic variation (i.e., allele frequency) were less able to be explained by environmental variation, as

evidenced by the non-significant BEST models for *P. australis* and, to a lesser extent, *P. subtriangulata*. Genetic measures based on F_{ST} and ϕ'_{ST} showed a stronger association with environmental variation. The *P. subtriangulata* GLZ models run with F_{ST} tended to result in lower *p*-values compared to ϕ'_{ST} but *p*-values were similar when both indices were tested for *P. australis*. F_{ST} and ϕ'_{ST} were often associated with different sets of variables for both species, although some commonalities were seen. Despite some issues with low significance, GLZ and BEST models did show some commonality in the variables that were included in best fit models; it is these common variables that I have focused on when attempting to summarise the results of all models in Table 5.10.

5.4.1 Key variables driving genetic population structure

5.4.1.1 Paphies subtriangulata

The most striking aspect of genetic population variation in *P. subtriangulata* was the differentiation of the Chatham Island population from the remainder of the mainland New Zealand locations. The large genetic distance between the Chatham Island population and mainland populations, in terms of F_{ST}, ϕ'_{ST} and allelic frequency, made this population an outlier. It was also an outlier in many of the environmental variables that were used. As a consequence it was difficult to assess the meaningfulness of results from GLZ models as they were all highly significant; addition and removal of different variables made little difference to model performance. Similar results were seen in BEST analyses where a large number of models with the same correlation coefficients were observed. Excluding the Chatham Island population resulted in a smaller number of significant models and smaller correlation coefficients, as lower levels of genetic variation among mainland locations made it difficult to identify significant correlations. However it was easier to assess the importance of different variables on model performance.

Results from GLZ and BEST analyses using all *P. subtriangulata* locations were in agreement that geospatial and environmental variables were both important in explaining genetic variation among populations. Both analyses identified geographic distance among populations as an important explanatory variable, suggesting that the large distance between mainland New Zealand and Chatham Island populations is the most significant factor determining large scale genetic population structure in *P. subtriangulata*. In terms of

Table 5.10: Summary and comparison of GLZ and BEST results. The variables that were most important in explaining genetic variation (as determined by model *p*-values, effect *p*-values, R_S coefficients and number of times a variable appeared in models) are shown along with any correlated variables that were not included in model runs. The relative importance of geospatial and environmental variables in explaining genetic variation is also summarised.

Test	Analysis Type	Important Variables	Correlated Variables	Comments
P. subtriangulata all populations	GLZ	geo_dist orb_v_mean SSTgrad bed_slope		Geospatial and environmental variables were both important in explaining genetic distance among populations
	BEST	geo_dist orb_v_mean rad_mean SSTgrad bed_slope	SSTanamp	Geospatial and environmental variables were both important in explaining allelic variation between populations
P. subtriangulata mainland populations	GLZ	geo_dist SSTanamp bed_slope	SSTwint, rad_mean	Environmental variables explained a larger amount of genetic distance than geospatial variables
	BEST	geo_dist bed_slope	SSTwint, rad_mean	Geospatial and environmental variables were both important in explaining allelic variation between populations
P. australis	GLZ	latitude geo_dist SSTgrad tidal intertidal tot_vol_SHT	rad_mean, SSTwint spring_TR mean_depth, river_discharge	EEC variables were able to better explain genetic distance than MEC variables
	BEST	latitude SSTgrad tidal bed_slope area_HT catch_area river_discharge	rad_mean, SSTwint tot_vol_SHT	MEC variables were able to better explain allelic variation among populations than EEC variables

environmental variables, both analyses identified a similar set of key variables (SSTgrad, orb_v_mean and bed_slope). Results were less clear when the Chatham Island population was excluded from analyses; GLZ analyses suggested that MEC variables, but not geospatial variables, were significantly correlated with population genetic structure, whereas BEST did not find any significant models and the importance of geospatial versus environmental variables was similar. In general, the results of these analyses using all locations versus only mainland locations have shown that a similar set of geospatial and environmental variables (geographic distance, SST, bed slope) are associated with genetic variation in *P. subtriangulata* regardless of spatial scale. Although model performance was poorer for mainland populations, due to lower levels of genetic variation among these locations, the results suggest that similar processes are working at different geographic scales to influence genetic population structure in this species.

Overall, distance among populations appears to be the most significant variable influencing *P. subtriangulata* genetic population structure, but there is evidence that at smaller spatial scales environmental variation might also be important. Sea surface temperature has been implicated in population genetic structuring of other marine species in New Zealand waters (Wei *et al.* 2013b) and elsewhere (Banks *et al.* 2007; Selkoe *et al.* 2010). Mean orbital velocity and bed slope are both related to the physical morphology of the sea bed. Given that *P. subtriangulata* are buried within the sea bed on high energy surf beaches, variation in current velocity and the slope of the bed between sampling locations could have the potential to affect the genetic makeup of the population via adaptation to local conditions.

5.4.1.2 Paphies australis

It is important to keep in mind when interpreting analyses for *P. australis* that many of the models (particularly for BEST) were not significant or showed low correlation coefficients. This problem is further compounded by the observation that for some models, GLZ and BEST analyses tended to highlight different sets of variables as being important. However, by taking a conservative approach of focussing only on those variables that were common between the two approaches we can be confident that the associations observed are meaningful.

GLZ results showed that models with geospatial and MEC variables tended to be less significant, whereas models with EEC variables were highly significant with most variables included in the best fit model. Since *P. australis* is an estuarine organism it is not unexpected that EEC variables were able to explain more genetic variation among populations as they are more representative of the physical environment experienced by this species, compared to MEC variables which are measured from the open coast. However, interpretation of model significance is challenging as the addition or removal of EEC variables had little effect on *p*-values. BEST analyses were unable to find significant models using any of the sets of variables. BEST results differed from GLZ results as BEST models that included MEC variables tended to have higher correlation coefficients and lower (although still non-significant) *p*-values. Results from the two approaches were similar in that models with EEC variables resulted in many different models with the same correlation coefficient.

GLZ and BEST analyses both identified latitude as the most important geospatial variable, and indicated that SST and radiance levels also play a role in shaping *P. australis* genetic variation. Similar to *P. subtriangulata*, variables associated with the characteristics of the intertidal area inhabited by *P. australis* (tidal, bed_slope, intertidal) and with the size of the estuaries/harbours sampled (catch_area, area_HT, tot_vol_SHT) were also common in best fit models. Both GLZ and BEST analyses also identified that the volume of freshwater entering the estuary, in terms of river discharge, was an important explanatory variable. These results indicate that, while latitude and distance among estuarine habitat appear to have an important influence on genetic population structure, differences in the physical nature of the beds where *P. australis* are found as well as the extent of the surrounding habitat could also be influencing genetic structure in this species. It is difficult to pinpoint specific habitat variables that could be driving population structure due to the aforementioned problem of highly significant or similarly correlated models that are unaffected by addition or removal of specific EEC variables.

5.4.2 Links between genetic and environmental variation

Chapters 3 and 4 reported isolation by distance (IBD) population structure for both *P. subtriangulata* and *P. australis*, the significance of which varied depending on the geographic scale being considered. Therefore it was not surprising that geospatial variables such as geographic distance and latitude appeared often in best fit models. This is

particularly evident for the Chatham Island *P. subtriangulata* population, where distance from external sources of mainland New Zealand recruits is likely to have resulted in genetic differentiation of this population. Distance between Chatham Island and mainland New Zealand has also been invoked as an explanation for genetic differentiation in other coastal marine species (Goldstien *et al.* 2006; Hickey *et al.* 2009; Stevens & Hogg 2004; Will *et al.* 2011). Among mainland *P. subtriangulata* many populations were not genetically differentiated from each other, meaning that geospatial and environmental variation probably do not have a large influence on the genetic structure of these populations. Where populations were differentiated it is possible that distance from other populations plays a role; the distance between many populations is further than larvae can potentially cover within their lifespan (Shanks 2009; Shanks *et al.* 2003). When coupled with coastal eddies that restrict larval dispersal (see chapter 3), this could drive differentiation of isolated populations.

Sampling of *P. australis* populations covered a large geographic and latitudinal range so patterns of genetic variation associated with geographic distance and latitude are not unexpected. In chapter 4 I discussed how the distance between suitable habitat (i.e., habitat availability) can drive the IBD pattern commonly reported for estuarine organisms. I concluded that patterns of population differentiation seen in *P. australis* provide some support for this concept, and the significance of latitude and geographic distance in the models presented here is consistent with this conclusion. However, I also reported that contrasting patterns of small scale genetic differentiation in some parts of New Zealand versus long distance dispersal events in other areas provide evidence that the relationship between geographic and genetic distance has become decoupled and is not a simple function of habitat availability. The results of GLZ models presented here provide further support for this observation: geospatial variables on their own were not significant, but were significant in combination with other EEC variables. These results suggest an interaction between processes operating at different stages during the larval life cycle (e.g., current patterns, distance between habitat, local adaptation), resulting in the complex patterns of observed genetic population structure.

Closely associated with latitude and distance are measures of SST. Sea surface temperature range from the north to the south of the New Zealand region is large, with

differences of up to 15°C observed during the summer months (Heath 1985). Therefore, SST is likely to play an important role in the genetic structure of New Zealand marine species (e.g., Wei *et al.* 2013b), possibly creating warm and cold-adapted populations, similar to patterns seen among Northern Hemisphere marine species (Case *et al.* 2005; Powers & Schulte 1998). Changes in SST associated with the convergence of currents and upwelling have also been proposed as an explanation for the genetic disjunction seen in many marine organisms around the Cook Strait area (Apte & Gardner 2002; Ayers & Waters 2005; Goldstien *et al.* 2006; Ross *et al.* 2011; Star *et al.* 2003; Stevens & Hogg 2004). Seascape genetic analyses of the New Zealand greenshell mussel *Perna canaliculus* provide evidence that genetically differentiated northern and southern groups may represent warm and cold-adapted populations of this species (Wei *et al.* 2013b). The three genetically distinct *P. australis* groups described in chapter 4 could similarly represent warm and cold-adapted populations.

There were also indications that SST could be influencing genetic structure among *P. subtriangulata* populations. The SST conditions experienced within the range of sampled populations is less extreme than those described above for *P. australis*, but SSTwint values still varied by approximately 4°C over the range sampled. In particular SST could help explain the differentiation of the Ruakaka population, which is known to be influenced by the warmer subtropical waters of the East Auckland current (Stanton *et al.* 1997), potentially resulting in warm-adapted northern populations. Variation in allele frequency and expression levels of heat shock proteins have been shown to play a role in responses to temperature gradients in other bivalve species (Dutton & Hofmann 2009) and marine fish (Hemmer-Hansen *et al.* 2007). Heat shock proteins are associated with general stress responses in many species (Hoffmann & Willi 2008), and are likely to respond in a similar way in *P. subtriangulata* and *P. australis*, presenting a method for testing the underlying genetic mechanisms of how these species respond to changes in SST.

More difficult to interpret are the results suggesting that the physical characteristics of the habitats of *P. subtriangulata* and *P. australis* are influencing their genetic population structures. For *P. subtriangulata*, orb_v_mean and bed_slope can be considered indices of beach type, which can range from high-energy reflective to low-energy dissipative beaches. Burrowing ability, shape and density of surf clams is known to vary according to beach type

(McLachlan *et al.* 1995) and in the deep water tuatua, *Paphies donacina*, large variability in terms of size frequency distribution has been observed along the length of a beach, possibly due to variation in environmental conditions (Marsden 2000). If there is an underlying genetic basis to variation in the physical characteristics of the organism (i.e., burrowing ability, size, shape, physiological stress tolerance, etc.), it is possible that variation in orbital velocity and bed slope could be placing selective pressure on populations, and individuals may become adapted to particular beach types. In particular there is a strong relationship between F_{ST}/ϕ'_{ST} whereby beaches with a steeper bed slope are characterised by lower F_{ST}/ϕ'_{ST} values, possibly because only select morphotypes are adapted to live in a reflective style beach. Physical characteristics of the beach could be particularly relevant to differentiation of the Kakamatua population, which is located in a sheltered harbour compared to the open coast habitat of other sampled locations. The observed relationship between F_{ST}/ϕ'_{ST} and bed slope suggests that one or more loci are somehow linked to physiological responses to bed slope in the organisms, a hypothesis that could be further investigated by testing how allele frequencies change with variations in bed slope.

Physical parameters of estuary habitat were also important in explaining genetic variation in *P. australis*. It is not unreasonable to expect estuaries of different size and shape to be characterised by variable bivalve community composition (Cole *et al.* 2000), and for these communities to experience different selective pressures. Variables associated with the intertidal area inhabited by *P. australis* often featured in best fit models. Intertidal areas are known for their extreme range of environmental conditions, so the size of the intertidal area and flow of tidal currents could represent strong selective agents (Schmidt *et al.* 2008). Indices of estuary size also featured in best fit models and it could be expected that larger estuaries could accommodate larger, more variable populations. However, there does not seem to be any clear relationship between these environmental variables and effective population size (data not shown), possibly due to the low statistical power for many of the *P. australis* models. Interestingly, freshwater discharge was associated with genetic variation in GLZ and BEST analyses. Exposure to freshwater is known to significantly decrease survivorship for *P. australis* (McLeod & Wing 2008) and has previously been used to explain patterns of genetic variation in marine organisms (Jørgensen *et al.* 2005; Puritz & Toonen 2011). Variation in freshwater discharge among estuaries could be a strong selective

force for *P. australis* populations depending on an individual's tolerance to low salinity conditions.

Variation in selective pressure related to the physical properties of different estuaries could help explain some of the small scale population differentiation observed in *P. australis* if it has resulted in local adaptation of populations. Selection has been implicated in previous studies of estuarine organisms where patterns of genetic variation are associated with changes in environmental variables (Schmidt *et al.* 2008). For example variation in the *Lap* locus is associated with variation in salinity in *Mytilus edulis* (Hilbish & Koehn 1987) and variation in the *Ldh-B* locus is associated with temperature in *Fundulus heteroclitus* (Crawford & Powers 1989; Powers & Place 1978). The effect that variable intertidal conditions can have on genetic variation is illustrated by *Littorina saxatilis*, where variation in the *Aat* locus is related to the amount of physiological stress experienced by low versus high intertidal organisms (Panova & Johannesson 2004).

The links discussed between genetic and environmental variation in *P. subtriangulata* and *P. australis* are purely speculative at present. However the candidate loci identified in previous studies form a good basis for further investigating the mechanisms behind how environmental variation drives genetic variation in these two species. Furthermore, the large amount of next generation sequence data generated for these two species provides an excellent genomic resource for identifying the candidate loci mentioned above if homologous sequence is available. Establishing the nature of these relationships provides a promising area of further research.

Finally, the comparative approach taken in this study has allowed for identification of common environmental drivers of genetic population structure (i.e., geographic distance, SST, physical habitat features) and forms a good basis for simplifying management/conservation strategies across multiple species (Riginos & Liggins 2013). However, it can also be useful to look at the ways that two closely related species contrast, i.e., are there any species specific environmental variables that can help to explain the differences seen in patterns of genetic population structure? The main difference observed between the two study species was that MEC variables were poorly correlated with genetic variation in *P. australis* whereas models containing EEC variables tended to be more

significant by comparison (particularly for GLZ analyses). Because the EEC is more likely to represent the environmental variables that *P. australis* populations are exposed to, this finding provides evidence that habitat is contributing to the different patterns of genetic population structure seen in these two species.

5.4.3 Limitations to the seascape genetics approach and future directions

Seascape genetics assumes that we have some prior expectation that the geospatial and environmental variables used will be able to explain the genetic variation seen among populations in the species of interest. In fact the opposite situation might be true; none of the variables used in an analysis may truly be associated with genetic variation and other unknown variables might be more important. Often there are limits to the variables available for testing as they must be drawn from existing and readily available data sets, such as the MEC and EEC variables used in this study. The upside of using such readily available data sets is that results from studies across multiple species using the same environmental data are comparable.

Use of the MEC data set has several drawbacks. Firstly, the nature of the environmental variables used in this data set is that they will fluctuate on a temporal basis and so are averaged for ease of use. However, we know that genetic structure in marine organisms can be temporally patchy, and it is often the type of extreme events that are not captured by 'average conditions' that can highly influence estimates of genetic structure, e.g., rare long-distance dispersal events (Selkoe *et al.* 2010). Secondly, the MEC data set was designed to cover the entire New Zealand exclusive economic zone (EEZ) and be applicable to a broad range of biological applications (New Zealand Ministry for the Environment 2005) so the resolution of some variables is coarse and in places does not extend all the way to the coast. This was particularly problematical for *P. australis* as often values were not available for sites within harbours or estuaries, meaning the closest available value had to be used (which is not necessarily a good proxy for the real value at the sampling location) or treated as a missing data point. Results reported here demonstrate that the EEC data set represents a more appropriate set of variables to describe genetic variation in *P. australis*. Additionally, EEC variables are largely based on relatively stable physical characteristics of the estuary that are unlikely to change over time.

The EEC data set contains a large number of variables that are correlated, as illustrated by PCA plots, presenting a challenge when determining an appropriate independent data set for testing. Few seascape genetic studies have attempted to test as many variables as in the present study, usually focussing on a small set that are believed *a priori* to have a relationship with the genetic measure of interest (Riginos & Liggins 2013). Results from the present study illustrate the effect of including too many variables in an analysis: runs with large numbers of variables (particularly EEC variables) performed poorly, resulting in models with highly significant *p*-values or similar correlation coefficients, which included most variables. Addition or removal of variables made little difference to results, suggesting that the large number of variables made it difficult for the true signal to be detected. However the risk of excluding too many variables is that important and unexpected associations could be missed. In the present study these problems were largely overcome by splitting variables into geospatial and environmental groupings. Furthermore, by using different but complementary approaches with several genetic measures and analysis methods, it was possible to detect variables that were deemed significant across methods. This approach can give us confidence that a true signal was detected.

Seascape genetics is still an emerging field and the number of studies using these methods is limited. Many different methods and measures of genetic distance are used (Storfer *et al.* 2007, 2010) and making meaningful comparisons among species is challenging. Few studies have used the GLZ and BEST approaches implemented here. Perhaps the most comparable study is that of Wei *et al.* (2013b), described previously, which used the same methods and variables (genetic, geospatial and MEC) to test for associations between genetic and environmental variation in the New Zealand greenshell mussel *Perna canaliculus*. Furthermore, comparative seascape genetics studies that assess patterns between two or more species (such as in the present study) are rare, presumably due to the difficulties of working with multiple species (Riginos & Liggins 2013). However, such studies can be highly valuable for detecting common drivers of genetic patterns within a biological system. An excellent example of this approach comes from Selkoe *et al.* (2010), who used a linear modelling method to show that the same environmental variables were driving similar patterns of genetic structure in three different coastal marine species in Southern California. This study demonstrates how the power to interpret subtle patterns of

population structure can be increased by comparing genetic patterns and environmental processes across multiple species. This level of understanding is needed if we want to make robust management and conservation decisions, particularly in a marine environment that is as environmentally complex and ecologically diverse as New Zealand.

Finally, seascape genetic analyses have begun to demonstrate the types of environmental processes that might be responsible for driving local adaption to marine environments (Moller-Hansen & Hemmer-Hansen 2007). The results presented here suggest that local adaptation may be occurring for *P. subtriangulata* and *P. australis* populations and previous studies have highlighted several promising candidate genes that could provide the mechanistic links between local adaptation and environmental variation. This allows us to formulate hypotheses that can be further tested, particularly by measuring allele variation or expression in candidate genes across environmental gradients.

5.4.4 Conclusions

Seascape genetic analyses for *P. subtriangulata* and *P. australis* highlight the value of using an integrative biological, ecological and physical oceanographic approach to investigate subtle patterns of population genetic variation. By using standardised methods across species it is possible to determine both the common and species-specific factors that may be responsible for shaping genetic variation of coastal marine populations in New Zealand. For both *P. subtriangulata* and *P. australis* geographic distance between populations, sea surface temperature and measures of the organism's physical habitat were associated with genetic variation. Species-specific patterns were also observed between the two species, and it is notable that estuarine environmental variables were much better at explaining genetic variation in *P. australis* than variables measured from an open coastal environment.

Studies of coastal marine species tend to emphasise how distance between suitable habitat drives population differentiation; a valid observation seeing as habitat availability is necessary for recruitment to occur. The present study supports this concept as geographic distance was identified as a potential driver of population differentiation. In chapter 4 I emphasised how habitat availability and oceanic processes can interact to modify population structure in coastal marine species. This is further supported by the associations

observed between distance, latitude, sea surface temperature and genetic variation, which are likely to be responsible for the three genetically differentiated groups observed for *P. australis* and some of the isolated cases of genetic differentiation observed in *P. subtriangulata*.

The physical measures of beach and estuary habitat that were identified as significant suggest that adaptation to local environments could also play a role in some of the small scale genetic differentiation seen. This observation is significant; it implies that not only is larval dispersal and recruitment modified by oceanographic processes and habitat availability, but once recruitment to suitable habitat has been successful, the environmental characteristics of the habitat might further modify genetic variation by favouring locally adapted individuals for post-settlement survival. Presently the mechanisms behind local adaptation are unknown but several promising candidate genes for further investigation have been identified.

The potential for seascape genetics to explore previously unknown associations between a species' genetic population structure and its environment are substantial. This type of comparative study shows how two species with differing but complex patterns of genetic variation can both be influenced by common environmental drivers. It seems likely that the same environmental variables could be influencing the population genetic patterns seen across a range of species and it is encouraging that results from this study correspond to those reported for *P. canaliculus* by Wei *et al.* (2013b). As this emerging field continues to be applied to other New Zealand marine species it will be of interest to see if the patterns reported here for *P. subtriangulata* and *P. australis* will be observed in populations of other species.

6 General Discussion

6.1 Context and aims of the research

The New Zealand marine realm is unique in its size, complexity of oceanographic/environmental processes and diversity of species (Gordon *et al.* 2010; Laing & Chiswell 2003), all of which present a challenging but fascinating situation in which to study the processes that shape the genetic population structure of marine species. Not only are genetic studies of marine species contributing to a growing body of knowledge on how and why genetic population structure is formed, but there is a growing realisation that this knowledge is required for the successful management and conservation of marine species. Increasingly there will be a demand for this information in the face of growing exploitation of marine resources and with the threat that climate change is expected to pose to the oceans and the species that inhabit them in the near future (Waples & Naish 2009). Population genetics represents one tool that can potentially be used to tackle these challenges. The rapid development of computational power and analytical approaches means that we are increasingly able to use genetic tools to address a broader range of ecologically relevant questions about populations of marine organisms.

In this study population genetic techniques have been applied to two New Zealand species of surf clam – *Paphies subtriangulata* (tuatua) and *P. australis* (pipi). Both species support recreational, customary and commercial fisheries but it is unknown how stocks of either species are structured and whether harvest is sustainable (New Zealand Ministry for Primary Industries 2013a; b). Furthermore, little is known about population dynamics of New Zealand surf clams in general, or for species that inhabit estuarine and sandy open coast habitats (Gardner *et al.* 2010). The overall goal of this thesis was to determine genetic population structure in *P. subtriangulata* and *P. australis*, estimate rates and patterns of migration among populations (including the location of possible barriers to gene flow), and to examine the relationship of these to environmental variation to identify some of the important processes that might result in similar or disparate patterns of genetic structure and migration.

