Genome-wide SNP analysis reveals patterns of population differentiation and adaptation of red rock lobster *Jasus edwardsii*

Irina Ilyushkina

A thesis submitted to Victoria University of Wellington

in fulfillment of the requirements for the degree of

Doctor of Philosophy

2018

Victoria

UNIVERSITY OF WELLINGTON

Te Whare Wānanga
o te Ūpoko o te Ika a Māui

Finally, I would like to thank my parents Vera Bondar and Anatoly Ilyushkin, and my partner Andy Biggerstaff for the endless support over the years, words of encouragement that allowed me to move forward. Also, I greatly appreciate support of all my friends and lab mates who gave me much needed advice and facilitated the work on this thesis.

Abstract

Declines in global marine finfish catches, which accounts for ~15% of the animal protein consumed by humans, has caused a 6-fold increase in total reported catch of invertebrates since 1950. This has led to the over-exploitation and decline of many marine invertebrate fisheries. The red rock lobster *(Jasus edwardsii)* fishery is New Zealand's most economically valuable inshore fishery. The current management strategy relies on the assumption that the stock is comprised of a single panmictic population. However, more recent studies have challenged the genetic homogeneity of *Jasus edwardsii* across the Tasman sea and described high levels of self-recruitment in a Stewart Island subpopulation. A disregard for the underlying genetic structure in the management of a fishery can lead to excessive removal of individuals from populations contributing to the overall genetic diversity of the stock and thus reduce the species adaptability. The ability to adapt to new environments is particularly important in the context of global climate change and can significantly affect the long-term sustainability of the stock. Thus, the goal of this study was to identify specific patterns of genetic diversity of Jasus edwardsii population and provide an interpretation and assessment of the impact on the NZ fishery.

The first objective was to optimize and validate molecular and bioinformatic protocols of Single Nucleotide Polymorphism (SNP) discovery for the red rock lobster *Jasus edwardsii.* The double digest restriction-site associated DNA (ddRADseq) protocol was optimized for the relatively large red rock lobster genome, which also has a high paralog content. The impact of bioinformatic processing on the population genetic inferences was then assessed by testing three different SNP discovery pipelines with the Rad-loci pipeline producing the most optimal marker discovery rate with a low level of missing data and a low SNP error rate. An analysis of technical replicates confirmed the reproducibility of both the molecular and bioinformatic protocols and also the validated the data generation process suitable for population genetic analyses.

The second objective of my thesis was to investigate the genetic structure and population connectivity of adult red rock lobsters. The SNPs discovered were characterised as selectively neutral or under divergent selection (outlier) and both types of markers were analysed using Bayesian model-based clustering (STRUCTURE), non model-based multivariate analysis (Discriminant Analysis of Principal Components (DAPC)) and F-statistics. A lack of population differentiation using neutral genetic markers indicated a high

level of gene flow and connectivity between populations. In contrast, there was evidence for selective pressure as a result of the analysis of outlier markers. Three main regions were identified: North-East NZ, North-West NZ and South NZ sub-populations, as part of a larger NZ metapopulation (F_{ST} ranged from 0.025 to 0.049, P < 0.001). The results of this study suggested that high levels of gene flow and connectivity are counteracted to some extent by the local selection that promotes the survival and reproduction of locally adapted genotypes. However, the strength of this selective pressure still permits low levels of survival and reproduction of non-optimal genotypes causing allele frequency homogenisation of the new generation of lobsters.

The third objective was to investigate the levels of connectivity and adaptive divergence of the red rock lobster pueruli/juvenile lobsters for comparison with pattern of divergence of adult lobster in order to investigate the mechanisms of population structure formation. A suite of Bayesian clustering, non-model multivariate analysis and F-statistics were employed in the assessment of neutral and outlier markers developed for pueruli/juveniles. Similar to adult lobsters, pueruli/juveniles were characterised by a low level of divergence of the neutral markers indicating effective larvae dispersal. Outlier markers detected population differentiation patterns likely to originate from a phenotype – environment mismatch resulting in post-settlement mortality of non-adapted genotypes. The similarity between patterns of genetic divergence of adult lobsters and late juvenile/early juveniles indicates that post-settlement mortality, driven by local environmental conditions, has most likely occurred on earlier developmental stages of *Jasus edwardsii*, which were not possible to sample in my study.

The final objective was to explore environment–genotype associations of *Jasus edwardsii*. Biological Environment Stepwise (BEST) analyses, redundancy analyses (RDA) and generalized linear modelling (GLM) consistently indicated a correlation between the annual amplitude of sea surface temperature (SST) and adaptive population divergence. In addition, an influence of spatial distribution on the patterns of adaptive population differentiation was also detected via RDA. From these results I propose a mechanism underlying the patterns of population differentiation discovered in Chapters 3 and 4: a latitudinal gradient of SST appears to be the selective force promoting the adaptive divergence of the lobster populations with local patterns of connectivity distorting the gradient and thus forming three distinct temperature adapted genotypes (North-West, North-East, and South). An environmental association analysis offered 43 candidate loci, which after alignment of transcriptome-

mapped reference catalog sequences to annotated protein databases identified a candidate gene for thermal adaptation - UDP-glycosyltransferase (UGT). UGT is a detoxification enzyme involved in the metabolization of a variety of endogenous and environmental compounds and its activity and gene expression patterns have been linked to temperature.