This thesis began with a description of novel microsatellite markers developed for *P. subtriangulata* and *P. australis* using next generation sequencing (NGS) methods (chapter 2). In chapter 3 the genetic population structure of *P. subtriangulata* was determined and levels of migration among locations estimated, focussing on the paradigm that species occupying open coast habitats have well connected populations and experience few restrictions on dispersal. Chapter 4 presented similar analyses for *P. australis*, which focussed on the interaction of habitat availability versus oceanographic processes in structuring populations of estuarine species. Finally, the analyses presented in chapter 5 were a test for associations between geographic patterns of genetic variation and environmental variation. The purpose of this latter section of the study was to understand how variation in environmental processes and the ecology of the organism influence patterns of genetic population structure, particularly where small scale 'chaotic' patterns of differentiation exist.

In this discussion I firstly review the major findings of the study. Next, I present a synthesis of the major findings of chapters 3, 4 and 5. These chapters addressed separately the patterns of population differentiation and migration observed for each species, and associations with environmental variation. To understand how patterns of genetic population structure arise and are maintained it is necessary to compare and contrast results from both species, and to compare these findings to what is already known about population structure and connectivity in other coastal marine species from New Zealand. Thirdly, I discuss the management and conservation implications that arise from the research. Finally, I discuss how this study has significantly contributed to furthering understanding in the field of marine population genetics in New Zealand and highlight areas of further research.

6.2 Major findings

The patterns of genetic diversity, population structure and connectivity observed for *P. subtriangulata* and *P. australis* were similar to what has been commonly reported for other species of highly fecund marine invertebrates (Hedgecock *et al.* 2007). In general, gene flow was high and genetic population structure weak. However, significant differentiation was observed at both large and small spatial scales in both species. These

findings also emphasise how connectivity between populations can be highly variable over time, as evidenced by the differences seen in short and long-term migration estimates.

Microsatellite markers developed for *P. subtriangulata* and *P. australis* were characterised by homozygote excess and deviations from Hardy-Weinberg equilibrium (HWE; chapter 2). The main cause of HWE deviation was likely to be null alleles, which was not unexpected given the high mutation rate that characterises the genomes of marine invertebrates (Hedgecock *et al.* 2004; Launey & Hedgecock 2001). However, null alleles were unlikely to be the sole contributor to HWE deviation; spatial/temporal population structure could also have played a role and there was evidence that one *P. australis* locus (*Paus_9*) may be strongly influenced by selection. Overall, levels of null alleles were unlikely to be significantly influencing the conclusions of this study (Chapuis & Estoup 2007); if anything they appeared to be masking population structure.

The main feature of *P. subtriangulata* genetic population structure (chapter 3) was the highly significant differentiation of the Chatham Island population from other locations on the New Zealand mainland. Contemporary migration between Chatham Island and the mainland appeared to be severely restricted. However, alleles were shared between Chatham Island and mainland populations, suggesting that historical gene flow or rare dispersal events have occurred. Mainland populations showed genetic structure that was consistent with an 'open' model of population connectivity. Many locations were undifferentiated and long-term migration estimates showed high levels of connectivity. However, this pattern was not consistent across the whole range sampled for this study; in some locations significant differentiation of populations was observed. This finding indicates that while large parts of the *P. subtriangulata* range were free from restriction to dispersal, barriers do exist in some locations, allowing for differentiation of isolated populations. In *P. subtriangulata* areas of restricted gene flow coincided with geographic distance among populations (i.e., between mainland New Zealand and Chatham Island), coastal current features (i.e., North Cape Eddy, East Cape Eddy, Wairarapa Eddy) and physical landforms (i.e., Kakamatua population located in the relatively enclosed Manukau Harbour).

When results from the present study were compared to a previous study of allozyme variation in *P. subtriangulata* by Smith *et al.* (1989) some differences in genetic population

structure were evident among mainland locations. However, both studies reported significant differentiation of the *P. subtriangulata* population at Chatham Island. Differences in genetic population structure are most likely due to the different marker types used in the two studies, which has previously been used as an explanation where differences are seen in genetic population structure within a species (e.g., Apte & Gardner 2002, 2001; Lidgard 2001; Ross *et al.* 2011). Additionally, low levels of differentiation among mainland locations or changes in population structure over time could also be responsible for the differences observed between the two studies. If gene flow patterns are variable over time it is conceivable that *P. subtriangulata* population structure could have changed in the 24 year period between this study and the one conducted by Smith *et al.* (1989).

A further finding from chapter 3 surrounds the accuracy of the reported distribution of *P. subtriangulata*. Reports that *P. subtriangulata* is distributed around the entire South Island of New Zealand (Powell 1979) were found to be inaccurate. In this study *P. subtriangulata* could not be found south of 42° S. Smith *et al.* (1989) also failed to include any populations further south than the Golden and Tasman Bay areas at the top of the South Island, apart from one Stewart Island population. Records held by the Museum of New Zealand Te Papa Tongarewa show that apart from a population on Stewart Island, there are limited reports of *P. subtriangulata* occurring south of 42° S (B. Marshall, Te Papa Wellington, personal communication). Many of these reports date from prior to 1960, so it is possible that species misidentification has occurred or that the species distribution has changed over time. If *P. subtriangulata* is present in south of 42° S it must be extremely rare or very patchy in its distribution.

Overall *P. australis* genetic population structure (chapter 4) was in agreement with the IBD model that is expected for estuarine species where distance between discrete habitats results in a stepping-stone model of migration (Bilton 2002). However, habitat availability alone cannot explain the patterns of genetic population differentiation that were observed. Three areas of reduced gene flow were observed, which coincided with coastal current features, resulting in three genetically differentiated population groups. Small scale patterns of differentiation were present among North Island locations, whereas South Island locations showed less genetic divergence. More complex current circulation around parts of the North Island could be contributing to this pattern of genetic differentiation by restricting

larval dispersal and promoting self-recruitment. Migration is likely to be highly variable, meaning that genetic connectivity among populations is present when averaged over long periods of time but in the short-term levels of connectivity could be more intermittent.

Seascape genetic analyses provide a further explanation for small scale population differentiation in *P. australis* as there was evidence suggesting that adaptation to local estuary conditions may have occurred (chapter 5). Sea surface temperature (SST) and the geospatial location of populations could influence genetic structure in *P. australis,* but the most significant finding was that the physical measures of the estuaries/harbours sampled were most strongly associated with genetic variation among locations. For *P. subtriangulata*, geographic distance between the Chatham Island and mainland New Zealand populations was strongly associated with genetic variation. In comparison, correlation between genetic variation and coastal environmental variables was not as strong among mainland populations (i.e., *p*-values were higher), probably due to lower levels of genetic differentiation among locations. Where correlations were detected, they were suggestive of local adaptation to beach characteristics.

6.3 Synthesis

When devising management or conservation strategies for a marine region it is useful to know where common patterns of genetic population structure and connectivity exist across multiple species. This information can be used to make predictions about patterns and drivers of genetic population structure, and/or to simplify management and conservation strategies across similar species. Likewise, if we can identify the ways in which genetic population structure and connectivity differ between species and understand why these differences arise, then this information can also assist with management. To achieve these outcomes comparative studies with standardised methods, such as the present one, are preferable (Bohonak 1999). *Paphies subtriangulata* and *P. australis* are closely related species that have similar reproductive strategies. However, they differ in terms of their habitat preferences, population densities and geographic distributions. Therefore, similarities in genetic population structure could be a result of reproductive strategy and it is reasonable to assume that similar population structure could occur in species with similar reproductive strategies, i.e., other species of surf clam. Differences in genetic population structure and connectivity are likely to be a result of habitat preference and species

occupying similar habitats could share similar patterns of population structure (e.g., estuarine species).

The patterns of genetic population structure and connectivity observed in *P. subtriangulata* and *P. australis* emphasise how population dynamics of marine species and the processes that drive them can be variable over spatial and temporal scales. Both species showed differing patterns of genetic population structure in different parts of the sampled range. In some areas connectivity was high and consequently, genetic population structure was weak. This was particularly evident along the west coast of the North Island for both species and continued along the west coast of the South Island for *P. australis*. In comparison, the north-east of the North Island showed more restricted dispersal and stronger population differentiation in both species. Current circulation systems off the coast of Northland, Bay of Plenty and Wairarapa were implicated in the restriction of larval dispersal for both species. Estimates of migration in both species provided evidence for temporal variation in larval dispersal and recruitment. Contemporary migration estimates showed that the number of source populations was limited, migration was often unidirectional, levels of migration were high among some locations (often over long distances) and absent among other locations (even among nearby locations), and that self-recruitment could be high. In comparison, historical migration estimates averaged over many generations showed that all locations were genetically well connected.

The differences seen between contemporary and historic migration estimates are consistent with a 'crinkled' model of connectivity (Ovenden 2013). Barriers to dispersal are present in similar locations for both species, but rather than being absolute, the barriers are permeable and dispersal is restricted, rather than prevented. This form of connectivity can lead to a situation where populations are characterised by 'chaotic' genetic patchiness (CGP) (Hellberg *et al.* 2002; Johnson & Black 1982; Selkoe *et al.* 2006). Species characterised by CGP often show significant genetic differentiation over small spatial scales (similar to that observed in *P. australis*, and to a lesser extent in *P. subtriangulata*) related to site-specific recruitment history, despite genetic similarity of populations at larger spatial scales (i.e., low global F_{ST}; Hedgecock & Pudovkin 2011). A further process likely to be contributing to CGP is 'sweepstakes reproductive success' (SRS; Hedgecock 1994), whereby the chance of larval cohorts successfully recruiting to an adult population depends on encountering favourable

oceanographic conditions, resulting in spatially and temporally variable patterns of dispersal and recruitment. As a consequence, a small number of breeders can be responsible for the majority of recruitment in a population. This can result in the small effective population sizes that were observed in *P. subtriangulata* and *P. australis*, relative to what has been reported in other marine invertebrates.

Temporal variability in recruitment events and oceanographic processes mean that the CGP patterns observed in the study species may be present on short time scales and change quickly depending on the balance of gene flow and genetic drift. The theoretical relationship between genetic drift and gene flow (gene flow overrides drift when $m > 1/4N_e$; Wright 1931) suggests that genetic drift should act relatively rapidly on *P. subtriangulata* and *P. australis* populations. Two populations may be isolated for a period of time long enough to allow for development of genetic differentiation via drift, mutation and selection. However, a sudden rare influx of migrants may homogenise allele frequencies among the two populations. For these reasons, the possibility of 'non-equilibrium' conditions must be kept in mind when considering areas where populations are genetically undifferentiated, particularly for *P. subtriangulata*. Just because two locations are genetically undifferentiated does not necessarily mean they are connected by consistent gene flow events. There is an increasing awareness that temporal genetic structure may be just as significant as spatial genetic structure in marine species and there is growing evidence for a lack of temporal stability in the processes that drive genetic population structure in many species (Hedgecock & Pudovkin 2011; Selkoe *et al.* 2006; Siegel *et al.* 2008). The results presented here provide further evidence that a lack of temporal stability for some *P. subtriangulata* and *P. australis* populations may need to be incorporated into management strategies.

Geographic distance between populations was identified as an important factor shaping genetic population structure in both species, supported by the significant IBD signal detected in Mantel tests. Although Mantel tests are known to be problematical due to high type I error and can be influenced by hierarchical structure (Legendre & Fortin 2010; Meirmans 2012) the correlation between genetic variation and geographic distance detected by generalised linear models (GLZ) provides further support for the importance of geographic distance. This finding that geographic distance between available habitat can restrict dispersal in the two study species provides support for the paradigm in coastal

marine species that connectivity is higher among neighbouring populations than it is among distant ones (i.e., a stepping-stone migration model; Slatkin 1993). However, results from this study suggest that the relevance of the stepping-stone model is dependent on the spatial scale being considered. In particular, distance between mainland New Zealand and Chatham Island acted as a barrier to dispersal for *P. subtriangulata* but distance was less of a barrier among mainland populations and IBD was weaker as a consequence. Despite an overall IBD pattern in *P. australis*, the barriers to dispersal and genetic patchiness observed at smaller spatial scales resulted in the formation of a hierarchical structure (i.e., three genetically distinct groups) that is likely to be a more significant contributor to genetic population structure in this species. These findings suggest that for both species a decoupling of the relationship between distance and migration has occurred in some locations, probably due to localised oceanographic processes that create restrictions to gene flow and/or local adaptation to specific beach/estuary types or temperature regimes.

Local adaptation has been proposed as a mechanism for genetic population differentiation where small scale, geographically unstructured genetic patterns have been observed (Moller-Hansen & Hemmer-Hansen 2007; Selkoe *et al.* 2010). There is likely to be an intricate relationship between local adaptation and complex circulation patterns (particularly in the north and east of the North Island) that are driving the observed patterns of genetic population structure. In *P. australis* and the Kakamatua *P. subtriangulata* population, the restrictive nature or harbour/estuary habitat is likely to be further complicating dispersal processes.

Overall, similarities were observed in the genetic population structure of *P. subtriangulata* and *P. australis* in terms of large scale IBD genetic structure coupled with small scale CGP. This structure is likely to be driven by temporally variable migration rates and SRS, modified by stochastic current patterns, semi-permeable barriers to dispersal and possibly influenced by local adaptation. These factors result in the type of non-equilibrium population dynamics that are often seen in marine invertebrates with large population sizes and high reproductive output (Hedgecock *et al.* 2007). In *P. subtriangulata* and *P. australis* these findings illustrate how similar reproductive strategies can result in similar patterns of genetic population structure and connectivity. However, the differences observed between the two species can probably be attributed to habitat preference and population size.

Despite similarities in pelagic larval duration, the relatively continuous nature of open coast beach habitat preferred by *P. subtriangulata* seemed to promote higher levels of connectivity, resulting in weaker genetic population structure consistent with an 'open' model of population structure. In comparison, the more separated estuary/harbour habitat of *P. australis* has resulted in stronger patterns of genetic population subdivision. The genetic demarcation seen around the Cook Strait area in *P. australis* was not observed in *P. subtriangulata* (i.e., *P. subtriangulata* populations north and south of Cook Strait were genetically undifferentiated). However, there are various reports from different species regarding the location of this break (Gardner *et al.* 2010); in *P. australis* the break appears to be north of Cook Strait. It is entirely possible that such a break occurs in *P. subtriangulata* but is located south of Cook Strait (i.e., south of the sampling range of this study) in a similar location to that reported for greenshell mussels (*Perna canaliculus*; Apte & Gardner 2002). The differences observed in genetic population structure between the two study species are supported by estimates of migration. Although estimated migration rates were higher for *P. australis*, there was a higher degree of variation in those estimates. Furthermore, *P. australis* populations were larger and more genetically diverse than *P. subtriangulata* populations, so even if migration is high the homogenising effect of gene flow is likely to have a less significant influence on allele frequencies in larger *P. australis* populations.

Selection and local adaptation could play a more significant role in shaping *P. australis* genetic population structure, compared to *P. subtriangulata*. Estuarine habitat is known for its extreme environmental variation, which can place strong selective pressure on organisms inhabiting such environments (Bilton 2002). Furthermore, within an estuary *P. australis* beds are often spatially localised with high population densities, compared to the more dispersed distribution of *P. subtriangulata* (personal observation). In locations where *P. australis* population density is high, locally adapted individuals could be favoured over immigrants, maintaining population subdivision even if levels of migration are high. Lower population density for *P. subtriangulata* means that this species is unlikely to experience the same degree of selective pressure but genetic population structure could change more rapidly in response to changes in the migration-drift equilibrium.

6.3.1 Comparison to other coastal marine species in New Zealand

Very few studies of coastal marine species that had a New Zealand-wide sampling range have failed to report some sort of genetic population subdivision. Studies that have reported no structure generally used less discriminatory allozyme markers (e.g., Apte & Gardner 2001; Lidgard 2001) or were species with long pelagic larval duration (e.g., Hickey *et al.* 2009; Ovenden *et al.* 1992). This observation suggests that the complexity of the New Zealand coastal marine environment is conducive to producing genetic differentiation among populations of coastal marine species to varying extent. In *P. subtriangulata* and *P. australis* areas of restricted dispersal were observed along the north-east coast of the North Island, at East Cape and along the Wairarapa coastline, at Cook Strait, on the west coast north of Fiordland and between mainland New Zealand and Chatham Island. All of these areas of restricted dispersal are associated with pronounced changes in coastal current direction or complex circulation patterns (or distance in the case of Chatham Island) and patterns of genetic population subdivision associated with these areas have been reported for many other coastal marine species (as described in chapters 3 and 4). Many studies have reported IBD population structure (Hickey *et al.* 2009; Perrin *et al.* 2004; Veale & Lavery 2012) and in estuarine/fiord species enhanced population subdivision is commonly reported (Hickey *et al.* 2009; Knox *et al.* 2011; Miller 1997; Ross *et al.* 2011).

The role of local adaption in structuring populations of New Zealand marine species has been poorly studied to date but Wei *et al.* (2013b) suggest that it has played a role in the genetic structuring of *Perna canaliculus* populations, resulting in a warm-adapted northern population and a cold-adapted southern population, similar to the genetic structure seen in *P. australis*. As seascape genetic studies become more commonplace, this is a pattern that is likely to hold up for other species showing a similar pattern of north-south differentiation (e.g., Goldstien *et al.* 2006; Jones *et al.* 2008; Sponer & Roy 2002; Stevens & Hogg 2004; Waters & Roy 2004; Will *et al.* 2011). Other studies that have investigated local adaptation in the marine environment illustrate how it can have a significant impact on genetic population structure, even over very small spatial scales (Dutton & Hofmann 2009; González-Wangüemert *et al.* 2009; Hemmer-Hansen *et al.* 2007). The present study has illustrated how a combination of genetic and environmental data can be used to make comparisons across species, and has revealed a surprising level of

complexity in marine population dynamics across spatial scales. This finding has implications for how we manage populations of marine species in New Zealand.

6.4 Implications of the research

6.4.1 Fishery management

The results from this study have implications for the management of fisheries for *P. subtriangulata* and *P. australis* and potentially for management of other coastal fishery species. One surprising finding from this study was that even for a species where weak genetic structure and high connectivity were predicted, such as *P. subtriangulata*, isolated cases of genetic population differentiation can occur. For *P. australis*, higher levels of genetic population subdivision were predicted, but this study found a surprising level of small scale differentiation, even among locations where it was reasonable to assume that dispersal may be unrestricted (i.e., over the 100 km between Huia and Raglan). While stochastic current patterns, intermittent gene flow and high self-recruitment can probably explain much of the small scale genetic differentiation that was observed, there was often no obvious geographic pattern to the distribution of this genetic variation, so local adaptation could also play a role. In addition, there was evidence that null alleles could be masking population structure and some populations might not meet migration-drift equilibrium assumptions. This means that genetic differentiation between populations might actually be higher than estimated and connectivity could be low even among non-differentiated locations. As a consequence it may be necessary to reassess the spatial scales at which marine species are currently managed.

When connectivity is intermittent and populations unlikely to meet equilibrium expectations, then populations will be genetically connected when migration is averaged over long time periods, but levels of migration will not be high enough to demographically connect populations over short time scales. Demographic connectivity is crucial for fisheries management; to ensure the on-going sustainability of a fishery it is necessary that the reduction in stock biomass due to fishing mortality is replenished by recruitment. The implication for *P. subtriangulata* and *P. australis* fisheries is that many stocks may not be demographically connected at the rates required for sustainable fishing. Instead, populations many be heavily reliant on self-recruitment and will need to be managed at a

localised scale as a consequence. While this study has provided evidence for historical genetic connectivity, the challenge is to determine if and where populations exist that exchange consistent levels of migrants at a high enough rate to achieve the demographic connectivity required for sustainable fishing. This could be achieved through oceanographic or Lagrangian particle modelling, or with temporally spaced sampling of genetic population structure. Where Lagrangian modelling has been used to track larvae of other New Zealand marine invertebrates it has shown that self-recruitment is likely to be high and sporadic, with larvae travelling no more than tens of kilometres from their natal sites (Broekhuizen *et al.* 2011; Stephens *et al.* 2006). If areas that receive consistent levels of recruits over time can be identified this will form the basis for a sustainable fishery and will help determine the appropriate geographic scale on which stocks should be managed.

North Island *P. australis* populations tended to show a higher degree of population differentiation, resulting in lower genetic and demographic connectivity. This is concerning for fisheries management as separate stocks may cover smaller geographic areas than currently accounted for by the quota management system. Lower differentiation among South Island populations suggests that separate stocks could cover larger areas. There was evidence for many populations of both study species that external sources of recruits may be intermittent and self-recruitment may play an important role in maintaining stocks. This reliance on self-recruitment poses a risk when stocks become depleted; it may be a long time before recruitment from external sources is able to rebuild biomass to sustainable levels (McClanahan *et al.* 2007; Worm *et al.* 2009), especially if recruits are not adapted to local conditions. This finding is concerning for recreational and customary fisheries for which very little is known about harvest rates and how biomass might be affected when self-recruitment is high. There is evidence that some shellfish stocks have become locally depleted around the Auckland and Coromandel areas and recreational quotas have been restricted as a result (New Zealand Ministry for Primary Industries 2013a; b). In general, the *P. australis* fishery is probably much more vulnerable to over-exploitation given the high densities and spatially discrete nature of their beds. Furthermore, estuary habitats are threatened by sedimentation, pollution and climate change (Cole *et al.* 2000; Thrush *et al.* 2003), which could also contribute to a decline in *P. australis* biomass. In comparison the

large spatial scale of *P. subtriangulata* habitat and higher levels of connectivity are likely to make this species more resilient to fishing pressure and changing environmental conditions.

The commercial *P. australis* fishery at Mair Bank is of particular concern. The nearby *P. subtriangulata* population at Ruakaka showed significant genetic differentiation from all other populations and relied heavily on self-recruitment. It is likely that a similar situation exists at Mair Bank, meaning that this fishery could be vulnerable to over-exploitation. Although harvest of this fishery has been sustainable in the past, the New Zealand Ministry for Primary Industries is unsure about whether the present level of harvest is sustainable in the future (New Zealand Ministry of Fisheries 2010). Anecdotal evidence suggests that *P. australis* biomass has recently declined at this location (R. Ford, MPI Wellington, personal communication). Restricted connectivity and the extent of self-recruitment are also of relevance to the *P. subtriangulata* commercial fishery located in the Kaipara Harbour entrance. As demonstrated by *P. australis* genetic population structure, populations within harbours are likely to experience lower connectivity to populations outside the harbour, resulting in genetic differentiation of harbour populations. This situation is illustrated by the nearby *P. subtriangulata* population from Kakamatua within the Manukau Harbour, which was significantly differentiated from some other populations. There is also the possibility that individuals within the Kaipara Harbour fishery are locally adapted to harbour conditions rather than the more typical surf beach environment of *P. subtriangulata*. The implication is that this fishery could be at risk from over-exploitation and stocks may take some time to recover if reliant on intermittent external sources of recruits that are suitably adapted to harbour conditions.

If commercial fishing of these two species (or other surf clam species) is expanded then source-sink dynamics, levels of self-recruitment and the appropriate scale for stock management need to be considered. In areas where larval dispersal is restricted and self-recruitment is high (e.g., between the North and South Islands, northern and eastern coasts of the North Island and Chatham Island) stocks should be managed with the assumption that connectivity to external sources of recruits is low. This finding can be extrapolated to other coastal fishery species where similar restrictions to larval dispersal are observed. However, the small scale CGP patterns that were observed in both species show that unexpected patterns of differentiation can arise and possibly change over time, and that it is

difficult to make predictions about the exact geographic locations where this differentiation may occur. Ideally, where new commercial fisheries are being considered, it is recommended that small scale investigations of spatial and temporal genetic population structure and connectivity be conducted to determine the appropriate scale on which to manage the fishery. Having knowledge of genetic stock structure (and the scale on which it could potentially change over time) prior to undertaking fishing activity means that management decisions can be underpinned by scientific information, and fisheries can be managed sustainably from the outset. A long history of exploitation in many New Zealand fisheries means that in practice this recommendation is difficult to implement, but small scale investigations of genetic connectivity have been attempted for fisheries where overexploitation is of concern and it is suspected that dispersal is limited or localised (e.g., South Island black-foot abalone *Haliotis iris*, McCowan 2013; Coromandel scallops *Pecten novaezealandiae*, C. Silva in preparation). Furthermore, when determining management units the degree of temporal variation in genetic stock structure and connectivity should also be quantified to ensure that these units persist over time. Such investigations of temporal genetic stock structure are rare and have provided contradictory results. Temporal stability of genetic stock structure has been suggested for coastal (*Pagrus auratus*; Bernal-Ramírez & Adcock 2003) and pelagic fish species (Palstra & Ruzzante 2010; Therkildsen *et al.* 2010). Studies that specifically investigate temporal genetic structure in coastal invertebrate species are few and tend to show variability over time (Barshis *et al.* 2011; Calderón *et al.* 2011), although temporal stability in genetic population structure was observed in Arctic surf clams (Cassista & Hart 2007).

6.4.2 Local adaptation and the evolution of coastal marine species

In addition to the implications for fisheries management discussed above, the discovery that local adaptation could be influencing genetic population structure has implications for our understanding of how coastal marine populations evolve and speciate. It is significant that the only published study of seascape genetics for a New Zealand marine species has implicated adaptation to SST as playing a role in the genetic structuring of populations (Wei *et al.* 2013b) and it is very likely that future studies will find a similar pattern in other species. In addition to evidence that SST is structuring *P. australis* populations, results from this study suggest that adaptation to the physical properties of

estuary and beach habitat are influencing population structure in both species. Selection and adaptation are widely recognised as playing important roles in structuring populations of intertidal organisms (see Schmidt *et al.* 2008 and references therein).

The suggestion of local selection in *P. australis* is not surprising given the range of differences in the physical parameters of the estuaries/harbours sampled for this study. Sampling sites included large enclosed harbours (Huia), coastal lagoons (Okuru), open embayments (Hakahaka Bay), deep narrow fiords (Doubtful Sound) and surf beaches (Bluff). If local adaptation is widespread among coastal marine organisms (a likely scenario for estuarine species) it has significant implications for connectivity among populations. Populations may receive large numbers of migrants but if a settlement site has a high population density then competition and post-settlement mortality are likely to be high. If new immigrants are not well adapted to environmental conditions at the settlement site they are likely to experience higher mortality than self-recruited individuals. If immigrants do not survive to reproductive age then migration is essentially ineffective and the genetic signal of connectivity will not exist. However, it is important to remember that without knowledge of the mechanistic links between genetic variation and adaptation to local environmental conditions, it is difficult to speculate further about the relative importance of local adaptation versus other processes that were identified as being potentially important in the genetic structuring of *P. subtriangulata* and *P. australis* populations (i.e., larval duration, distance between habitats, current patterns).

The prospect of locally adapted populations also has consequences for conservation of marine species. The preservation of locally adapted traits is often a high priority in conservation strategies (Conover *et al.* 2006; Hauser & Carvalho 2008; Nielsen *et al.* 2009). This study suggests that local adaptation may be more widespread and significant for coastal marine species than previously imagined and locally adapted populations could easily be lost through over-exploitation. This is particularly pertinent given that climate change and associated changes in sea level are expected to have a significant impact on the physical dynamics of estuarine systems (Thrush *et al.* 2003). Furthermore, changes in SST gradients are expected to occur as a result of climate change, which is likely to have an impact on populations adapted to particular SST regimes. As a consequence, locally adapted populations may face extinction due to changes in their environment. There is potential for

re-colonisation by individuals that are better adapted to the altered estuary and SST conditions, but this likely to result in an overall decrease in the genetic diversity of a species.

The presence of locally adapted populations provides insight to the larger scale processes that might be contributing to the evolution of marine species in New Zealand. As mentioned previously, species with a New Zealand-wide distribution often exhibit genetic differentiation between northern and southern populations, and there is evidence that this could be related to SST adaptation. There are species that are generally more common in the north (red moki, snapper, parore) or south (bull kelp, black cod, blue cod, black-foot abalone, dredge oysters) of New Zealand and it is often assumed that their distributions are restricted by adaptation to warm or cold temperature regimes (Francis & Nelson 2003). It is conceivable that adaptation to the warm or cold water masses that characterise different parts of the New Zealand region could represent a first step towards speciation.

Local adaptation to SST is a plausible scenario leading to speciation of the two tuatua species, *P. subtriangulata* and *P. donacina*. While both species supposedly have a New Zealand-wide distribution, *P. subtriangulata* is more common in the north and *P. donacina* more common in the south. In the past it was suggested that these species represent a geographic cline driven by water temperature until it was established that they are actually separate species (Richardson *et al.* 1982). It is plausible that speciation may have occurred from a common tuatua ancestor when warm and cold adapted populations were isolated, possibly during the last glacial maximum, which has been proposed as a mechanism of speciation in other New Zealand marine and terrestrial species (Wallis & Trewick 2009). The fact that cross-amplification of microsatellite markers was not possible between these species suggests that genetic divergence is reasonably high (and therefore presumably relatively ancient) and morphological similarity could be maintained due to their similar habitat. The implication for *P. australis* and the many other coastal marine species that are characterised by north-south divergence is that local adaptation to SST could potentially drive speciation in the future if the average current patterns of the present day were to change (e.g., in response to climate change).

A further method of speciation is illustrated by the Chatham Island *P. subtriangulata* population, which showed strong genetic differentiation and was rarely connected to

mainland populations by gene flow. The results of the present study are consistent with a previous suggestion that this population represents a separate species of tuatua (Beu & de Rooij-Schuiling 1982). Furthermore the Chatham Islands are known for high levels of endemism in both their marine and terrestrial flora and fauna (Heenan *et al.* 2010; McDowall 2004; Nelson & Broom 2008). The genetic differentiation of Chatham Island *P. subtriangulata* illustrates how distance can form a barrier to dispersal and provides an example of how allopatry can drive speciation in the New Zealand marine environment.

6.5 Contribution to the field and future research

The findings of this research are significant in two primary ways. Firstly, this study has provided valuable and previously unknown information about genetic stock structure and connectivity in *P. subtriangulata* and *P. australis* that can be used for management of these species, and could potentially be applicable to fisheries of other surf clam species with similar life history and ecological characteristics. The findings and recommendations of this study may also be applicable to fisheries of estuarine species. Secondly, it has contributed to a growing body of knowledge about population processes and evolution of coastal marine species in New Zealand.

Few previous genetic studies have focussed on bivalves or species that inhabit soft-shore environments; the information provided by the present study helps to fill a gap in our knowledge of population genetics for these types of species. This study has shown that, despite the challenges of working with multiple species, comparative population genetics studies are highly valuable. By using standardised methods to compare two closely related species it is possible to reveal a more detailed understanding of how similar or differing patterns of genetic population structure can arise. The approach taken in this study was analytically comprehensive compared to the types of population genetic studies previously undertaken for New Zealand coastal marine species, primarily as a consequence of the recent development of the methods and computing capabilities required to handle large multilocus microsatellite data sets. Previous studies have presented standard population differentiation statistics, although recently Bayesian methods have become more commonplace. Very few studies have attempted to quantify rates of connectivity or to incorporate environmental variables into their analyses. In this study I have followed the recommendations of several authors who advocate using an integrative approach,

combining traditional population genetics methodology with more advanced (but complementary) computational methods that incorporate ecologically relevant information (Hedgecock *et al.* 2007; Pearse & Crandall 2004; Riginos & Liggins 2013; Selkoe *et al.* 2008). This approach increases our power to tease apart the complexities of how and why populations of marine organisms are structured over time and space, and to provide biologically meaningful information to assist with their management and conservation. The present research is significant as it represents one of the first studies of marine population genetics in New Zealand to implement such an approach.

Additionally, this study has highlighted several exciting directions for future research. Firstly, this study showed that most of the sampled populations are genetically connected for both species, but that many might not be demographically connected at the short time scales relevant to fishery management. To improve fishery management strategies a better understanding of demographic connectivity patterns is required. This could be achieved by incorporating data from ecological studies, modelling of propagule dispersal/oceanographic processes or by temporal genetic sampling of populations. Temporal sampling would confirm non-equilibrium population dynamics in the two study species and would further demonstrate how genetic population structure changes over time in response to variable levels of connectivity. Comparison of *P. subtriangulata* population structure reported in this study to that reported by Smith *et al.* (1989) has already suggested that genetic population structure could be variable over time in this species. Unfortunately samples from the Mair Bank and Kaipara Harbour commercial fisheries were unable to be obtained in the present study. Inclusion of these populations would provide useful information about the impact of commercial fishing on population dynamics and the future sustainability of these stocks.

While the present study has provided an overview of New Zealand-wide genetic population structure in *P. subtriangulata* and *P. australis*, small scale sampling would also provide new insight to patterns of demographic connectivity. While large scale investigations of genetic population structure are highly valuable, a draw-back of this approach (as highlighted by Gardner *et al.* 2010) is that they often do not have the resolution to pinpoint the exact location of barriers to dispersal. In this study it was only possible to report a broad geographic range for the location of potential barriers. This study has suggested that it may be necessary to manage fisheries on a smaller scale than present,

so it is crucial to accurately pinpoint the location of stock boundaries, something that can be achieved by conducting population genetics studies on a small spatial scale. For this type of small scale investigation higher resolution SNP markers would be highly beneficial, and existing genomic sequence data provides a good basis for development of these markers. A further advantage of SNP markers is that they are unaffected by null alleles and it is possible to sample genetic variation from a higher proportion of the genome compared to microsatellite markers (Morin *et al.* 2004). Consequently, portions of the genome affected by selection are more likely to be sampled, which could shed more light on the role that local adaptation plays in the genetic population structuring of marine species.

Perhaps the most interesting line of future research is to investigate the extent to which local adaptation influences genetic population structure and to confirm links to specific loci that might be under selection. A number of loci have already been implicated in adaptation to environmental variation in other marine species. These loci could easily be identified in *P. subtriangulata* and *P. australis* from existing genomic sequence data. Variation in allele frequency or expression level across an environmental gradient could be investigated. Evidence of selection was observed in one *P. australis* locus, suggesting that this locus may be linked to a gene undergoing selection. Further investigation of the genetic population structure revealed by this locus and comparison to results from 'neutral' loci would be a good starting point for investigating the role of natural selection in shaping *P. australis* population structure.

Finally, further investigation of genetic population structure in *P. subtriangulata* and *P. australis* would benefit from establishing the phylogenetic relationships among the species within the *Paphies* genus. This would help to establish whether Chatham Island *P. subtriangulata* is indeed a separate species. Knowledge of the phylogenetic relationships within this genus would provide a link between the microevolutionary local adaptation processes observed in this study and the macroevolutionary processes that lead to speciation within the marine environment.

References

Abdelkrim, J., Robertson, B.C., Stanton, J.A.L. & Gemmell, N.J. (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, **46**, 185–191.

Addison, J.A. & Hart, M.W. (2005) Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biology Letters*, **1**, 450–453.

An, H.S., Park, K.J., Cho, K.C., Han, H.S. & Myeong, J.I. (2012) Genetic structure of Korean populations of the clam *Ruditapes philippinarum* inferred from microsatellite marker analysis. *Biochemical Systematics and Ecology*, **44**, 186–195.

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

Apte, S. & Gardner, J.P.A. (2001) Absence of population genetic differentiation in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791) as assessed by allozyme variation. *Journal of Experimental Marine Biology and Ecology*, **258**, 173–194.

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

Apte, S., Star, B. & Gardner, J.P.A. (2003) A comparison of genetic diversity between cultured and wild populations, and a test for genetic introgression in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791). *Aquaculture*, **219**, 193–220.

Arias-Pérez, A., Fernández-Tajes, J., Gaspar, M.B. & Méndez, J. (2012) Isolation of microsatellite markers and analysis of genetic diversity among east Atlantic populations of the sword razor shell *Ensis siliqua*: a tool for population management. *Biochemical Genetics*, **50**, 397–415.

Arnaud-Haond, S., Bonhomme, F. & Blanc, F. (2003) Large discrepancies in differentiation of allozymes, nuclear and mitochondrial DNA loci in recently founded Pacific populations of the pearl oyster *Pinctada margaritifera*. *Journal of Evolutionary Biology*, **16**, 388–398.

Ayers, K.L. & Waters, J.M. (2005) Marine biogeographic disjunction in central New Zealand. *Marine Biology*, **147**, 1045–1052.

Baker, P., Austin, J.D., Bowen, B.W. & Baker, S.M. (2008) Range-wide population structure and history of the northern quahog (*Merceneria merceneria*) inferred from mitochondrial DNA sequence data. *ICES Journal of Marine Science*, **65**, 155–163.

Banks, S.C., Piggott, L.M.P., Williamson, J.E., Bové, U., Holbrook, N.J. & Beheregaray, L.B. (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, **88**, 3055–3064.

Barshis, D.J., Sotka, E.E., Kelly, R.P., Sivasundar, A., Menge, B.A., Barth, J.A. & Palumbi, S.R. (2011) Coastal upwelling is linked to temporal genetic variability in the acorn barnacle *Balanus glandula*. *Marine Ecology Progress Series*, **439**, 139–150.

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

Beerli, P. (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Benzie, J. & Smith-Keune, C. (2006) Microsatellite variation in Australian and Indonesian pearl oyster *Pinctada maxima* populations. *Marine Ecology Progress Series*, **314**, 197–211.

Bernal-Ramírez, J. & Adcock, G. (2003) Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents. *Marine Biology*, **142**, 567–574.

Beu, A.G. & de Rooij-Schuiling, L.A. (1982) Subgeneric classification of New Zealand and Australian species of *Paphies* Lesson (Bivalvia, Mesodesmatidae), and names for the two species of tuatua in New Zealand. *New Zealand Journal of Zoology*, **9**, 211–229.

Bilton, D. (2002) Dispersal, genetic differentiation and speciation in estuarine organisms. *Estuarine, Coastal and Shelf Science*, **55**, 937–952.

Bird, C.E., Karl, S.A., Smouse, P.E. & Toonen, R.J. (2011) Detecting and measuring genetic differentiation. *Crust Issues*, **19**, 39–63.

Bohonak, A.J. (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, **74**, 21–45.

Bradbury, I.R., Campana, S.E. & Bentzen, P. (2008) Low genetic connectivity in an estuarine fish with pelagic larvae. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 147–158.

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

Broekhuizen, N., Lundquist, C.J., Hadfield, M.G. & Brown, S.N. (2011) Dispersal of oyster (*Ostrea chilensis*) larvae in Tasman Bay inferred using a verified particle tracking model that incorporates larval behavior. *Journal of Shellfish Research*, **30**, 643–658.

Broquet, T. & Petit, E.J. (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology Evolution and Systematics*, **40**, 193–216.

Buchanan, J. & Zuccarello, G.C. (2012) Decoupling of short- and long-distance dispersal pathways in the endemic New Zealand seaweed *Carpophyllum maschalocarpum* (Phaeophyceae, Fucales). *Journal of Phycology*, **48**, 518–529.

Calderón, I., Pita, L., Brusciotti, S., Palacín, C. & Turon, X. (2011) Time and space: genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Marine Biology*, **159**, 187–197.

Carlton, J.T. (2003) Community assembly and historical biogeography in the North Atlantic Ocean: the potential role of human-mediated dispersal vectors. *Hydrobiologia*, **503**, 1–8.

Carroll, J. & Wells, R. (1995) Strategies of anaerobiosis in New Zealand infaunal bivalves: adaptations to environmental and functional hypoxia. *New Zealand Journal of Marine and Freshwater Research*, **29**, 137–146.

Case, R.A.J., Hutchinson, W.F., Hauser, L., Van Oosterhout, C. & Carvalho, G.R. (2005) Macro- and micro-geographic variation in pantophysin (*PanI*) allele frequencies in NE Atlantic cod *Gadus morhua*. *Marine Ecology Progress Series*, **301**, 267–278.

Cassista, M.C. & Hart, M.W. (2007) Spatial and temporal genetic homogeneity in the Arctic surfclam (*Mactromeris polynyma*). *Marine Biology*, **152**, 569–579.

Caudill, C.C. & Bucklin, A. (2004) Molecular phylogeography and evolutionary history of the estuarine copepod, *Acartia tonsa*, on the Northwest Atlantic coast. *Hydrobiologia*, **511**, 91–102.

Chapuis, M. & Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Chiswell, S.M. (2003) Circulation within the Wairarapa Eddy, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **37**, 691–704.

Chiswell, S.M. & Rickard, G.J. (2011) Larval connectivity of harbours via ocean currents: a New Zealand study. *Continental Shelf Research*, **31**, 1057–1074.

Chiswell, S. & Roemmich, D. (1998) The East Cape Current and two eddies: a mechanism for larval retention? *New Zealand Journal of Marine and Freshwater Research*, **32**, 385–397.

Clarke, K.R. & Gorley, G.N. (2006) *PRIMER v6: User Manual/Tutorial*. PRIMER-E, Plymouth.

Cole, R.G., Hull, P.J. & Healy, T.R. (2000) Assemblage structure, spatial patterns, recruitment, and post-settlement mortality of subtidal bivalve molluscs in a large harbour in north-eastern New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **34**, 317–329.

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

Coscia, I., Robins, P.E., Porter, J.S., Malham, S.K. & Ironside, J.E. (2012) Modelled larval dispersal and measured gene flow: seascape genetics of the common cockle *Cerastoderma edule* in the southern Irish Sea. *Conservation Genetics*, **14**, 451–466.

Cowen, R.K., Gawarkiewicz, G., Pineda, J., Thorrold, S.R. & Werner, F.E. (2007) Population connectivity in marine systems: an overview. *Oceanography*, **20**, 14–21.

Cowen, R.K. & Sponaugle, S. (2009) Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, **1**, 443–466.

Crandall, E.D., Treml, E.A. & Barber, P.H. (2012) Coalescent and biophysical models of stepping-stone gene flow in neritid snails. *Molecular Ecology*, **21**, 5579–98.

Crawford, D.L. & Powers, D.A. (1989) Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 9365–9369.

Cruz, F., Pérez, M. & Presa, P. (2005) Distribution and abundance of microsatellites in the genome of bivalves. *Gene*, **346**, 241–247.

Dempster, A., Laird, N. & Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, **39**, 1–38.

DeWoody, J., Nason, J.D. & Hipkins, V.D. (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, **6**, 951–957.

Dutton, J.M. & Hofmann, G.E. (2009) Biogeographic variation in *Mytilus galloprovincialis* heat shock gene expression across the eastern Pacific range. *Journal of Experimental Marine Biology and Ecology*, **376**, 37–42.

Earl, D.A. & vonHoldt, B.M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Einfeldt, A.L. & Addison, J.A. (2013) Hydrology influences population genetic structure and connectivity of the intertidal amphipod *Corophium volutator* in the northwest Atlantic. *Marine Biology*, **160**, 1015–1027.

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

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

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

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

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

Falush, D., Stephens, M. & Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Faubet, P., Waples, R.S. & Gaggiotti, O.E. (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Faurby, S. & Barber, P.H. (2012) Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology*, **21**, 3419–3432.

Francis, M.P. & Nelson, W. (2003) Biogeography. *The Living Reef*. (eds N.L. Andrew & M.P. Francis), pp. 32–37. Craig Potton Publishing, Nelson.

Frankham, R. (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.

Fraser, D.J., Hansen, M.M., Ostergaard, S., Tessier, N., Legault, M. & Bernatchez, L. (2007) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology*, **16**, 3866–3889.

Galindo, H.M., Olson, D.B. & Palumbi, S.R. (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology*, **16**, 1622–1626.

Galindo, H.M., Pfeiffer-Herbert, A.S., McManus, M.A., Chao, Y., Chai, F. & Palumbi, S.R. (2010) Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology*, **19**, 3692–3707.

Gao, X. & Starmer, J.D. (2008) AWclust: point-and-click software for non-parametric population structure analysis. *BMC Bioinformatics*, **9**, 77.

Gardner, J.P.A., Bell, J.J., Constable, H.B., Hannan, D.A., Ritchie, P.A. & Zuccarello, G.C. (2010) Multi-species coastal marine connectivity: a literature review with recommendations for further research. *New Zealand Aquatic Environment and Biodiversity Report No. 58*, Pp. 47, Wellington.

Goldstien, S., Gemmell, N. & Schiel, D. (2009) Colonisation and connectivity by intertidal limpets among New Zealand, Chatham and Sub-Antarctic Islands. I. Genetic connections. *Marine Ecology Progress Series*, **388**, 111–119.

Goldstien, S.J., Schiel, D.R. & Gemmell, N.J. (2006) Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology*, **15**, 3259–3268.

González-Wangüemert, M., Cánovas, F., Marcos, C. & Pérez-Ruzafa, Á. (2009) Phosphoglucose isomerase variability of *Cerastoderma glaucum* as a model for testing the influence of environmental conditions and dispersal patterns through quantitative ecology approaches. *Biochemical Systematics and Ecology*, **37**, 325–333.

Gordon, D.P., Beaumont, J., MacDiarmid, A., Robertson, D.A. & Ahyong, S.T. (2010) Marine biodiversity of Aotearoa New Zealand. *PloS One*, **5**, e10905.

Goudet, J. (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Goudet, J., Raymond, M., DeMeeüs, T. & Rousset, F. (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Gow, J.L., Noble, L.R., Rollinson, D. & Jones, C.S. (2005) A high incidence of clustered microsatellite mutations revealed by parent-offspring analysis in the African freshwater snail, *Bulinus forskalii* (Gastropoda, Pulmonata). *Genetica*, **124**, 77–83.

Grant, C.M. & Creese, R.G. (1995) The reproductive cycle of the tuatua, *Paphies subtriangulata* (Wood, 1828), in New Zealand. *Journal of Shellfish Research*, **14**, 287–292.

Grant, C.M., Hooker, S.H., Babcock, R.C. & Creese, R.G. (1998) Synchronous spawning and reproductive incompatibility of two bivalve species: *Paphies subtriangulata* and *Paphies australis*. *Veliger*, **41**, 148–156.

Hadfield, M.G., Rickard, G.J. & Uddstrom, M.J. (2007) A hydrodynamic model of Chatham Rise, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **41**, 239–264.

Harris, L.N., Howland, K.L., Kowalchuk, M.W., Bajno, R., Lindsay, M.M. & Taylor, E.B. (2012) Microsatellite and mtDNA analysis of lake trout, *Salvelinus namaycush*, from Great Bear Lake, Northwest Territories: impacts of historical and contemporary evolutionary forces on Arctic ecosystems. *Ecology and Evolution*, **3**, 145–161.

Hastings, A. (1993) Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecology*, **74**, 1362–1372.

Hauser, L. & Carvalho, G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.

Heath, R. (1985) A review of the physical oceanography of the seas around New Zealand—1982. *New Zealand Journal of Marine and Freshwater Research*, **19**, 79–124.

Hedgecock, D. (1994) Does variance in reproductive success limit effective population size of marine organisms? *Genetics and Evolution of Aquatic Organisms*. (ed A. Beaumont), pp. 122–134. Chapman & Hall, London.

Hedgecock, D., Barber, P.H. & Edmands, S. (2007) Genetic approaches to measuring connectivity. *Oceanography*, **20**, 70–79.

Hedgecock, D., Li, G., Hubert, S., Bucklin, K. & Ribes, V. (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research*, **23**, 379–385.

Hedgecock, D. & Pudovkin, A.I. (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, **87**, 971–1002.

Hedrick, P.W. (2005) Large variance in reproducive success and the N_e/N ratio. *Evolution*, **59**, 1596–1599.

Heenan, P., Mitchell, A., DeLange, P., Keeling, J. & Paterson, A. (2010) Late-Cenozoic origin and diversification of Chatham Islands endemic plant species revealed by analyses of DNA sequence data. *New Zealand Journal of Botany*, **48**, 83–136.

Hellberg, M.E. (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics*, **40**, 291–310.

Hellberg, M.E., Burton, R.S., Neigel, J.E. & Palumbi, S.R. (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, **70**, 273–290.

Hemmer-Hansen, J., Nielsen, E.E., Frydenberg, J. & Loeschcke, V. (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity*, **99**, 592–600.

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

Hilbish, T. & Koehn, R. (1987) The adaptive importance of genetic variation. *American Scientist*, **75**, 134–141.

Hoffmann, A.A. & Willi, Y. (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Holleley, C.E. & Geerts, P.G. (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques*, **46**, 511–517.

Hooker, S.H. (1997) Larval and postlarval development of the New Zealand pipi, *Paphies australis* (Bivalvia : Mesodesmatidae). *Bulletin of Marine Science*, **61**, 225–240.

Hooker, S.H. (1995) Preliminary evidence for post-settlement movement of juvenile and adult pipi, *Paphies australis* (Gmelin, 1790) (Bivalvia: Mesodesmatidae). *Marine and Freshwater Behaviour and Physiology*, **27**, 37–47.

Hooker, S.H. & Creese, R.G. (1995a) The reproductive biology of pipi, *Paphies australis* (Gmelin, 1790) (Bivalvia, Mesodesmatidae). 1. Temporal patterns of the reproductive cycle. *Journal of Shellfish Research*, **14**, 7–15.

Hooker, S.H. & Creese, R.G. (1995b) The reproductive biology of pipi, *Paphies australis* (Gmelin, 1790) (Bivalvia, Mesodesmatidae). 2. Spatial patterns of the reproductive cycle. *Journal of Shellfish Research*, **14**, 17–24.

Hou, L., Lu, H., Zou, X., Bi, X., Yan, D. & He, C. (2006) Genetic characterizations of *Mactra veneriformis* (Bivalve) along the Chinese coast using ISSR-PCR markers. *Aquaculture*, **261**, 865–871.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hull, P.J., Cole, R.G., Creese, R.G. & Healy, T.R. (1998) An experimental investigation of the furrowing behaviour of *Paphies australis* (Bivalvia : Mesodesmatidae). *Marine and Freshwater Behaviour and Physiology*, **31**, 167–183.

Hume, T., Snelder, T., Weatherhead, M. & Liefting, R. (2007) A controlling factor approach to estuary classification. *Ocean and Coastal Management*, **50**, 905–929.

Ilves, K.L., Huang, W., Wares, J.P. & Hickerson, M.J. (2010) Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation. *Molecular Ecology*, **19**, 4505–4519.

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

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

Jolly, M., Guyard, P., Ellien, C. & Gentil, F. (2009) Population genetics and hydrodynamic modeling of larval dispersal dissociate contemporary patterns of connectivity from historical expansion into European shelf seas in the polychaete *Pectinaria koreni*. *Limnology and Oceanography*, **54**, 2089–2106.

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

Jones, G.P., Milicich, M.J., Emslie, M.J. & Lunow, C. (1999) Self-recruitment in a coral reef fish population. *Nature*, **402**, 802–804.

Jørgensen, H.B.H., Hansen, M.M., Bekkevold, D., Ruzzante, D.E. & Loeschcke, V. (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology*, **14**, 3219–3234.

Kang, J.H., Kim, B.H., Park, J.Y., Lee, J.M., Jeong, J.E., Lee, J.S., Ko, H.S. & Lee, Y.S. (2012) Novel microsatellite markers of *Meretrix petechialis* and cross-species amplification in related taxa (Bivalvia: Veneroida). *International Journal of Molecular Sciences*, **13**, 15942–15954.

Kelly, D.W., MacIsaac, H.J. & Heath, D.D. (2006) Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. *Evolution*, **60**, 257–267.

Kelly, R.P. & Palumbi, S.R. (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PloS One*, **5**, e8594.

Kenchington, E.L., Patwary, M.U., Zouros, E. & Bird, C.J. (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology*, **15**, 1781–1796.

Kimura, M. & Weiss, G. (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 561–576.

Kinlan, B.P. & Gaines, S.D. (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology*, **84**, 2007–2020.

Knox, M.A., Hogg, I.D. & Pilditch, C.A. (2011) The role of vicariance and dispersal on New Zealand's estuarine biodiversity: the case of *Paracorophium* (Crustacea: Amphipoda). *Biological Journal of the Linnean Society*, **103**, 863–874.

Knutsen, H., Olsen, E.M., Jorde, P.E., Espeland, S.H., André, C. & Stenseth, N.C. (2011) Are low but statistically significant levels of genetic differentiation in marine fishes "biologically meaningful"? A case study of coastal Atlantic cod. *Molecular Ecology*, **20**, 768–783.

Kong, L., Li, Q. & Qiu, Z. (2007) Genetic and morphological differentiation in the clam *Coelomactra antiquata* (Bivalvia: Veneroida) along the coast of China. *Journal of Experimental Marine Biology and Ecology*, **343**, 110–117.

Laing, A. & Chiswell, S.M. (2003) The ocean medium. *The Living Reef*. (eds N.L. Andrew & M.P. Francis), pp. 24–31. Craig Potton Publishing, Nelson.

Laudien, J., Flint, N.S., van der Bank, F.H. & Brey, T. (2003) Genetic and morphological variation in four populations of the surf clam *Donax serra* (Röding) from southern African sandy beaches. *Biochemical Systematics and Ecology*, **31**, 751–772.

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

Legendre, P. & Fortin, M.J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Lessios, H.A. & Robertson, D.R. (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society B - Biological Sciences.*, **273**, 2201–2208.

Levin, L.A. (2006) Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, **46**, 282–297.

Li, J., Foighil, D.O. & Park, J.K. (2013) Triton's trident: cryptic Neogene divergences in a marine clam (*Lasaea australis*) correspond to Australia's three temperate biogeographic provinces. *Molecular Ecology*, **22**, 1933–1946.

Lidgard, C.W. (2001) *Evaluating the population genetic structure of the New Zealand cockle* Austrovenus stutchburyi *using allozyme electrophoresis.* MSc Thesis, University of Waikato, Hamilton, New Zealand.

Lind, C.E., Evans, B.S., Taylor, J.J.U. & Jerry, D.R. (2007) Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits. *Molecular Ecology*, **16**, 5193–5203.

Lowe, W.H. & Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Marins, F.O., Novaes, R.L.M., Rocha, R.M. & Junqueira, A.O.R. (2010) Non indigenous ascidians in port and natural environments in a tropical Brazilian bay. *Zoologia*, **27**, 213–221.

Marsden, I.D. (2000) Variability in low tide populations of tuatua, *Paphies donacina*, in Pegasus Bay, Canterbury, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **34**, 359–370.

Martínez, L., Arias, A., Méndez, J., Insua, A. & Freire, R. (2009) Development of twelve polymorphic microsatellite markers in the edible cockle *Cerastoderma edule* (Bivalvia: Cardiidae). *Conservation Genetics Resources*, **1**, 107–109.

McCairns, R.J.S. & Bernatchez, L. (2008) Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment. *Molecular Ecology*, **17**, 3901–3916.

McClanahan, T.R., Graham, N.A.J., Calnan, J.M. & MacNeil, M.A. (2007) Toward pristine biomass: reef fish recovery in coral reef marine protected areas in Kenya. *Ecological Applications*, **17**, 1055–1067.

McCowan, T.A. (2013) *Genetic approaches to management of New Zealand's blackfoot pāua,* Haliotis iris. PhD Thesis, University of Otago, Dunedin, New Zealand.

McCraney, W.T., Goldsmith, G., Jacobs, D.K. & Kinziger, A.P. (2010) Rampant drift in artificially fragmented populations of the endangered tidewater goby (*Eucyclogobius newberryi*). *Molecular Ecology*, **19**, 3315–3327.

McDowall, R.M. (2004) The Chatham Islands endemic galaxiid: A Neochanna mudfish (Teleostei: Galaxiidae). *Journal of the Royal Society of New Zealand*, **34**, 315–331.

McLachlan, A., Jaramillo, E., Defeo, O., Dugan, J.E., DeRuyck, A. & Coetzee, P. (1995) Adaptations of bivalves to different beach types. *Journal of Experimental Marine Biology and Ecology*, **187**, 147–160.

McLeod, R.J. & Wing, T.R. (2008) Influence of an altered salinity regime on the population structure of two infaunal bivalve species. *Estuarine Coastal and Shelf Science*, **78**, 529–540.

Meglécz, 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. (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Meirmans, P.G. (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Meirmans, P.G. & Hedrick, P.W. (2010) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Meirmans, P.G. & van Tienderen, P.H. (2004) Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Melo, M.A.D., Silva, A.R.B., Varela, E.S., Sampaio, I. & Tagliaro, C.H. (2012) Development and characterization of ten microsatellite markers for population studies of the native Brazilian oyster *Crassostrea gasar*. *Conservation Genetics Resources*, **4**, 583–586.

Miller, K.J. (1997) Genetic structure of black coral populations in New Zealand's fiords. *Marine Ecology Progress Series*, **161**, 123–132.

Miller, K.J., Mundy, C.N. & Chadderton, L.W. (2004) Ecological and genetic evidence of the vulnerability of shallow-water populations of the stylasterid hydrocoral *Errina novaezelandiae* in New Zealand's fiords. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **14**, 75–94.

Mladenov, P., Allibon, R. & Wallis, G. (1997) Genetic differentiation in the New Zealand sea urchin *Evechinus chloroticus* (Echinodermata: Echinoidea). *New Zealand Journal of Marine and Freshwater Research*, **31**, 261–269.

Moller-Hansen, M. & Hemmer-Hansen, J. (2007) Landscape genetics goes to sea. *Journal of Biology*, **6**, 6.

Morin, P.A., Luikart, G., Wayne, R.K. & the SNP workshop group. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Morton, J.E. & Miller, M.C. (1968) *The New Zealand Sea Shore*. Collins, Auckland.

Murray-Jones, S.E. & Ayre, D.J. (1997) High levels of gene flow in the surf bivalve *Donax deltoides* (Bivalvia: Donacidae) on the east coast of Australia. *Marine Biology*, **128**, 83–89.

Neigel, J.E. (2002) Is F_{ST} obsolete? *Molecular Ecology*, **3**, 167–173.

Nelson, W. & Broom, J. (2008) New Zealand Gigartinaceae (Rhodophyta): Resurrecting *Gigartina grandifida* endemic to the Chatham Islands. *New Zealand Journal of Botany*, **45**, 177–187.

New Zealand Ministry for the Environment. (2005) *The New Zealand Marine Environment Classification*. Wellington, Pp.80.

New Zealand Ministry of Fisheries. (2010) *Pipi (PPI 1A) Mair Bank (Whangarei Harbour) plenary document*. Wellington, Pp. 7.

New Zealand Ministry for Primary Industries. (2013a) *Pipi (PPI) plenary document*. Wellington, Pp. 11.

New Zealand Ministry for Primary Industries. (2013b) *Tuatua (TUA) plenary document*. Wellington, Pp. 5.

Ni, L., Li, Q. & Kong, L. (2011) Microsatellites reveal fine-scale genetic structure of the Chinese surf clam *Mactra chinensis* (Mollusca, Bivalvia, Mactridae) in Northern China. *Marine Ecology*, **32**, 488–497.

Ni, G., Li, Q., Kong, L. & Zheng, X. (2012) Phylogeography of bivalve *Cyclina sinensis*: testing the historical glaciations and Changjiang River outflow hypotheses in northwestern Pacific. *PloS one*, **7**, e49487.

Nielsen, E.E., Hemmer-Hansen, J., Larsen, P.F. & Bekkevold, D. (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

Ovenden, J.R. (2013) Crinkles in connectivity: combining genetics and other types of biological data to estimate movement and interbreeding between populations. *Marine and Freshwater Research*, **64**, 201–207.

Ovenden, J., Brasher, D. & White, R. (1992) Mitochondrial DNA analyses of the red rock lobster *Jasus edwardsii* supports an apparent absence of population subdivision throughout Australasia. *Marine Biology*, **326**, 319–326.

Paetkau, D., Slade, R., Burden, M. & Estoup, A. (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.

Paetkau, D., Waits, L., Clarkson, P., Craighead, L. & Strobeck, C. (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, **147**, 1943–1957.

Palstra, F.P. & Ruzzante, D.E. (2010) A temporal perspective on population structure and gene flow in Atlantic salmon (*Salmo salar*) in Newfoundland, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 225–242.

Palstra, F.P. & Ruzzante, D.E. (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, **17**, 3428–3447.

Palumbi, S.R. (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, S146–S158.

Pampoulie, C., Gysels, E.S., Maes, G.E., Hellemans, B., Leentjes, V., Jones, A.G. & Volckaert, F.A.M. (2004) Evidence for fine-scale genetic structure and estuarine colonisation in a potential high gene flow marine goby (*Pomatoschistus minutus*). *Heredity*, **92**, 434–45.

Panova, M. & Johannesson, K. (2004) Microscale variation in *Aat* (aspartate aminotransferase) is supported by activity differences between upper and lower shore allozymes of *Littorina saxatilis*. *Marine Biology*, **144**, 1157–1164.

Pearse, D.E. & Crandall, K.A. (2004) Beyond F_{ST}: Analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585–602.

Pelc, R.A., Warner, R.R. & Gaines, S.D. (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, **36**, 1881–1890.

Perrin, C., Wing, S.R. & Roy, M.S. (2004) Effects of hydrographic barriers on population genetic structure of the sea star *Coscinasterias muricata* (Echinodermata, Asteroidea) in the New Zealand fiords. *Molecular Ecology*, **13**, 2183–2195.

Pineda, J., Hare, J.A. & Sponaugle, S. (2007) Larval transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanography*, **20**, 22–39.

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

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

Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Powell, A.W.B. (1979) *New Zealand Mollusca: Marine, Land and Freshwater Shells.* Collins, Auckland.

Powers, D. & Place, A. (1978) Biochemical genetics of *Fundulus heteroclitus* (L.). I. Temporal and spatial variation in gene frequencies of *Ldh-B*, *Mdh-A*, *Gpi-B*, and *Pgm-A*. *Biochemical Genetics*, **16**, 593–607.

Powers, D.A. & Schulte, P.M. (1998) Evolutionary adaptations of gene structure and expression in natural populations in relation to a changing environment: a multidisciplinary approach to address the million-year saga of a small fish. *The Journal of Experimental Zoology*, **282**, 71–94.

Preda, C., Memedemin, D., Skolka, M. & Cogălniceanu, D. (2012) Early detection of potentially invasive invertebrate species in *Mytilus galloprovincialis* Lamarck, 1819 dominated communities in harbours. *Helgoland Marine Research*, **66**, 545–556.

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

Pulliam, H. (1988) Sources, sinks, and population regulation. *American Naturalist*, **132**, 652–661.

Puritz, J.B. & Toonen, R.J. (2011) Coastal pollution limits pelagic larval dispersal. *Nature Communications*, **2**, 226.

Rannala, B. & Mountain, J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9197–9201.

Redfearn, P. (1987) Larval shell development of the northern tuatua, *Paphies subtriangulata* (Bivalvia, Mesodesmatidae). *New Zealand Journal of Marine and Freshwater Research*, **21**, 65–70.

Richards, V.P., Thomas, J.D., Stanhope, M.J. & Shivji, M.S. (2007) Genetic connectivity in the Florida reef system: comparative phylogeography of commensal invertebrates with contrasting reproductive strategies. *Molecular Ecology*, **16**, 139–157.

Richardson, J.R., Aldridge, A.E. & Smith, P.J. (1982) Analyses of tuatua populations - *Paphies subtriangulata* and *Paphies donacina*. *New Zealand Journal of Zoology*, **9**, 231–237.

Riginos, C. & Liggins, L. (2013) Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass*, **7**, 197–216.

Rosenberg, N.A. (2003) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

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

Ross, P.M., Hogg, I.D., Pilditch, C.A., Lundquist, C.J. & Wilkins, R.J. (2011) 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. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

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

Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. (eds S. Krawetz & S. Misener), pp. 365–386. Humana Press, Totowa, NJ.

Schmidt, P.S., Serrão, E.A., Pearson, G.A., Riginos, C., Rawson, P.D., Hilbish, T.J., Brawley, S.H., Trussell, G.C., Carrington, E., Wethey, D.S., Grahame, J.W., Bonhomme, F. & Rand,

D.M. (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, **89**, S91–107.

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

Schwartz, M.K. & McKelvey, K.S. (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441–452.

Selkoe, K.A., Gaines, S.D., Caselle, J.E. & Warner, R.R. (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, **87**, 3082–3094.

Selkoe, K.A., Henzler, C.M. & Gaines, S.D. (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

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

Selkoe, K.A. & Toonen, R.J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Selkoe, K.A., Watson, J.R., White, C., Horin, T.B., Iacchei, M., Mitarai, S., Siegel, D.A., Gaines, S.D. & Toonen, R.J. (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.

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

Shanks, A.L., Grantham, B.A. & Carr, M.H. (2003) Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, **13**, S159–S169.

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

Siegel, D.A., Mitarai, S., Costello, C.J., Gaines, S.D., Kendall, B.E., Warner, R.R. & Winters, K.B. (2008) The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8974–8979.

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

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 nonequilibrium populations. *Evolution*, **47**, 264–279.

Smith, P.J., MacArthur, G.J. & Michael, K.P. (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. & Roy, M. (2002) Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation. *Evolution*, **56**, 1954–1967.

Stanton, B.R., Goring, D.G. & Bell, R.G. (2001) Observed and modelled tidal currents in the New Zealand region. *New Zealand Journal of Marine and Freshwater Research*, **35**, 397–415.

Stanton, B.R., Sutton, P.J.H. & Chiswell, S.M. (1997) The East Auckland Current, 1994–95. *New Zealand Journal of Marine and Freshwater Research*, **31**, 537–549.

Star, B., Apte, S. & Gardner, J. (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.

Stenseth, N.C., Jorde, P.E., Chan, K.S., Hansen, E., Knutsen, H., André, C., Skogen, M.D. & Lekve, K. (2006) Ecological and genetic impact of Atlantic cod larval drift in the Skagerrak. *Proceedings of the Royal Society B - Biological Sciences.*, **273**, 1085–1092.

Stephens, S.A., Broekhuizen, N., MacDiarmid, A.B., Lundquist, C.J., McLeod, L. & Haskew, R. (2006) Modelling transport of larval New Zealand abalone (*Haliotis iris*) along an open coast. *Marine and Freshwater Research*, **57**, 519–532.

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

St-Onge, P., Sévigny, J.M., Strasser, C. & Tremblay, R. (2013) Strong population differentiation of softshell clams (*Mya arenaria*) sampled across seven biogeographic marine ecoregions: possible selection and isolation by distance. *Marine Biology*, **160**, 1065–1081.

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R.J., Delmelle, E., Vierling, L. & Waits, L.P. (2007) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–142.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R. & Waits, L.P. (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.

Strasser, C.A. & Barber, P.H. (2008) Limited genetic variation and structure in softshell clams (*Mya arenaria*) across their native and introduced range. *Conservation Genetics*, **10**, 803–814.

Su, X.J., Pan, Y., Zuo, Z.R., Kong, L.F. & Li, Q. (2012) Isolation and characterization of microsatellite loci in the clam *Lutraria sieboldii* Reeve (Bivalvia: Veneroida). *Conservation Genetics Resources*, **5**, 223–225.

Sui, L., Zhang, F., Wang, X., Bossier, P., Sorgeloos, P. & Hänfling, B. (2009) Genetic diversity and population structure of the Chinese mitten crab *Eriocheir sinensis* in its native range. *Marine Biology*, **156**, 1573–1583.

Sutton, P. & Bowen, M. (2011) Currents off the west coast of Northland, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **45**, 609–624.

Swearer, S.E., Caselle, J.E., Lea, D.W. & Warner, R.R. (1999) Larval retention and recruitment in an island population of a coral-reef fish. *Nature*, **402**, 799–802.

Swearer, S.E., Shima, J.S., Hellberg, M.E., Thorrold, S.R., Jones, G.P., Robertson, D.R., Morgan, S.G., Selkoe, K.A., Ruiz, G.M. & Warner, R.R. (2002) Evidence of self-recruitment in demersal marine populations. *Bulletin of Marine Science*, **70**, 251–271.

Tarnowska, K., Chenuil, A., Nikula, R., Féral, J.P. & Wolowicz, M. (2010) Complex genetic population structure of the bivalve *Cerastoderma glaucum* in a highly fragmented lagoon habitat. *Marine Ecology Progress Series*, **406**, 173–184.

Therkildsen, N.O., Nielsen, E.E., Swain, D.P. & Pedersen, J.S. (2010) Large effective population size and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 1585–1595.

Thrush, S., Hewitt, J., Norkko, A., Nicholls, P., Funnell, G. & Ellis, J. (2003) Habitat change in estuaries: predicting broad-scale responses of intertidal macrofauna to sediment mud content. *Marine Ecology Progress Series*, **263**, 101–112.

Torkkola, J., Riginos, C. & Liggins, L. (2013) Regional patterns of mtDNA diversity in *Styela plicata*, an invasive ascidian, from Australian and New Zealand marinas. *Marine and Freshwater Research*, **64**, 139–145.

Trewick, S.A. & Wallis, G.P. (2001) Bridging the "Beech Gap": New Zealand invertebrate phylogeography implicates Pleistocence glaciation and Pliocence isolation. *Evolution*, **55**, 2170–2180.

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. *Molecular Ecology Notes*, **4**, 535–538.

van Oosterhout, C., Weetman, D. & Hutchinson, W.F. (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes*, **6**, 255–256.

Varela, M.A., Martínez-Lage, A. & González-Tizón, A.M. (2009) Temporal genetic variation of microsatellite markers in the razor clam *Ensis arcuatus* (Bivalvia: Pharidae). *Journal of the Marine Biological Association of the United Kingdom*, **89**, 1703–1707.

Veale, A. & Lavery, S. (2012) The population genetic structure of the waratah anemone (*Actinia tenebrosa*) around New Zealand. *New Zealand Journal of Marine and Freshwater Research*, **46**, 523–536.

Verhoeven, K., Simonsen, K. & McIntyre, L. (2005) Implementing false discovery rate control: increasing your power. *Oikos*, **108**, 643–647.

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

Waples, R.S. & Gaggiotti, O.E. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Waples, R.S. & Naish, K.A. (2009) Genetic and evolutionary considerations in fishery management: research needs for the future. *The Future of Fisheries Science in North America*. (eds R.J. Beamish & B.J. Rothschild), pp. 427–451. Springer Netherlands.

Waters, J.M. & Roy, M.S. (2004) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*, **13**, 2797–2806.

Watts, R.J. & Johnson, M.S. (2004) Estuaries, lagoons and enclosed embayments: habitats that enhance population subdivision of inshore fishes. *Marine and Freshwater Research*, **55**, 641–651.

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

Weetman, D., Hauser, L., Bayes, M.K., Ellis, J.R. & Shaw, P.W. (2006) Genetic population structure across a range of geographic scales in the commercially exploited marine gastropod *Buccinum undatum*. *Marine Ecology Progress Series*, **317**, 157–169.

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

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

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

Williams, D.A., Brown, S.D. & Crawford, D.L. (2008) Contemporary and historical influences on the genetic structure of the estuarine-dependent Gulf killifish *Fundulus grandis*. *Marine Ecology Progress Series*, **373**, 111–121.

Williams, J.R., Cryer, M., Hooker, S.H., Smith, M.D., Watson, T.G., Mackay, G. & Tasker, R. (2007) *Biomass survey and stock assessment of pipi (*Paphies australis*) on Mair Bank, Whangarei Harbour, 2005*. New Zealand Fisheries Assessment Report 2007/3, Wellington, Pp. 29.

Wilson, G.A. & Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A.M., Ricard, D., Rosenberg, A.A., Watson, R. & Zeller, D. (2009) Rebuilding global fisheries. *Science*, **325**, 578–585.

Wright, S. (1931) Evolution in Mendelian Populations. *Genetics*, **16**, 97–159.

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

Xiao, J., Cordes, J.F., Wang, H., Guo, X. & Reece, K.S. (2010) Population genetics of *Crassostrea ariakensis* in Asia inferred from microsatellite markers. *Marine Biology*, **157**, 1767–1781.

Zhan, A., Hu, J., Hu, X., Zhou, Z., Hui, M., Wang, S., Peng, W., Wang, M. & Bao, Z. (2009) Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): do local marine currents drive geographical differentiation? *Marine Biotechnology*, **11**, 223–235.

Appendix 1: Raw allelic frequency data

Supplementary Table 1: Raw allele frequencies for 11 *Paphies subtriangulata* microsatellite loci for each location and total frequency for all individuals sampled.

Supplementary Table 2: Raw allele frequencies for 13 *Paphies australis* microsatellite loci for each location and total frequency for all individuals sampled.

Supplementary Table 1.

Locus	Allele	Location										Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_1	140	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	148	0.010	0.020	0.020	0.018	0.009	0.010	0.064	0.009	0.000	0.083	0.024
	152	0.529	0.618	0.500	0.500	0.491	0.588	0.538	0.395	0.490	0.250	0.490
	156	0.067	0.029	0.010	0.054	0.036	0.020	0.013	0.026	0.020	0.000	0.028
	160	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.135	0.014
	164	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.004
	168	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.002
	192	0.010	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	196	0.000	0.010	0.000	0.000	0.000	0.010	0.000	0.000	0.010	0.010	0.004
	200	0.029	0.039	0.120	0.036	0.018	0.039	0.064	0.079	0.040	0.031	0.050
	204	0.048	0.010	0.050	0.027	0.018	0.020	0.013	0.035	0.010	0.031	0.026
	208	0.221	0.167	0.190	0.259	0.300	0.255	0.231	0.307	0.320	0.083	0.233
	212	0.058	0.088	0.070	0.054	0.082	0.049	0.064	0.096	0.070	0.083	0.071
	216	0.010	0.000	0.030	0.009	0.009	0.010	0.013	0.053	0.030	0.062	0.023
	220	0.019	0.000	0.010	0.036	0.009	0.000	0.000	0.000	0.010	0.073	0.016
	224	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.042	0.005
	232	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.010	0.003
	236	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.002
	240	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.002
	244	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.001
Psub_2	107	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001
	110	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.001
	112	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	115	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001
	117	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.001
	118	0.000	0.020	0.010	0.009	0.018	0.000	0.013	0.000	0.009	0.000	0.008
	120	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.004

Supplementary Table 1 continued.

Locus	Allele	Location									Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_2	121	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.009	0.000	0.000	0.003
	122	0.009	0.059	0.010	0.045	0.045	0.020	0.050	0.053	0.028	0.053	0.037
	123	0.000	0.000	0.000	0.000	0.000	0.010	0.025	0.000	0.000	0.000	0.004
	124	0.000	0.039	0.000	0.000	0.000	0.010	0.013	0.000	0.019	0.000	0.008
	125	0.009	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	126	0.557	0.490	0.529	0.536	0.518	0.549	0.575	0.509	0.713	0.351	0.533
	127	0.028	0.020	0.118	0.036	0.073	0.069	0.037	0.035	0.037	0.000	0.045
	128	0.132	0.176	0.127	0.080	0.100	0.127	0.087	0.114	0.065	0.255	0.126
	129	0.009	0.020	0.020	0.000	0.018	0.010	0.000	0.044	0.009	0.032	0.016
	130	0.038	0.098	0.088	0.116	0.064	0.127	0.050	0.079	0.037	0.064	0.076
	131	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.002
	132	0.057	0.039	0.049	0.080	0.073	0.049	0.050	0.061	0.056	0.138	0.065
	133	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	134	0.047	0.000	0.010	0.036	0.018	0.020	0.000	0.018	0.009	0.064	0.022
	136	0.028	0.029	0.000	0.018	0.027	0.000	0.037	0.035	0.000	0.021	0.020
	138	0.000	0.010	0.029	0.000	0.000	0.000	0.050	0.009	0.000	0.011	0.011
	140	0.019	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	142	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	144	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.009	0.000	0.004
	146	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	148	0.000	0.000	0.000	0.009	0.009	0.000	0.013	0.000	0.000	0.000	0.003
Psub_3	186	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.001
	189	0.040	0.098	0.078	0.111	0.130	0.090	0.163	0.132	0.093	0.031	0.097
	192	0.600	0.186	0.324	0.139	0.213	0.130	0.138	0.211	0.231	0.688	0.286
	195	0.340	0.637	0.529	0.657	0.519	0.710	0.562	0.623	0.602	0.271	0.545

Supplementary Table 1 continued.

Locus	Allele	Location										Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_3	198	0.020	0.078	0.069	0.093	0.130	0.070	0.138	0.035	0.074	0.000	0.071
	201	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.001
Psub_4	204	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.011	0.002
	207	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	210	0.009	0.031	0.020	0.009	0.009	0.031	0.025	0.070	0.019	0.000	0.022
	213	0.047	0.010	0.031	0.009	0.009	0.020	0.013	0.000	0.120	0.000	0.026
	216	0.906	0.939	0.949	0.953	0.982	0.918	0.950	0.921	0.852	0.989	0.936
	219	0.038	0.010	0.000	0.019	0.000	0.020	0.013	0.000	0.009	0.000	0.011
	222	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.009	0.000	0.000	0.002
Psub_5	89	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.010	0.000	0.004
	91	0.000	0.010	0.000	0.018	0.000	0.000	0.000	0.009	0.000	0.000	0.004
	93	0.153	0.051	0.100	0.127	0.147	0.053	0.113	0.164	0.154	0.000	0.106
	95	0.398	0.388	0.220	0.309	0.314	0.330	0.212	0.209	0.375	0.638	0.339
	97	0.347	0.398	0.440	0.464	0.343	0.415	0.512	0.445	0.375	0.234	0.397
	99	0.031	0.071	0.100	0.036	0.137	0.170	0.075	0.082	0.048	0.117	0.087
	101	0.000	0.041	0.040	0.018	0.000	0.000	0.000	0.027	0.029	0.000	0.016
	103	0.031	0.000	0.040	0.018	0.039	0.000	0.075	0.055	0.010	0.011	0.028
	105	0.000	0.020	0.020	0.000	0.010	0.000	0.000	0.009	0.000	0.000	0.006
	107	0.020	0.020	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	111	0.000	0.000	0.030	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.004
	113	0.020	0.000	0.000	0.009	0.010	0.000	0.000	0.000	0.000	0.000	0.004
Psub_6	212	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.001
	216	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.001
	220	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.001
	222	0.000	0.000	0.052	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	224	0.220	0.021	0.104	0.009	0.028	0.038	0.000	0.045	0.010	0.000	0.048
	226	0.073	0.000	0.104	0.009	0.019	0.000	0.000	0.009	0.029	0.160	0.040

Supplementary Table 1 continued.

Locus	Allele	Location									Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_6	230	0.049	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.008
	234	0.000	0.000	0.000	0.009	0.009	0.029	0.000	0.000	0.010	0.000	0.006
	236	0.317	0.426	0.312	0.473	0.417	0.375	0.333	0.357	0.373	0.606	0.399
	238	0.341	0.511	0.417	0.500	0.519	0.538	0.653	0.545	0.539	0.000	0.456
	240	0.000	0.043	0.010	0.000	0.009	0.010	0.014	0.018	0.020	0.202	0.033
	244	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.010	0.000	0.002
	250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.001
Psub_7	103	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	135	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	143	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	167	0.000	0.011	0.000	0.019	0.009	0.000	0.000	0.000	0.000	0.000	0.004
	203	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.001
	209	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	211	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.000	0.000	0.004
	215	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.002
	217	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	219	0.112	0.000	0.037	0.019	0.000	0.024	0.000	0.071	0.012	0.000	0.028
	221	0.194	0.054	0.300	0.115	0.111	0.107	0.125	0.153	0.195	0.298	0.165
	223	0.031	0.054	0.100	0.087	0.102	0.071	0.062	0.112	0.134	0.083	0.084
	225	0.071	0.022	0.037	0.077	0.065	0.107	0.016	0.041	0.061	0.000	0.050
	227	0.296	0.522	0.250	0.394	0.370	0.405	0.375	0.306	0.305	0.369	0.359
	229	0.122	0.228	0.138	0.212	0.250	0.190	0.297	0.082	0.085	0.024	0.163
	231	0.000	0.033	0.100	0.029	0.009	0.095	0.062	0.010	0.098	0.048	0.048
	233	0.061	0.054	0.025	0.000	0.000	0.000	0.000	0.051	0.000	0.143	0.033
	235	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.003
	237	0.031	0.000	0.000	0.000	0.056	0.000	0.062	0.051	0.061	0.000	0.026
	239	0.010	0.000	0.000	0.029	0.009	0.000	0.000	0.041	0.000	0.000	0.009

Supplementary Table 1 continued.

Locus	Allele	Location										Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_7	241	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.024	0.000	0.003
	245	0.010	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.024	0.000	0.004
	247	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.020	0.000	0.000	0.003
Psub_8	202	0.015	0.079	0.021	0.000	0.009	0.150	0.059	0.018	0.129	0.052	0.053
	220	0.015	0.000	0.000	0.000	0.000	0.075	0.000	0.000	0.000	0.000	0.009
	222	0.000	0.000	0.000	0.026	0.000	0.050	0.000	0.000	0.000	0.000	0.008
	226	0.045	0.026	0.000	0.026	0.000	0.050	0.000	0.000	0.000	0.000	0.015
	228	0.197	0.026	0.000	0.026	0.000	0.000	0.000	0.000	0.014	0.083	0.035
	230	0.197	0.237	0.074	0.372	0.142	0.175	0.191	0.109	0.200	0.083	0.178
	232	0.015	0.026	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.333	0.040
	234	0.045	0.053	0.074	0.013	0.085	0.075	0.044	0.073	0.086	0.000	0.055
	236	0.470	0.434	0.745	0.423	0.594	0.375	0.544	0.718	0.500	0.438	0.524
	238	0.000	0.118	0.085	0.090	0.160	0.025	0.162	0.064	0.071	0.000	0.078
	240	0.000	0.000	0.000	0.026	0.009	0.000	0.000	0.009	0.000	0.010	0.005
	248	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001
Psub_9	133	0.038	0.029	0.049	0.009	0.027	0.010	0.050	0.009	0.009	0.000	0.023
	135	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.003
	137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.004
	139	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	141	0.858	0.873	0.882	0.929	0.882	0.942	0.863	0.825	0.880	0.812	0.875
	143	0.057	0.069	0.000	0.045	0.055	0.038	0.062	0.061	0.046	0.104	0.054
	145	0.028	0.000	0.039	0.000	0.009	0.010	0.000	0.026	0.065	0.021	0.020
	147	0.000	0.000	0.010	0.009	0.000	0.000	0.000	0.000	0.000	0.021	0.004
	149	0.000	0.000	0.010	0.000	0.009	0.000	0.013	0.009	0.000	0.000	0.004
	151	0.009	0.020	0.010	0.000	0.009	0.000	0.013	0.009	0.000	0.000	0.007
	153	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.035	0.000	0.000	0.004
	155	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.002

Supplementary Table 1 continued.

Locus	Allele	Location										Total
		RUA	PAP	WMR	MAR	PKR	COL	OAK	KAK	WPK	CHA	
Psub_10	171	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	173	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	175	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.010	0.002
	181	0.038	0.029	0.020	0.018	0.009	0.029	0.025	0.035	0.028	0.000	0.023
	183	0.048	0.049	0.020	0.000	0.019	0.000	0.013	0.035	0.019	0.000	0.020
	189	0.010	0.000	0.059	0.018	0.009	0.010	0.025	0.044	0.037	0.000	0.021
	191	0.529	0.539	0.627	0.652	0.676	0.625	0.662	0.614	0.574	0.906	0.640
	193	0.058	0.049	0.039	0.080	0.093	0.058	0.013	0.061	0.037	0.073	0.056
	195	0.154	0.157	0.069	0.098	0.093	0.048	0.075	0.053	0.083	0.000	0.083
	197	0.000	0.000	0.020	0.000	0.000	0.010	0.000	0.018	0.009	0.010	0.007
	199	0.000	0.000	0.020	0.009	0.000	0.000	0.000	0.026	0.028	0.000	0.008
	201	0.048	0.088	0.059	0.071	0.037	0.029	0.025	0.044	0.065	0.000	0.047
	203	0.019	0.029	0.010	0.018	0.019	0.058	0.050	0.018	0.019	0.000	0.024
	205	0.058	0.039	0.029	0.000	0.037	0.077	0.075	0.035	0.065	0.000	0.042
	207	0.000	0.000	0.020	0.009	0.009	0.010	0.025	0.000	0.009	0.000	0.008
	209	0.019	0.020	0.010	0.009	0.000	0.029	0.013	0.018	0.000	0.000	0.012
	215	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001
	217	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.002
	221	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.001
	235	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Psub_11	96	0.000	0.098	0.020	0.054	0.018	0.019	0.087	0.044	0.000	0.000	0.034
	100	0.623	0.431	0.657	0.464	0.509	0.471	0.525	0.561	0.648	1.000	0.589
	104	0.358	0.451	0.314	0.429	0.445	0.490	0.375	0.377	0.324	0.000	0.356
	108	0.019	0.020	0.010	0.054	0.018	0.019	0.000	0.018	0.028	0.000	0.019
	112	0.000	0.000	0.000	0.000	0.009	0.000	0.013	0.000	0.000	0.000	0.002

Supplementary Table 2.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_1	153	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	165	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	169	0.045	0.056	0.064	0.037	0.047	0.020	0.036	0.073	0.087	0.060	0.009	0.057	0.010	0.046
	173	0.091	0.028	0.091	0.037	0.009	0.090	0.080	0.045	0.115	0.070	0.038	0.043	0.020	0.058
	177	0.382	0.417	0.291	0.444	0.311	0.370	0.464	0.418	0.462	0.510	0.462	0.386	0.471	0.414
	181	0.273	0.231	0.318	0.315	0.340	0.310	0.232	0.273	0.183	0.230	0.236	0.271	0.304	0.270
	185	0.091	0.148	0.155	0.130	0.151	0.130	0.107	0.136	0.115	0.070	0.198	0.157	0.127	0.132
	189	0.055	0.074	0.064	0.037	0.094	0.060	0.062	0.055	0.038	0.030	0.047	0.071	0.049	0.057
	193	0.036	0.019	0.000	0.000	0.009	0.010	0.000	0.000	0.000	0.010	0.009	0.014	0.020	0.010
	197	0.009	0.000	0.009	0.000	0.019	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.004
	201	0.000	0.009	0.009	0.000	0.009	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.003
	203	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	205	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	217	0.009	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Paus_2	227	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.001
	229	0.000	0.000	0.000	0.019	0.000	0.000	0.019	0.010	0.000	0.000	0.000	0.000	0.000	0.004
	231	0.010	0.000	0.000	0.019	0.000	0.011	0.009	0.000	0.000	0.000	0.010	0.000	0.000	0.005
	232	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.002
	233	0.010	0.010	0.000	0.000	0.029	0.033	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.008
	235	0.010	0.000	0.009	0.019	0.019	0.000	0.009	0.010	0.000	0.000	0.000	0.000	0.000	0.006
	236	0.000	0.019	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	237	0.000	0.019	0.000	0.009	0.038	0.000	0.019	0.000	0.020	0.000	0.010	0.000	0.000	0.009
	238	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.001
	239	0.000	0.000	0.000	0.009	0.000	0.033	0.028	0.019	0.010	0.010	0.019	0.000	0.020	0.011
	241	0.020	0.000	0.038	0.009	0.019	0.011	0.000	0.029	0.000	0.000	0.000	0.000	0.000	0.010
	243	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.010	0.000	0.000	0.002
	247	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001

Supplementary Table 2 continued.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_2	248	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.001
	251	0.010	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	253	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.002
	255	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.010	0.000	0.000	0.000	0.002
	257	0.000	0.010	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.005
	259	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.002
	261	0.029	0.048	0.038	0.019	0.000	0.076	0.019	0.019	0.039	0.010	0.019	0.000	0.020	0.026
	263	0.049	0.048	0.075	0.075	0.010	0.076	0.066	0.000	0.020	0.060	0.096	0.030	0.071	0.052
	264	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.002
	265	0.039	0.125	0.019	0.009	0.096	0.076	0.047	0.038	0.029	0.030	0.038	0.061	0.000	0.047
	267	0.147	0.250	0.236	0.151	0.269	0.207	0.283	0.087	0.206	0.160	0.154	0.258	0.265	0.206
	268	0.000	0.010	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.029	0.015	0.000	0.006
	269	0.137	0.202	0.085	0.038	0.077	0.141	0.132	0.010	0.108	0.060	0.173	0.121	0.133	0.109
	270	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.010	0.000	0.010	0.000	0.000	0.000	0.003
	271	0.059	0.038	0.066	0.085	0.058	0.076	0.057	0.010	0.069	0.080	0.087	0.076	0.041	0.062
	272	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.003
	273	0.078	0.029	0.057	0.075	0.048	0.000	0.047	0.058	0.020	0.070	0.038	0.091	0.031	0.049
	274	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.004
	275	0.020	0.010	0.047	0.057	0.048	0.011	0.038	0.019	0.010	0.010	0.010	0.030	0.051	0.028
	276	0.000	0.000	0.000	0.000	0.010	0.033	0.000	0.010	0.000	0.000	0.010	0.000	0.000	0.005
	277	0.118	0.038	0.057	0.028	0.000	0.033	0.038	0.048	0.059	0.030	0.067	0.015	0.061	0.046
	278	0.000	0.000	0.000	0.009	0.010	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.020	0.004
	279	0.000	0.010	0.009	0.000	0.000	0.022	0.009	0.010	0.010	0.000	0.000	0.000	0.061	0.010
	280	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.010	0.000	0.010	0.000	0.000	0.000	0.002
	281	0.000	0.010	0.000	0.019	0.000	0.000	0.000	0.038	0.020	0.010	0.000	0.000	0.000	0.007
	282	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.001
	283	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.020	0.000	0.000	0.000	0.000	0.002

Supplementary Table 2 continued.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_2	284	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.020	0.002
	286	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.001
	287	0.265	0.125	0.226	0.311	0.202	0.152	0.179	0.404	0.324	0.310	0.231	0.273	0.204	0.247
	289	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
	290	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.002
	292	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
	293	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001
	303	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001
Paus_3	131	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.010	0.000	0.000	0.014	0.000	0.003
	134	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	137	0.000	0.010	0.036	0.000	0.000	0.000	0.000	0.012	0.020	0.000	0.028	0.014	0.020	0.011
	140	0.018	0.060	0.036	0.019	0.000	0.021	0.010	0.000	0.029	0.056	0.000	0.000	0.000	0.019
	143	0.027	0.030	0.091	0.009	0.010	0.000	0.010	0.000	0.000	0.089	0.094	0.000	0.071	0.033
	146	0.009	0.000	0.055	0.037	0.029	0.021	0.049	0.037	0.010	0.033	0.009	0.000	0.020	0.024
	149	0.027	0.010	0.009	0.009	0.029	0.052	0.020	0.012	0.039	0.022	0.009	0.000	0.051	0.022
	152	0.082	0.010	0.018	0.046	0.019	0.052	0.059	0.061	0.010	0.011	0.028	0.014	0.071	0.037
	155	0.082	0.110	0.073	0.065	0.115	0.052	0.108	0.037	0.049	0.056	0.075	0.043	0.102	0.074
	158	0.509	0.580	0.445	0.611	0.596	0.583	0.431	0.598	0.471	0.444	0.491	0.500	0.337	0.507
	164	0.027	0.050	0.027	0.056	0.038	0.094	0.098	0.098	0.069	0.078	0.132	0.129	0.122	0.078
	167	0.073	0.040	0.064	0.046	0.077	0.031	0.118	0.000	0.069	0.100	0.066	0.029	0.051	0.059
	170	0.045	0.030	0.055	0.009	0.048	0.010	0.039	0.098	0.157	0.033	0.066	0.114	0.071	0.060
	173	0.055	0.010	0.045	0.046	0.010	0.010	0.010	0.024	0.000	0.022	0.000	0.014	0.041	0.022
	176	0.000	0.030	0.000	0.028	0.000	0.031	0.010	0.012	0.039	0.011	0.000	0.014	0.020	0.015
	179	0.000	0.000	0.018	0.000	0.029	0.000	0.000	0.000	0.010	0.011	0.000	0.000	0.000	0.005
	182	0.027	0.020	0.000	0.000	0.000	0.010	0.000	0.000	0.020	0.022	0.000	0.029	0.020	0.011
	188	0.009	0.010	0.000	0.009	0.000	0.010	0.000	0.000	0.000	0.011	0.000	0.057	0.000	0.008
	191	0.009	0.000	0.000	0.000	0.000	0.021	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.004

Supplementary Table 2 continued.

Locus	Allele	Location												Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_3	197	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.000	0.003
	206	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	209	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.001
Paus_4	258	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	261	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.002
	264	0.000	0.000	0.000	0.012	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	267	0.000	0.000	0.062	0.023	0.044	0.056	0.020	0.010	0.000	0.011	0.000	0.000	0.021	0.019
	270	0.000	0.022	0.083	0.000	0.011	0.014	0.041	0.010	0.000	0.022	0.000	0.017	0.021	0.019
	271	0.000	0.044	0.000	0.000	0.000	0.000	0.020	0.000	0.010	0.000	0.000	0.000	0.021	0.007
	272	0.000	0.044	0.031	0.000	0.033	0.000	0.020	0.019	0.000	0.033	0.000	0.000	0.000	0.014
	273	0.010	0.156	0.125	0.047	0.022	0.083	0.092	0.077	0.094	0.000	0.070	0.100	0.117	0.076
	276	0.088	0.044	0.073	0.058	0.067	0.167	0.122	0.067	0.104	0.087	0.050	0.183	0.191	0.100
	278	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	279	0.667	0.500	0.438	0.616	0.467	0.431	0.500	0.394	0.573	0.554	0.640	0.467	0.479	0.517
	282	0.088	0.056	0.052	0.070	0.178	0.111	0.082	0.183	0.083	0.076	0.050	0.050	0.053	0.087
	285	0.127	0.044	0.083	0.151	0.122	0.097	0.071	0.173	0.062	0.163	0.120	0.183	0.074	0.113
	288	0.000	0.056	0.031	0.000	0.011	0.028	0.020	0.029	0.052	0.033	0.010	0.000	0.021	0.022
	291	0.020	0.022	0.010	0.000	0.000	0.014	0.000	0.019	0.000	0.011	0.010	0.000	0.000	0.008
	294	0.000	0.011	0.000	0.023	0.011	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.004
	297	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.002
	300	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.040	0.000	0.000	0.005
Paus_5	186	0.000	0.000	0.000	0.019	0.000	0.000	0.009	0.018	0.000	0.000	0.000	0.000	0.010	0.004
	190	0.000	0.009	0.000	0.009	0.000	0.000	0.000	0.009	0.000	0.000	0.019	0.014	0.010	0.005
	194	0.056	0.009	0.038	0.028	0.057	0.052	0.019	0.036	0.067	0.108	0.085	0.100	0.078	0.056
	198	0.037	0.056	0.048	0.065	0.066	0.021	0.075	0.064	0.077	0.069	0.057	0.129	0.078	0.065
	202	0.056	0.056	0.087	0.037	0.057	0.062	0.038	0.055	0.029	0.078	0.028	0.043	0.010	0.049
	206	0.815	0.676	0.673	0.704	0.708	0.740	0.717	0.691	0.692	0.588	0.689	0.614	0.716	0.694

Supplementary Table 2 continued.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_5	210	0.009	0.046	0.010	0.009	0.019	0.062	0.038	0.027	0.010	0.039	0.038	0.014	0.020	0.026
	214	0.028	0.093	0.096	0.093	0.075	0.052	0.085	0.064	0.096	0.059	0.085	0.086	0.059	0.075
	218	0.000	0.028	0.010	0.037	0.009	0.010	0.019	0.027	0.010	0.020	0.000	0.000	0.010	0.014
	222	0.000	0.028	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.004
	226	0.000	0.000	0.010	0.000	0.009	0.000	0.000	0.009	0.019	0.029	0.000	0.000	0.000	0.006
	230	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.002
Paus_7	86	0.009	0.009	0.009	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.003
	92	0.064	0.037	0.027	0.009	0.151	0.040	0.027	0.000	0.019	0.000	0.009	0.000	0.029	0.032
	94	0.000	0.000	0.000	0.000	0.000	0.050	0.009	0.009	0.000	0.010	0.019	0.014	0.000	0.009
	96	0.082	0.111	0.100	0.111	0.066	0.070	0.161	0.009	0.125	0.167	0.094	0.071	0.088	0.097
	98	0.682	0.565	0.582	0.713	0.396	0.560	0.536	0.664	0.538	0.559	0.547	0.500	0.441	0.560
	100	0.018	0.028	0.000	0.019	0.009	0.020	0.027	0.055	0.010	0.000	0.000	0.043	0.010	0.018
	102	0.000	0.028	0.027	0.019	0.009	0.000	0.018	0.018	0.029	0.029	0.019	0.014	0.000	0.016
	104	0.018	0.000	0.000	0.000	0.009	0.000	0.018	0.000	0.019	0.000	0.028	0.000	0.029	0.009
	106	0.127	0.213	0.255	0.130	0.340	0.250	0.196	0.236	0.250	0.235	0.283	0.329	0.382	0.248
	108	0.000	0.000	0.000	0.000	0.019	0.010	0.009	0.009	0.000	0.000	0.000	0.029	0.020	0.007
	110	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Paus_8	210	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.003
	216	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
	219	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	222	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.011	0.000	0.021	0.000	0.000	0.000	0.005
	225	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.002
	228	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.002
	231	0.000	0.057	0.000	0.000	0.000	0.027	0.000	0.011	0.064	0.062	0.000	0.068	0.000	0.022
	234	0.000	0.029	0.013	0.019	0.000	0.000	0.061	0.034	0.000	0.031	0.009	0.000	0.015	0.016
	237	0.036	0.086	0.062	0.077	0.032	0.027	0.082	0.080	0.021	0.083	0.019	0.000	0.044	0.050
	240	0.073	0.071	0.025	0.087	0.129	0.027	0.061	0.205	0.021	0.052	0.151	0.045	0.029	0.075

Supplementary Table 2 continued.

Locus	Allele	Location												Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_8	243	0.427	0.371	0.263	0.404	0.355	0.527	0.327	0.398	0.362	0.365	0.292	0.318	0.456	0.374
	246	0.218	0.186	0.250	0.240	0.226	0.216	0.255	0.136	0.245	0.208	0.151	0.318	0.206	0.220
	249	0.073	0.071	0.150	0.048	0.161	0.068	0.122	0.057	0.064	0.062	0.198	0.000	0.103	0.091
	252	0.082	0.043	0.062	0.067	0.016	0.041	0.071	0.023	0.149	0.052	0.028	0.000	0.074	0.054
	255	0.073	0.000	0.000	0.010	0.000	0.000	0.000	0.011	0.011	0.031	0.094	0.023	0.044	0.023
	258	0.000	0.029	0.025	0.000	0.000	0.000	0.000	0.011	0.021	0.010	0.057	0.045	0.015	0.016
	261	0.000	0.000	0.013	0.000	0.000	0.000	0.010	0.000	0.021	0.010	0.000	0.023	0.000	0.006
	264	0.000	0.000	0.025	0.029	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	267	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.008
	273	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.003
	276	0.018	0.057	0.050	0.000	0.081	0.041	0.000	0.023	0.000	0.000	0.000	0.000	0.000	0.021
Paus_9	103	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	109	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.010	0.000	0.038	0.042	0.008
	112	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.010	0.000	0.000	0.000	0.002
	115	0.043	0.010	0.019	0.012	0.023	0.024	0.000	0.064	0.000	0.052	0.000	0.000	0.028	0.021
	118	0.098	0.020	0.019	0.047	0.034	0.000	0.068	0.074	0.071	0.083	0.029	0.058	0.014	0.047
	121	0.011	0.060	0.067	0.070	0.068	0.048	0.149	0.106	0.114	0.177	0.103	0.077	0.139	0.091
	124	0.054	0.110	0.077	0.070	0.080	0.167	0.149	0.053	0.129	0.115	0.088	0.327	0.069	0.114
	127	0.685	0.670	0.644	0.523	0.580	0.595	0.473	0.255	0.243	0.156	0.279	0.115	0.250	0.421
	130	0.076	0.070	0.067	0.151	0.080	0.083	0.095	0.245	0.100	0.229	0.147	0.212	0.208	0.136
	133	0.011	0.010	0.010	0.012	0.068	0.060	0.068	0.106	0.243	0.094	0.088	0.038	0.167	0.075
	136	0.022	0.010	0.058	0.093	0.068	0.000	0.000	0.011	0.014	0.062	0.221	0.000	0.042	0.046
	139	0.000	0.040	0.000	0.012	0.000	0.000	0.000	0.032	0.000	0.000	0.029	0.096	0.028	0.018
	142	0.000	0.000	0.038	0.000	0.000	0.024	0.000	0.011	0.057	0.010	0.015	0.019	0.014	0.014
	145	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.019	0.000	0.002
	148	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.002
	169	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.001

Supplementary Table 2 continued.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_10	164	0.000	0.000	0.009	0.009	0.009	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.005
	166	0.009	0.019	0.000	0.009	0.009	0.000	0.000	0.055	0.010	0.020	0.000	0.014	0.010	0.012
	170	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.009	0.000	0.000	0.002
	174	0.000	0.000	0.009	0.009	0.000	0.000	0.018	0.009	0.000	0.020	0.000	0.014	0.000	0.006
	178	0.018	0.000	0.009	0.009	0.009	0.031	0.018	0.018	0.010	0.020	0.028	0.014	0.010	0.015
	180	0.027	0.010	0.009	0.009	0.000	0.000	0.009	0.009	0.000	0.010	0.009	0.000	0.000	0.007
	182	0.236	0.212	0.343	0.352	0.368	0.265	0.080	0.127	0.163	0.090	0.160	0.186	0.133	0.209
	184	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	186	0.109	0.144	0.185	0.204	0.198	0.173	0.223	0.109	0.154	0.150	0.104	0.114	0.133	0.154
	188	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001
	190	0.309	0.250	0.111	0.093	0.179	0.265	0.330	0.300	0.212	0.250	0.349	0.357	0.276	0.252
	194	0.182	0.192	0.194	0.194	0.113	0.173	0.223	0.291	0.317	0.300	0.264	0.214	0.306	0.228
	196	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.003
	198	0.036	0.067	0.056	0.093	0.057	0.051	0.054	0.045	0.087	0.100	0.038	0.071	0.031	0.060
	202	0.036	0.038	0.046	0.019	0.009	0.020	0.027	0.009	0.038	0.010	0.000	0.014	0.031	0.023
	206	0.009	0.048	0.019	0.000	0.019	0.010	0.000	0.000	0.000	0.020	0.000	0.000	0.031	0.012
	210	0.009	0.010	0.000	0.000	0.019	0.000	0.000	0.009	0.010	0.000	0.009	0.000	0.010	0.006
	212	0.000	0.000	0.009	0.000	0.009	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.002
	216	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
Paus_11	218	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.001
	236	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.011	0.000	0.000	0.000	0.002
	239	0.000	0.038	0.010	0.019	0.000	0.000	0.018	0.000	0.000	0.056	0.009	0.000	0.000	0.012
	242	0.055	0.057	0.049	0.009	0.020	0.033	0.045	0.033	0.054	0.044	0.009	0.135	0.033	0.044
	245	0.073	0.085	0.049	0.120	0.088	0.089	0.045	0.033	0.054	0.022	0.094	0.096	0.056	0.070
	248	0.255	0.274	0.294	0.204	0.353	0.244	0.259	0.233	0.196	0.333	0.274	0.308	0.211	0.264
	251	0.500	0.500	0.520	0.574	0.363	0.478	0.518	0.533	0.587	0.444	0.547	0.442	0.533	0.503
	254	0.082	0.019	0.059	0.056	0.127	0.111	0.071	0.122	0.065	0.056	0.066	0.019	0.111	0.074

Supplementary Table 2 continued.

Locus	Allele	Location												Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_11	257	0.009	0.019	0.010	0.000	0.049	0.011	0.036	0.022	0.033	0.011	0.000	0.000	0.011	0.016
	260	0.027	0.009	0.010	0.019	0.000	0.022	0.009	0.000	0.011	0.000	0.000	0.000	0.044	0.012
	263	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.001
Paus_12	173	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	179	0.000	0.020	0.000	0.000	0.000	0.010	0.009	0.018	0.010	0.020	0.009	0.029	0.000	0.010
	183	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	185	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	187	0.000	0.049	0.028	0.009	0.047	0.021	0.018	0.009	0.029	0.078	0.009	0.014	0.010	0.025
	189	0.000	0.000	0.009	0.000	0.009	0.010	0.000	0.009	0.010	0.000	0.000	0.000	0.010	0.004
	191	0.536	0.353	0.537	0.454	0.462	0.552	0.518	0.491	0.471	0.539	0.509	0.457	0.363	0.480
	193	0.400	0.108	0.389	0.463	0.208	0.198	0.188	0.400	0.337	0.304	0.415	0.200	0.304	0.301
	195	0.027	0.108	0.019	0.009	0.047	0.062	0.045	0.000	0.038	0.039	0.009	0.114	0.069	0.045
	197	0.018	0.275	0.009	0.037	0.208	0.115	0.196	0.055	0.077	0.010	0.038	0.100	0.186	0.102
	199	0.000	0.069	0.000	0.009	0.009	0.031	0.027	0.018	0.029	0.000	0.000	0.086	0.039	0.024
	201	0.009	0.000	0.009	0.009	0.009	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.020	0.005
	209	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001
Paus_13	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001
	106	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.010	0.000	0.000	0.000	0.003
	108	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.001
	110	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	112	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.020	0.003
	114	0.764	0.778	0.718	0.722	0.632	0.720	0.714	0.818	0.692	0.833	0.830	0.743	0.716	0.745
	116	0.164	0.176	0.209	0.222	0.236	0.190	0.214	0.118	0.212	0.147	0.094	0.214	0.167	0.182
	118	0.036	0.037	0.027	0.028	0.085	0.050	0.045	0.018	0.048	0.010	0.028	0.043	0.059	0.040
	120	0.000	0.009	0.018	0.009	0.028	0.020	0.018	0.009	0.019	0.000	0.009	0.000	0.029	0.013
	122	0.018	0.000	0.018	0.000	0.019	0.000	0.000	0.000	0.010	0.000	0.009	0.000	0.000	0.006
	124	0.000	0.000	0.009	0.019	0.000	0.000	0.009	0.000	0.000	0.000	0.009	0.000	0.000	0.004

Supplementary Table 2 continued.

Locus	Allele	Location													Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_13	130	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.002
	132	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	136	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.010	0.000	0.000	0.000	0.000	0.001
Paus_14	138	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	141	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	147	0.027	0.028	0.000	0.047	0.020	0.011	0.009	0.009	0.050	0.030	0.038	0.015	0.021	0.023
	150	0.045	0.074	0.019	0.000	0.020	0.011	0.028	0.009	0.020	0.040	0.009	0.074	0.000	0.027
	153	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.029	0.000	0.004
	156	0.027	0.000	0.000	0.009	0.000	0.000	0.000	0.009	0.000	0.010	0.000	0.000	0.000	0.004
	159	0.000	0.009	0.000	0.000	0.010	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	165	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.002
	168	0.000	0.019	0.019	0.009	0.020	0.054	0.120	0.037	0.040	0.000	0.009	0.088	0.010	0.033
	171	0.000	0.000	0.000	0.019	0.000	0.043	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.006
	174	0.018	0.019	0.000	0.000	0.000	0.022	0.009	0.019	0.000	0.000	0.009	0.015	0.000	0.009
	176	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	177	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.015	0.000	0.003
	180	0.145	0.130	0.173	0.075	0.112	0.185	0.056	0.102	0.020	0.110	0.085	0.059	0.115	0.105
	183	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	186	0.000	0.000	0.000	0.000	0.000	0.011	0.009	0.000	0.010	0.000	0.000	0.000	0.000	0.002
	189	0.227	0.250	0.394	0.330	0.347	0.250	0.370	0.407	0.340	0.300	0.396	0.309	0.438	0.335
	192	0.091	0.185	0.144	0.123	0.235	0.130	0.130	0.120	0.080	0.150	0.094	0.118	0.156	0.135
	194	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	195	0.291	0.074	0.173	0.236	0.082	0.141	0.120	0.176	0.220	0.280	0.179	0.176	0.073	0.171
	197	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	198	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.001
	200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
	209	0.009	0.019	0.000	0.009	0.000	0.033	0.000	0.000	0.010	0.010	0.019	0.000	0.000	0.008

Supplementary Table 2 continued.

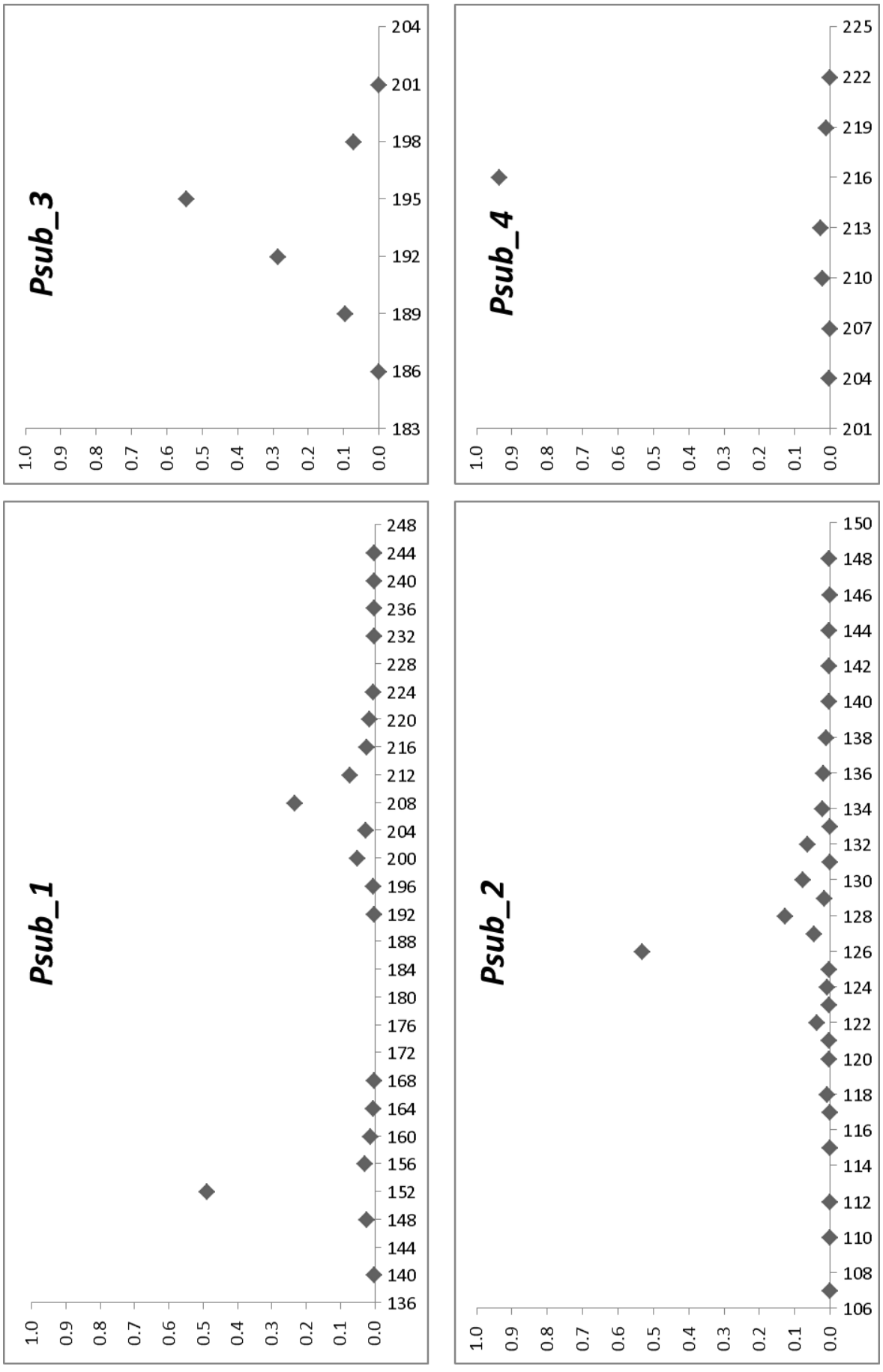
Locus	Allele	Location												Total
		RAG	HUI	TAP	WAI	TAU	NAP	PET	HAK	LYT	BLU	DBT	OKU	KAR	
Paus_14	215	0.009	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	222	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	227	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001
	228	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	229	0.091	0.130	0.029	0.123	0.122	0.087	0.111	0.074	0.160	0.040	0.142	0.103	0.188	0.108
	233	0.000	0.000	0.000	0.000	0.010	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	241	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.001
	242	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	253	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.002
	257	0.000	0.019	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002

Appendix 2: Allele frequency graphs

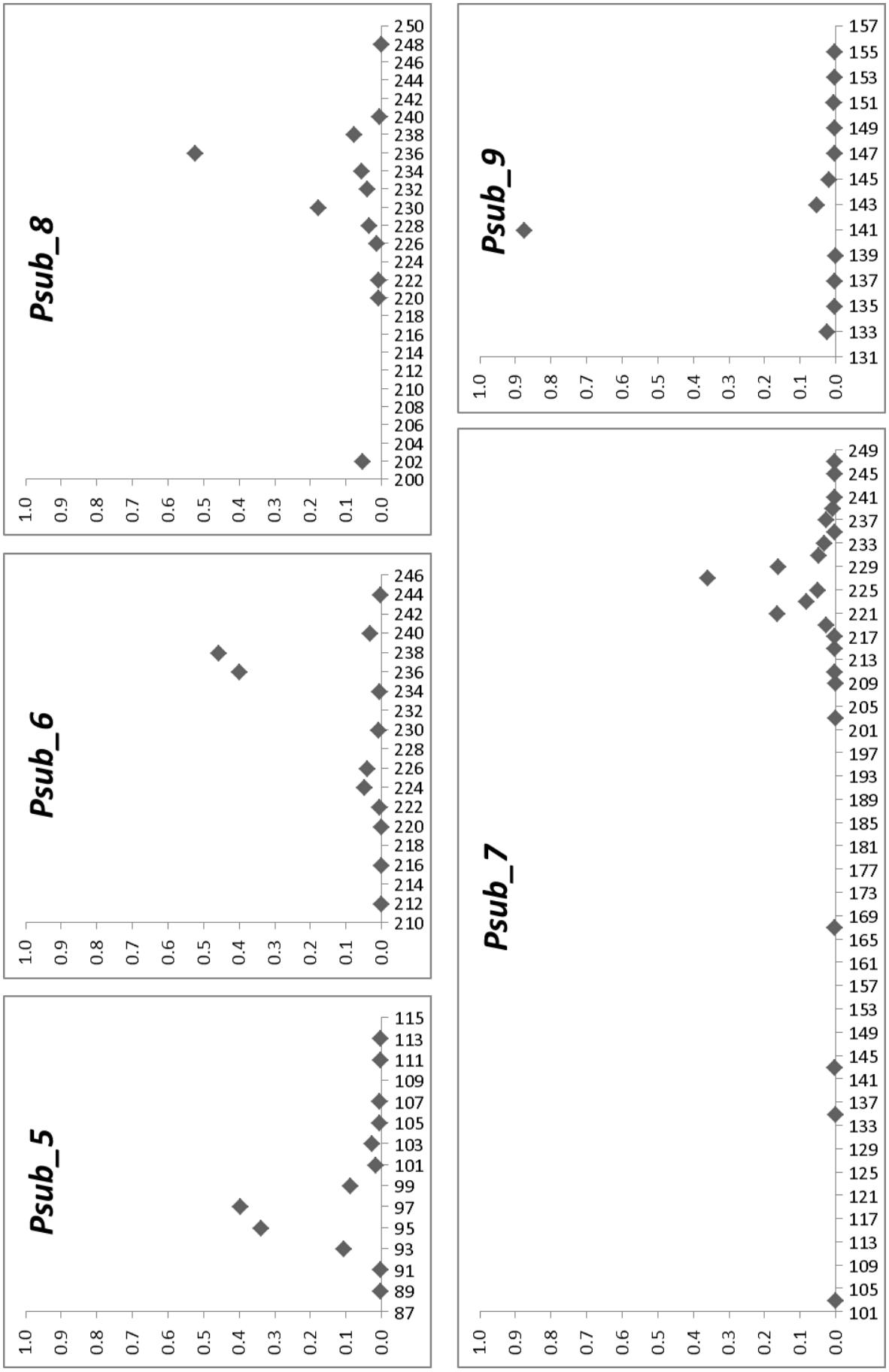
Supplementary Figure 1: Total frequency of observed alleles at each locus for all locations of *Paphies subtriangulata*. X axis = allele length, Y axis = observed frequency of allele.

Supplementary Figure 2: Total frequency of observed alleles at each locus for all locations of *Paphies australis*. X axis = allele length, Y axis = observed frequency of allele.

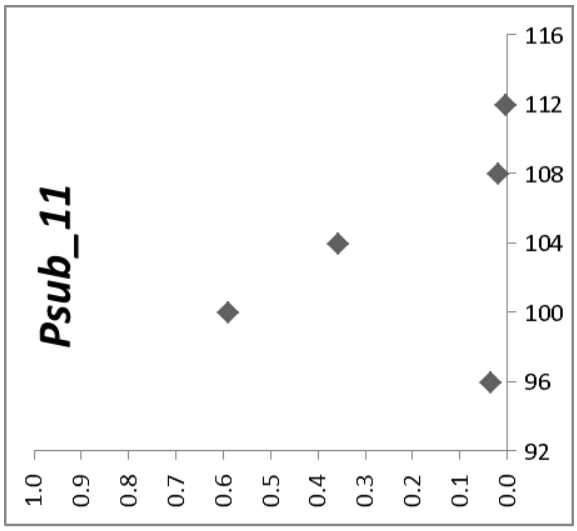
Supplementary Figure 1.

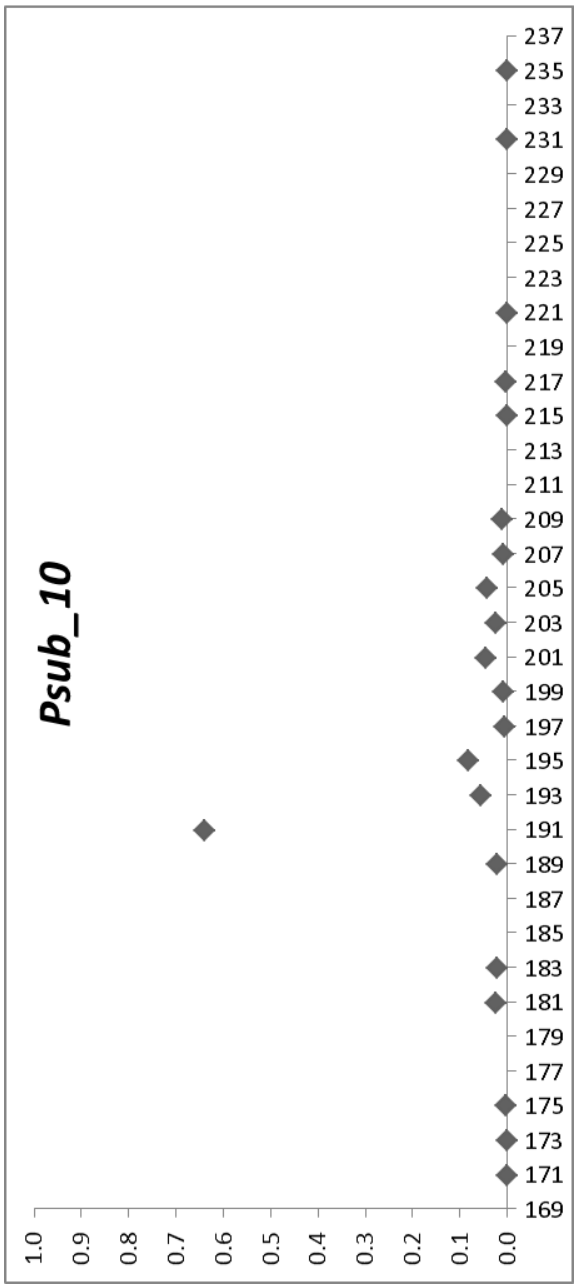


Supplementary Figure 1 continued.

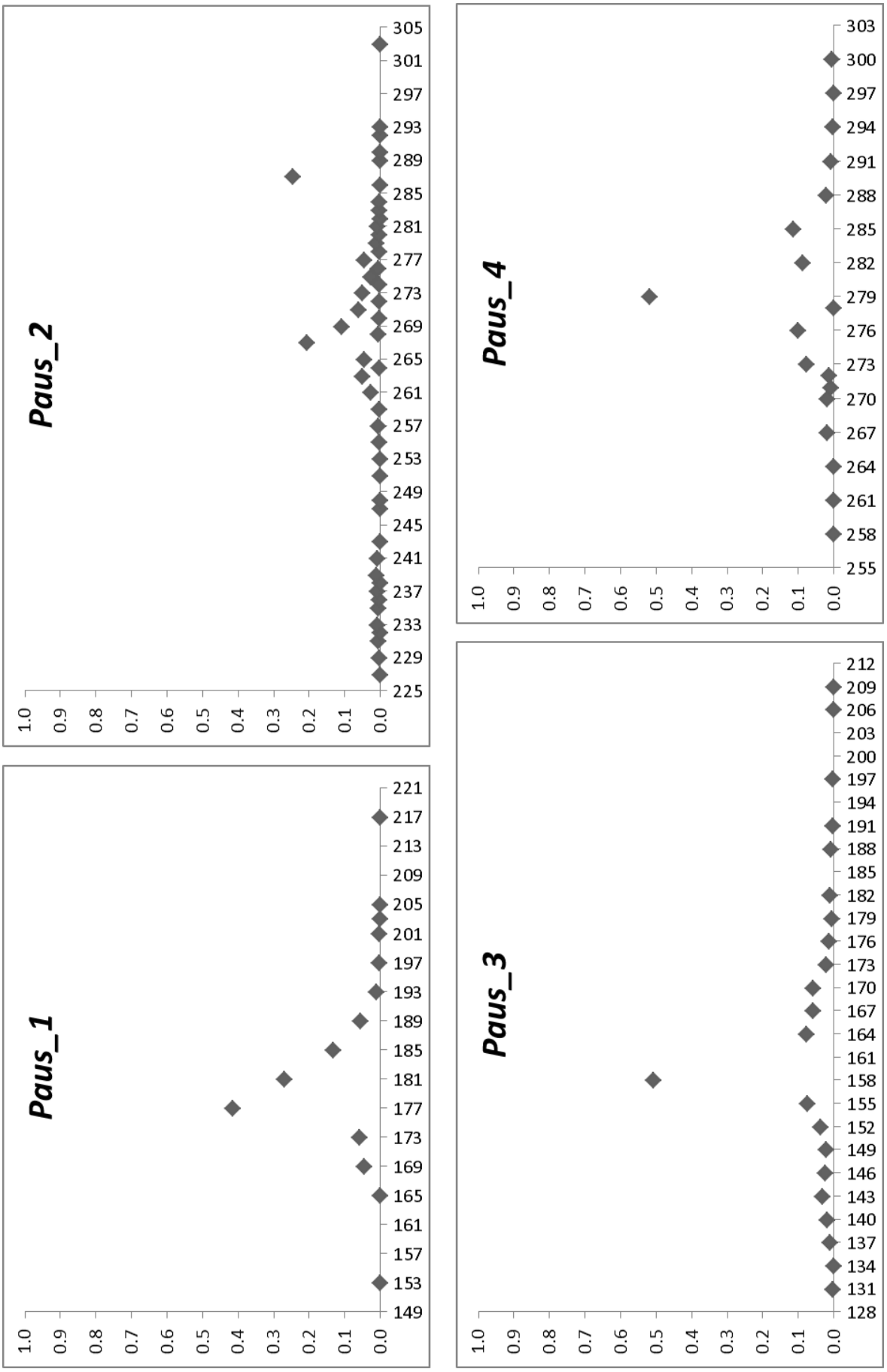


Supplementary Figure 1 continued.

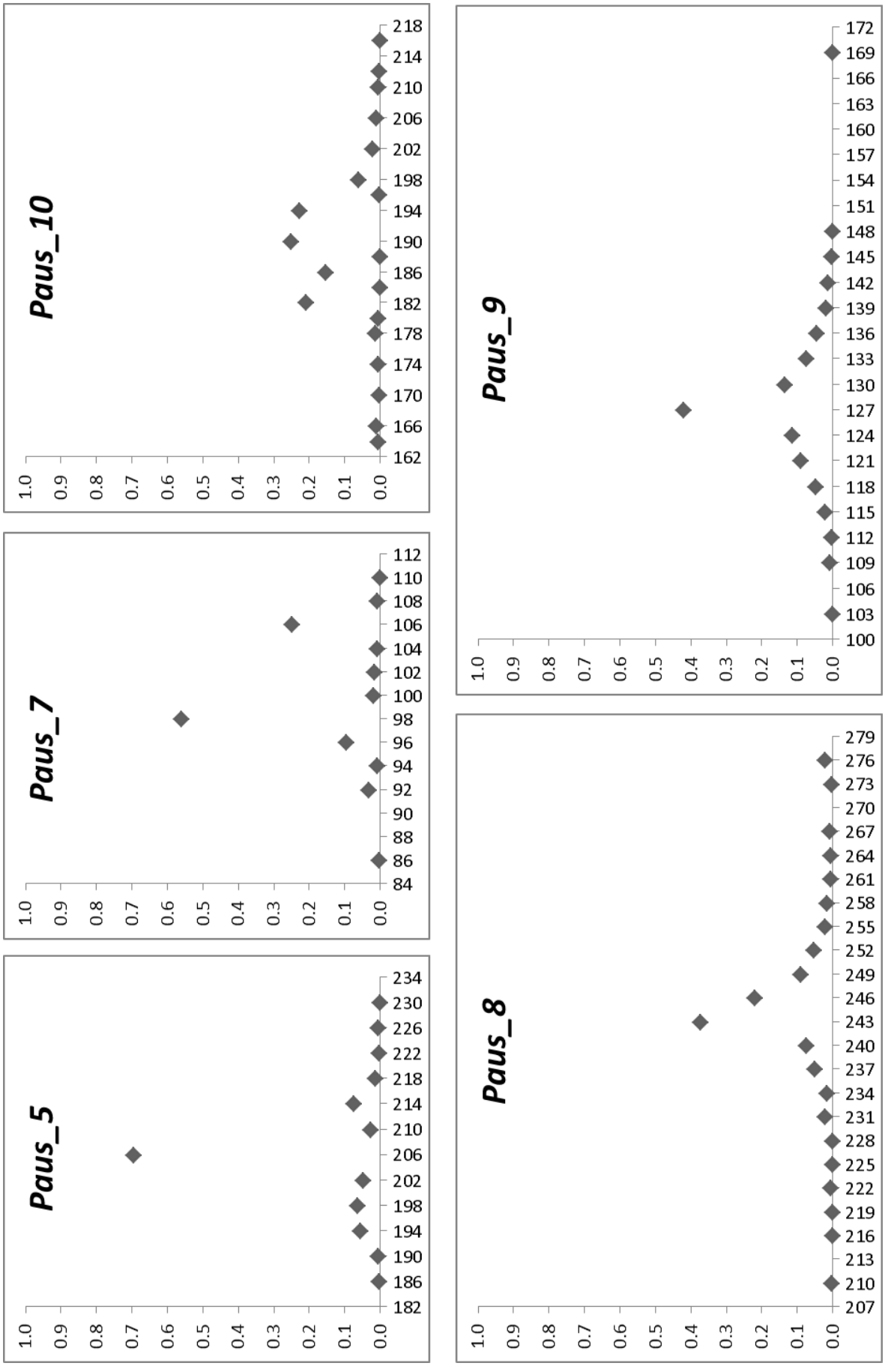




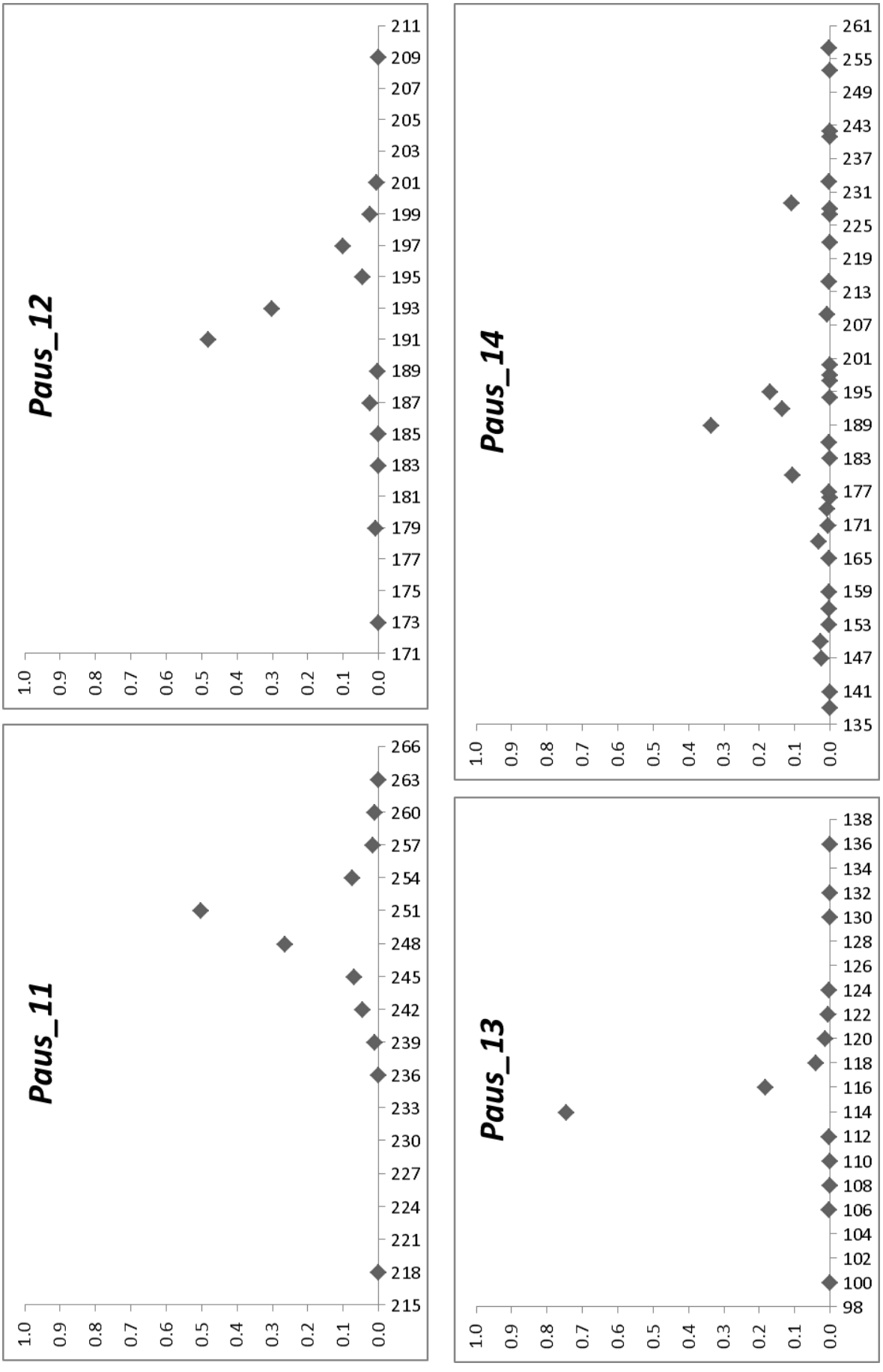
Supplementary Figure 2.



Supplementary Figure 2 continued.



Supplementary Figure 2 continued.

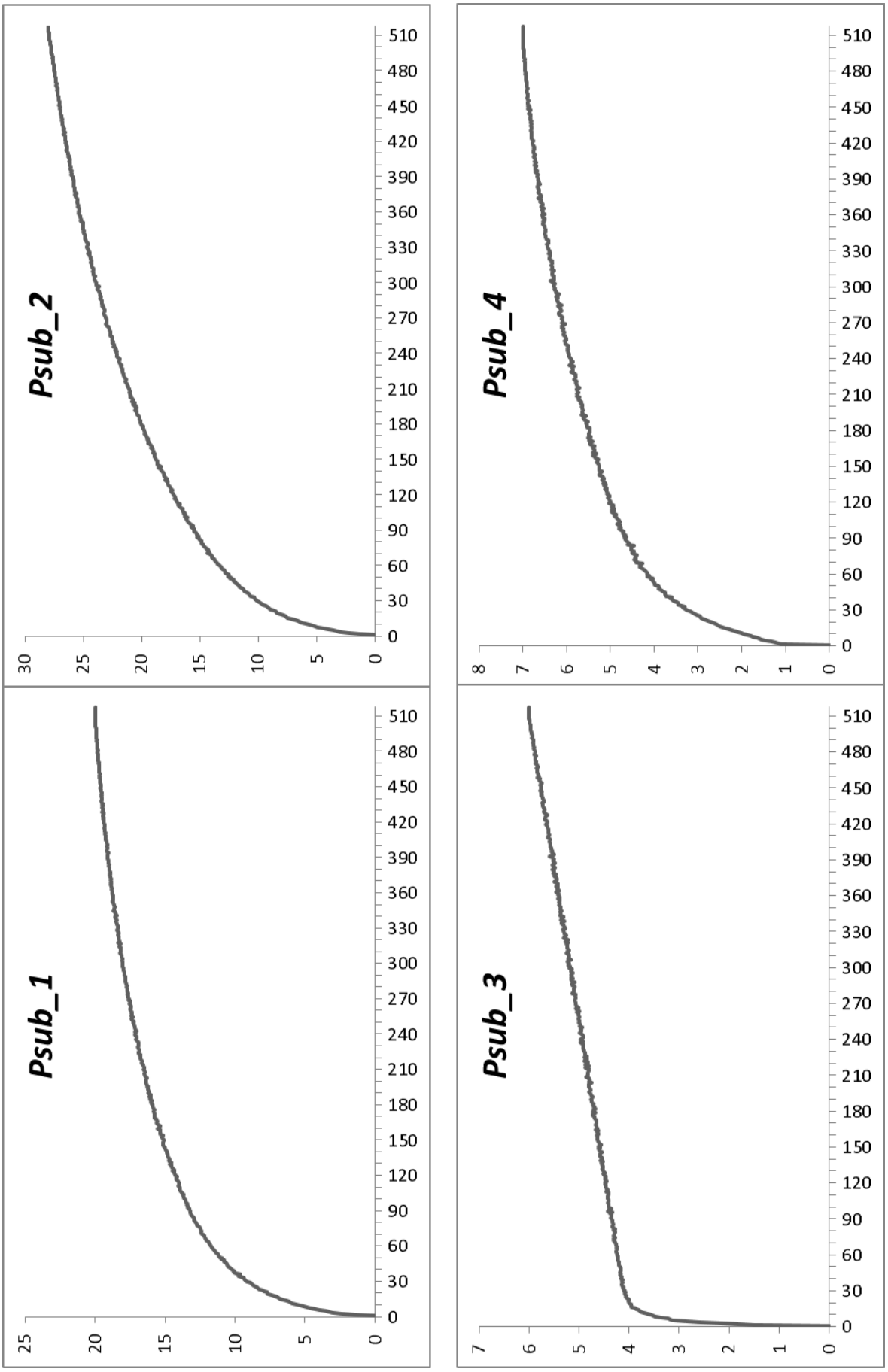


Appendix 3: Allele discovery curves

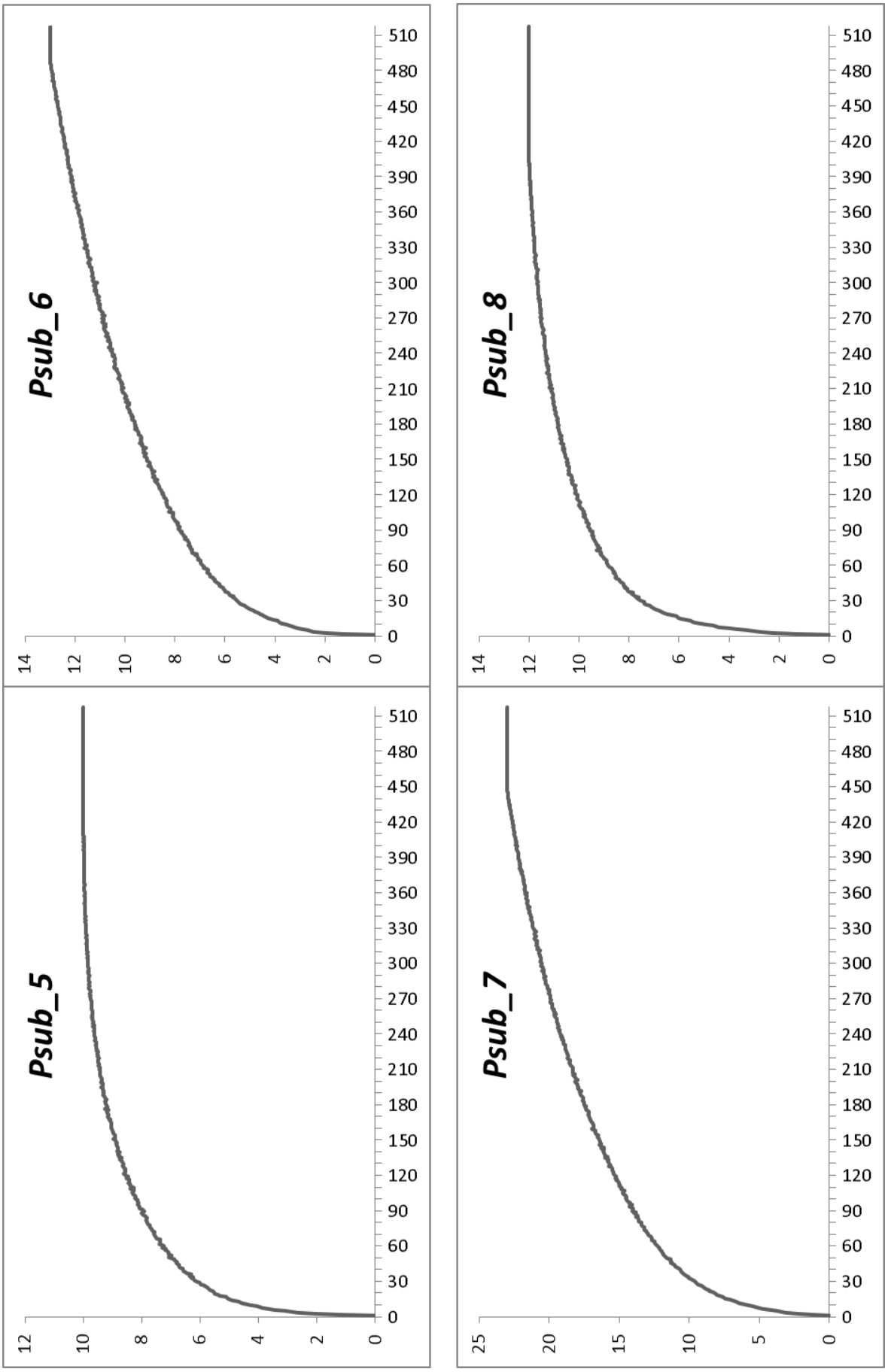
Supplementary Figure 3: Allele discovery curves for all *Paphies subtriangulata* loci generated in the R package PopGenKit (Rousset 2008). X axis = number of individuals samples, Y axis = number of alleles observed.

Supplementary Figure 4: Allele discovery curves for all *Paphies australis* loci generated in the R package PopGenKit (Rousset 2008). X axis = number of individuals samples, Y axis = number of alleles observed.

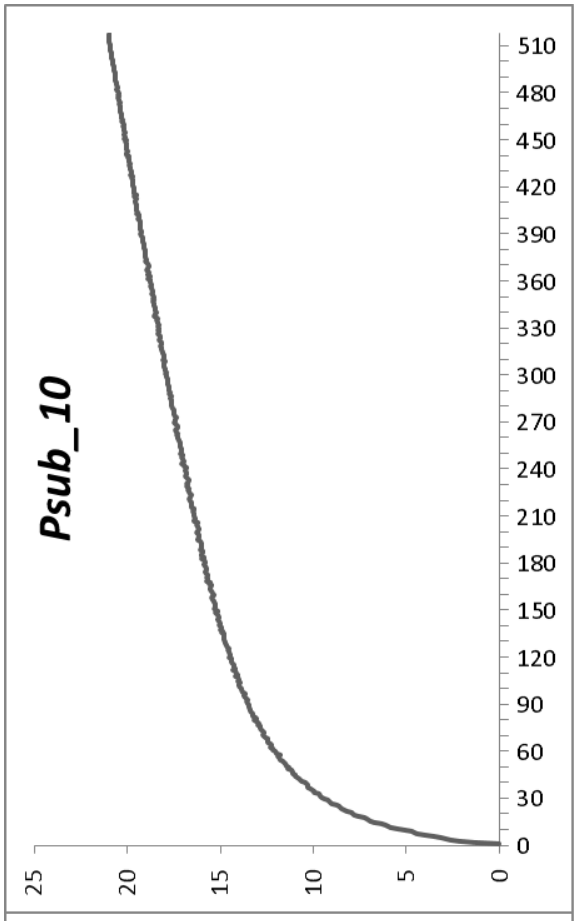
Supplementary Figure 3.

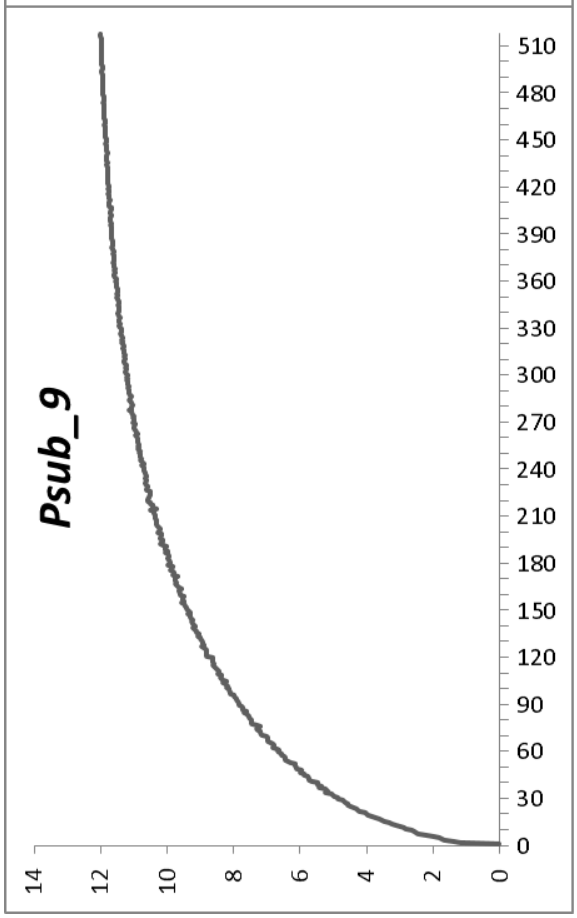


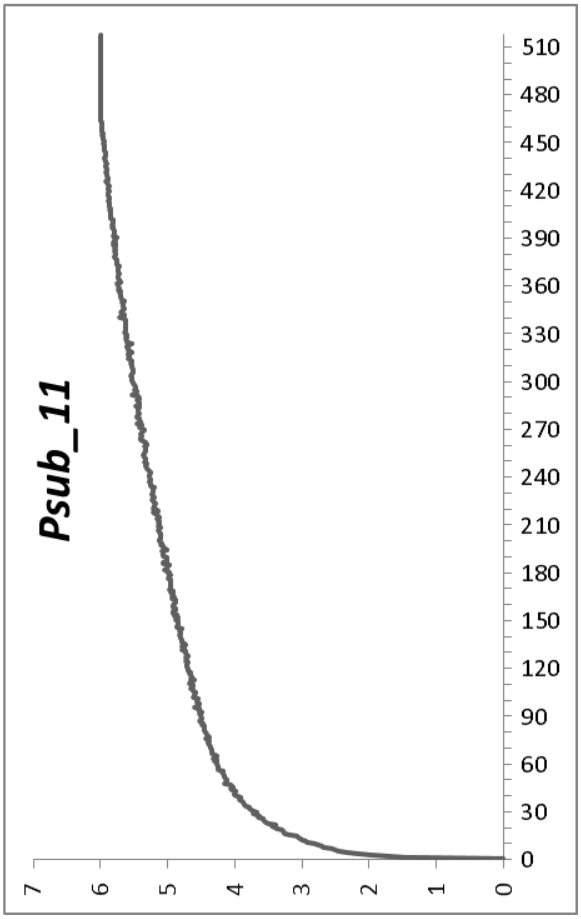
Supplementary Figure 3 continued.



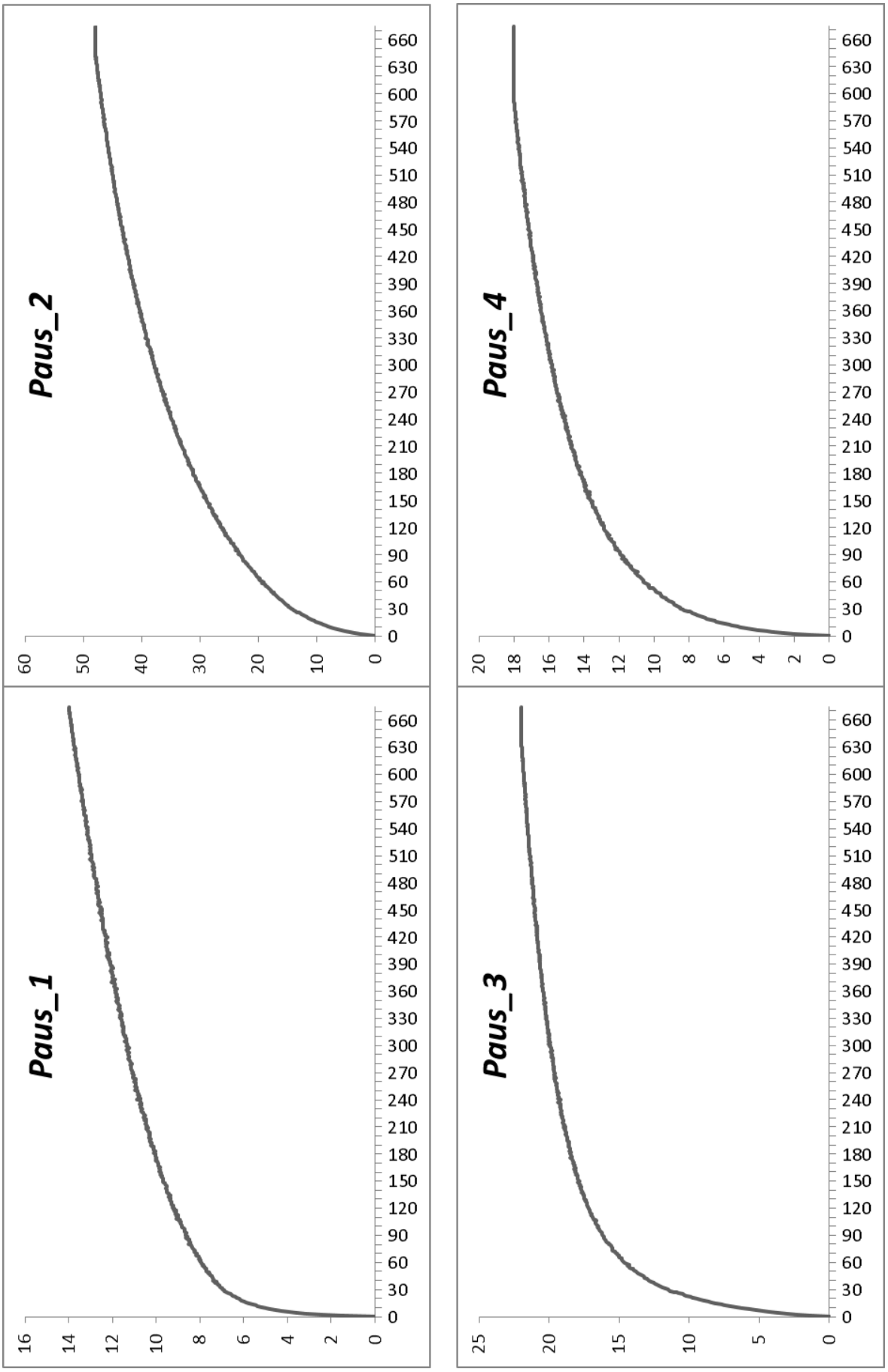
Supplementary Figure 3 continued.



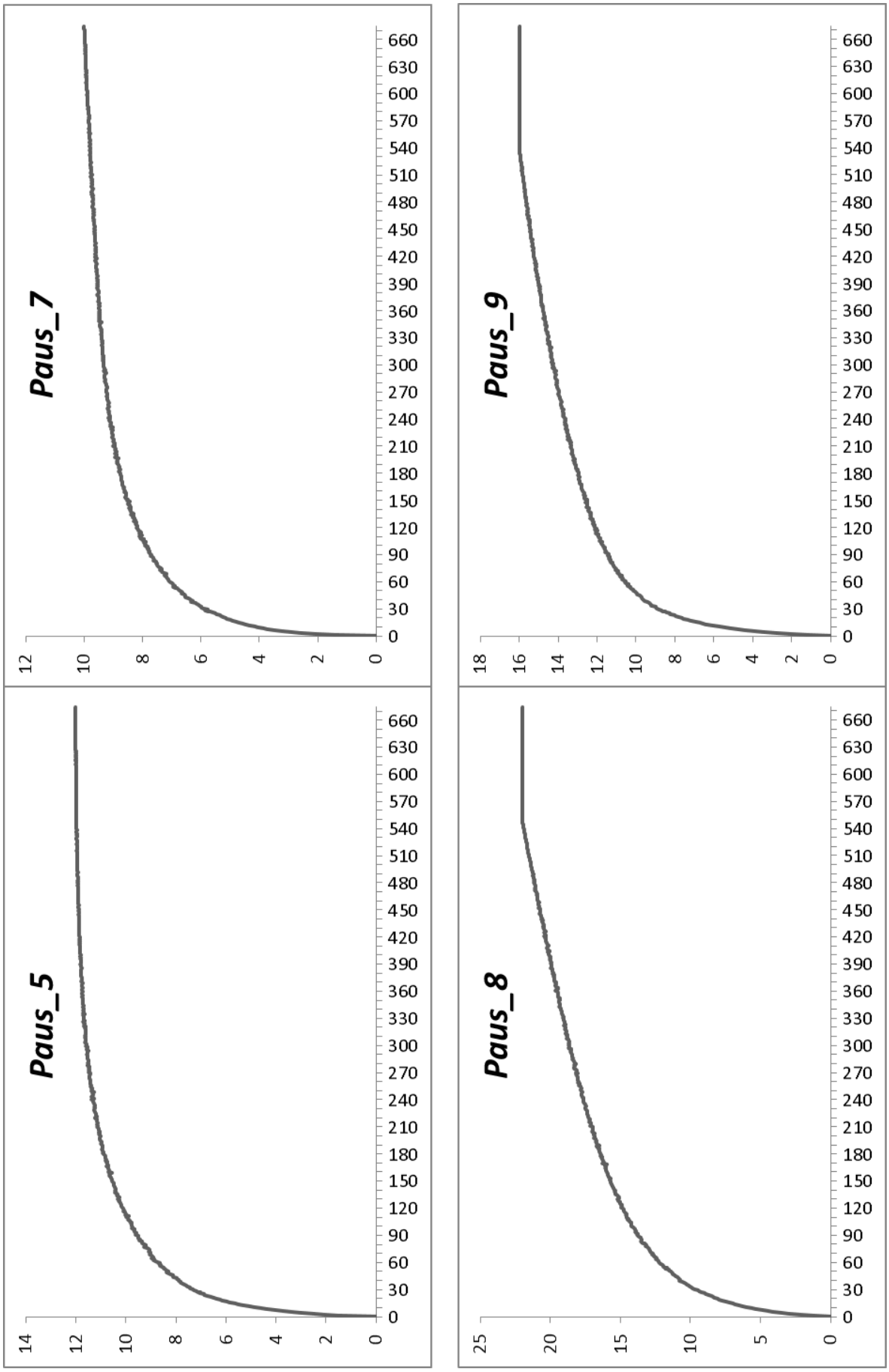




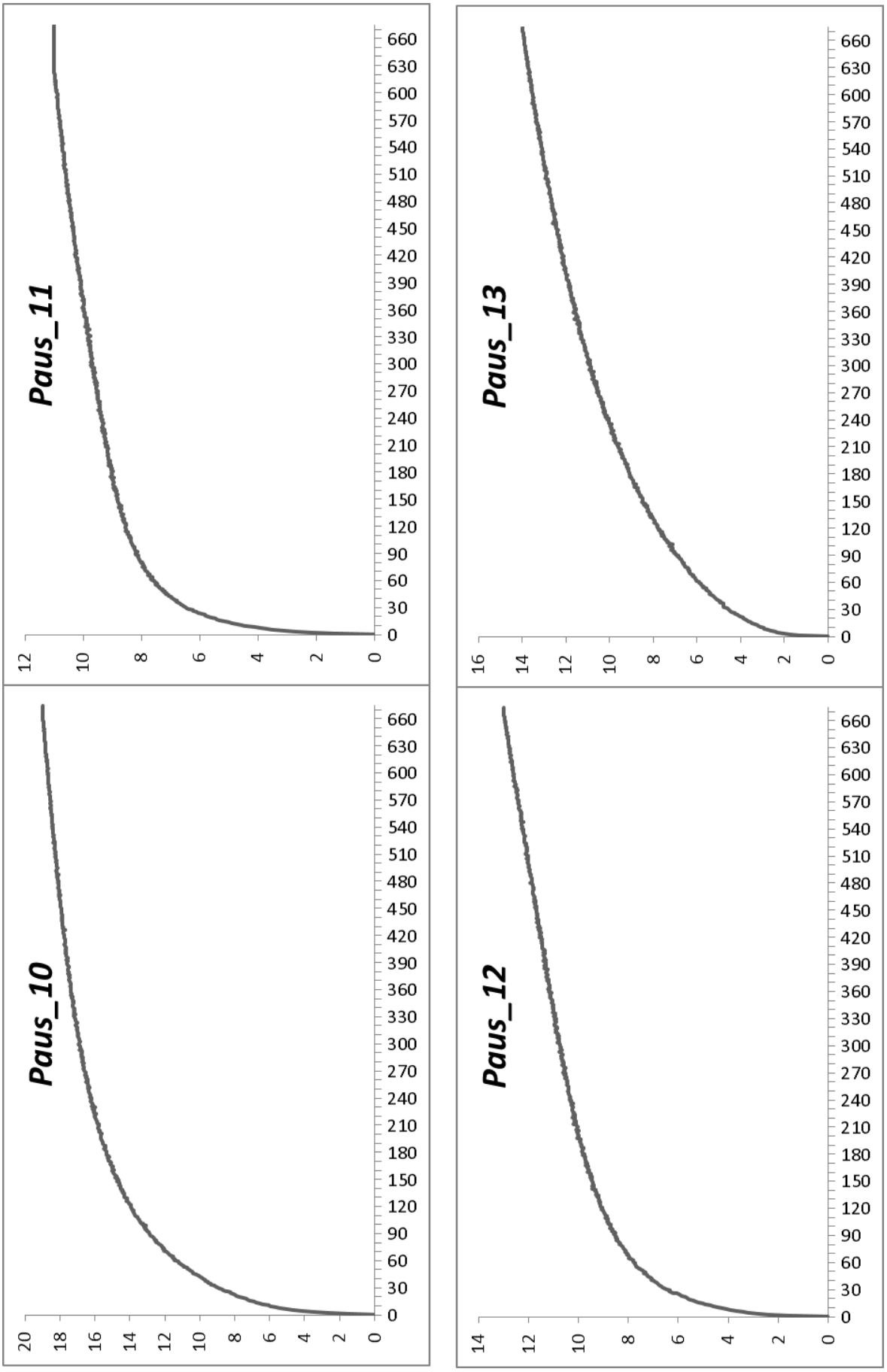
Supplementary Figure 4.



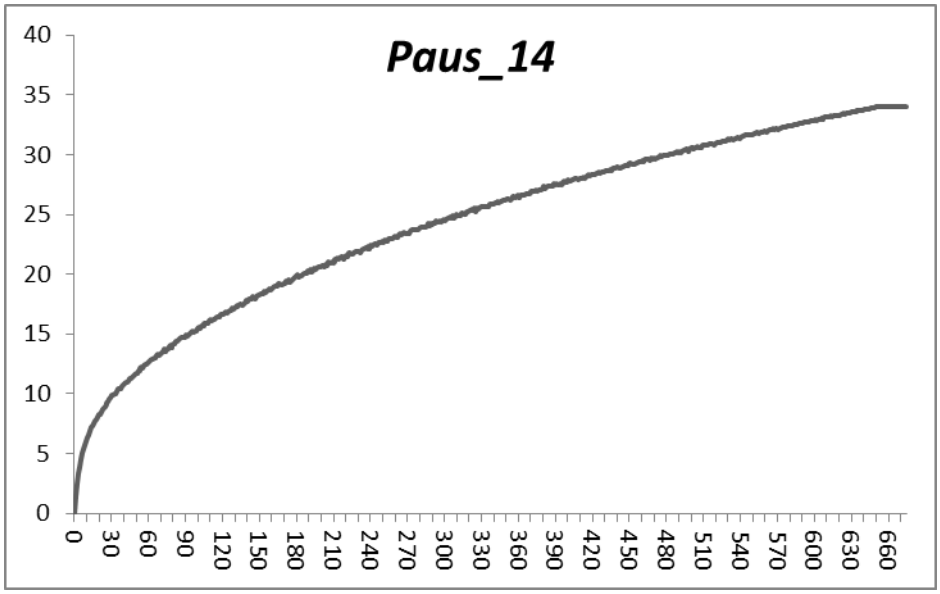
Supplementary Figure 4 continued.



Supplementary Figure 4 continued.



Supplementary Figure 4 continued.



Appendix 4: GLM and BEST input data

Supplementary Table 3: Raw input data used in generalised linear model (GLM) and biological environmental stepwise (BEST) analyses for *Paphies subtriangulata*, as implemented in Statistica v.10 (StatSoft Inc.) and PRIMER v.6 (Clarke & Gorley 2006). F_{ST} and ϕ'_{ST} were derived from the average of pairwise multilocus F_{ST} or ϕ'_{ST} estimates for each location. Latitude, longitude and geo_dist were derived from Google Earth. Environmental variables were derived from the New Zealand Marine Environment Classification system (New Zealand Ministry for the Environment 2005).

Supplementary Table 4: Raw input data used in generalised linear model (GLM) and biological environmental stepwise (BEST) analyses for *Paphies australis*, as implemented in Statistica v.10 (StatSoft Inc.) and PRIMER v.6 (Clarke & Gorley 2006). F_{ST} and ϕ'_{ST} were derived from the average of pairwise multilocus F_{ST} or ϕ'_{ST} estimates for each population. Latitude, longitude and geo_dist were derived from Google Earth. Environmental variables were derived from the New Zealand Marine Environment Classification system[a] (New Zealand Ministry for the Environment 2005) and the New Zealand Estuarine Environment Classification system[b] (Hume *et al.* 2007).

Supplementary Table 3.

Location	F_{ST}	Φ'_{ST}	Latitude	Longitude	geo_dist (km)	orb_v _mean (m/s)	rad_ mean (Wm^{-2})	SST anamp (°C)	SSTgrad ($°C\ km^{-1}$)	SSTwint (°C)	tidal (m/s)	bed_ slope ($0.01m^{-1}$)
RUA	0.040	0.064	-35.908	174.461	7241	424.65	16	2.718	0.012	14.274	0.032	0.131
PAP	0.027	0.046	-37.699	176.295	7197	260.98	15	2.999	0.016	13.938	0.000	0.371
WMR	0.029	0.047	-39.815	176.995	5695	589.58	14	2.830	0.017	12.917	0.000	0.126
MAR	0.026	0.044	-41.724	174.200	4930	793.13	14	2.373	0.015	10.808	0.260	0.262
PKR	0.021	0.036	-40.985	174.951	4873	605.78	14	2.601	0.021	12.564	0.332	0.315
COL	0.027	0.035	-40.678	172.687	5364	1011.40	14	3.080	0.017	12.326	0.103	0.233
OAK	0.026	0.043	-39.115	173.946	5061	705.92	14	2.701	0.017	13.410	0.099	0.379
KAK	0.026	0.047	-37.006	174.597	5667	0.00	15	3.405	0.009	13.931	0.000	0.031
WPK	0.021	0.029	-35.040	173.168	6439	153.41	16	2.503	0.011	14.641	0.090	0.455
CHA	0.121	0.230	-43.801	183.652	9465	2000.70	11	2.021	0.003	10.621	0.000	0.350

Supplementary Table 4.

Location	F_{ST}	Φ'_{ST}	Latitude	Longitude	geo_dist (km)	orb_v_mean (m/s) [a]	rad_mean (Wm^{-2}) [a]	SST anamp (°C) [a]	SSTgrad (°C km^{-1}) [a]	SSTwint (°C) [a]	tidal (m/s) [a]	bed_slope ($0.01m^{-1}$) [a]
RAG	0.011	0.025	-37.819	174.827	8581	0.00	14.981	3.461	0.013	13.523	0.000	0.003
HUI	0.010	0.017	-37.006	174.573	8798	0.00	15.293	3.382	0.010	13.870	0.000	0.022
TAP	0.009	0.014	-34.435	172.714	10141	394.00	15.902	2.147	0.025	14.747	0.000	0.376
WAI	0.013	0.029	-36.543	174.712	12043	438.02	15.524	3.695	0.024	13.678	0.057	0.328
TAU	0.014	0.031	-37.659	176.132	12195	429.35	15.095	3.042	0.016	13.943	0.733	0.002
NAP	0.006	0.007	-39.483	176.888	9644	1025.50	14.459	3.334	0.009	12.489	0.063	0.002
PET	0.006	0.013	-41.226	174.864	7300	0.00	14.158	2.286	0.039	12.462	0.010	0.043
HAK	0.012	0.023	-41.298	174.114	7279	0.00	14.081	2.167	0.078	11.720	0.022	0.002
LYT	0.006	0.011	-43.638	172.750	9234	1174.60	13.051	3.806	0.023	10.307	0.000	0.106
BLU	0.008	0.017	-46.570	168.491	12834	381.70	11.240	2.323	0.010	10.243	1.048	0.002
DBT	0.009	0.022	-45.282	166.905	12201	2168.30	11.831	1.613	0.007	11.797	0.078	0.207
OKU	0.005	0.006	-43.894	168.920	10626	706.83	12.894	2.063	0.018	11.869	0.122	0.381
KAR	0.010	0.018	-41.257	172.105	8066	0.00	13.752	2.852	0.018	12.380	0.000	0.194

Supplementary Table 4 continued.

Location	spring_TR (m) [b]	spring_TP (m) [b]	mouth_width (m) [b]	catch_area (km^2) [b]	shore_length (m) [b]	inter tidal (%) [b]	mean_depth (m) [b]	area_LT (m^2) [b]	area_HT (m^2) [b]	tot_vol_SHT (m^3) [b]	river_discharge (cumecs) [b]
RAG	2.921	60944181	437	522	142267	69.00	0.868	9874636	31853664	27652903	26.859
HUI	2.811	710146881	2294	1022	459539	61.80	6.061	139660198	365602613	2215803524	42.521
TAP	2.312	557184	543	13	5191	0.94	3.292	239859	242135	797044	0.494
WAI	2.467	1659432	161	37	11722	64.49	2.382	352533	992771	2364498	1.878
TAU	1.716	211514717	1128	1299	355135	77.01	2.122	46081061	200439588	425300509	75.904
NAP	1.471	3853629	125	143	33792	9.18	0.205	2493704	2745765	561571	4.453
PET	1.034	88321084	2801	638	70484	0.01	16.032	85412639	85421181	1369490185	41.396
HAK	1.270	30943063	1926	38	50581	1.15	12.013	24216083	24497808	294283650	1.842
LYT	1.804	70438844	2233	106	73987	16.15	5.718	35623897	42485268	242920351	4.075
BLU	2.222	89628434	789	99	91973	52.22	2.235	26078587	54580551	121988796	3.392
DBT	1.863	254867547	1284	963	324505	0.82	138.19	136212455	137338631	18978270538	239.491
OKU	2.130	3128239	146	514	30312	25.08	2.612	1258317	1679548	4386557	117.013
KAR	2.973	7809114	726	1307	32648	67.63	2.615	1284665	3968691	10378445	183.508

Ian Chesterton *"Doctor, why do you always show the greatest interest in the least important things?"*

The Doctor *"The least important things, sometimes, my dear boy, lead to the greatest discoveries."*

"The Space Museum", BBC, 1965