This study provides evidence for the local adaptations of the NZ population of *Jasus edwardsii* to SST, which together with the efficient mechanism of larval dispersal creates a system likely resilient to changes in temperature. This feature is important in the light of climate change-induced range shifts and supports the long-term sustainability of the red rock lobster fishery. The three genetically distinct regions identified coincide with existing boundaries of the management units and therefore do not require an adjustment of the current management regime.

Table of Contents

Chapter 1

General introduction

Fisheries play an essential role in the global provision of food, directly accounting for at least 15% of the animal protein consumed by humans (Muir 2012). As human populations have grown, an increased need for animal protein has led to higher demand for fish as a source of food. Despite this increased need for fish protein, global marine fisheries catches have declined in recent decades and at least 28% of the world's fish stocks are described as overexploited or depleted (Mora et al. 2009). Declining finfish catches have led to increased interest in the enhancement of invertebrate fisheries. Since 1950, the total reported catch of invertebrates has increased 6-fold from 2 to 12 million tons, with 18% of invertebrate fisheries already being fully exploited, 21% over-exploited or restrictively managed, and 13% have collapsed or closed (Anderson et al. 2011). In order to prevent further declines in these fisheries and ensure sustainable development of invertebrate fisheries, there is an urgent need for a robust scientific basis upon which to base management recommendations. Global fishing industries need more rigorous scientific assessment for precautionary management and sustainable exploitation to ensure the long-term resilience of fished populations and ocean ecosystems.

Processes affecting population structure

Conservation and sustainable management of harvested species require an understanding of the levels of population connectivity and patterns of differentiation as these processes are vital for maintaining genetic diversity and demographical stability generated by mutation (Kough et al. 2013; Burgess et al. 2014). Most marine invertebrates have a bi-partite life cycle that includes relatively sedentary adults and larvae performing the function of dispersal (Strathmann et al. 2002; Kritzer et al. 2010). Patterns of larval dispersal can influence the population structure and connectivity to a great extent. A number of factors affect the dispersal potential of the larvae: the length of the pelagic larval duration (Weersing et al. 2009), larval behaviour (Jeffs et al. 2005) and currents flow patterns (White et al. 2010; Chiswell et al. 2011). However, not only dispersal, but also settlement and survival to reach reproductive maturity affect levels of population differentiation and effective migration between populations (Pineda et al. 2009). Successful settlement is highly dependent on larval experience (Shima et al. 2010), substrate availability (Pineda et al. 2009) and any potential phenotype – environment mismatch that can lead to post-settlement mortality and reduce gene flow between populations (Marshall et al. 2010). Larval dispersal between genetically

diverse populations promotes the introduction of novel alleles into the genetic pool thus increasing resilience to changing environments and ensuring fisheries productivity (Aguirre et al. 2012). Thus, knowledge about the processes influencing survival to reproductive maturity from early life stages of development into adulthood can provide an invaluable insight into the mechanisms of formation and maintenance of patterns of population connectivity.

There are four main processes influencing genetic variability in a population: mutation, genetic drift, gene flow and selection (Hellberg et al. 2002; Hauser et al. 2008). Though mutation occurs at the level of an individual, at the population level they become easier identifiable over a significant time period due to effects of the genetic drift (thousands of years). Genetic drift can be defined as a change in allele frequencies in reproductively isolated populations due to a random sampling of the individuals in a population and continuously accumulates over each generation (Liggins et al. 2013). Genetic drift is dependent on population size and can result in a significant loss of genetic variation in small populations over a short period of time (Hellberg et al. 2002). Gene flow represents migration between populations, which results in the transfer of new genes/alleles between genetically distinct populations. In populations, where selection and genetic drift are small, gene flow can eventually homogenise allele frequencies (Hellberg et al. 2002; Masel 2011). Genetic drift, as well as genetic flow, are processes leading to neutral genetic structure, contrasting with selection. Genes or genomic regions under selection can be subject to either divergent or balancing selection, which are described by increased or decreased levels of differentiation (correspondingly) compared to neutrally evolving genetic loci (Nosil et al. 2009).

One of the first approaches employed to determine genetic diversity and relationships between and within different species was allozyme technology. Allozymes are different protein variants coded by different alleles at the same locus and can be separated by flat-bed electrophoresis on either a starch or cellulose acetate medium. However, the cost-effectiveness of this method comes at a price: only a small fraction of the genome is surveyed and the regions studied are coding and thus tend to be more conservative, and less polymorphic compared to other molecular markers (Seeb et al. 1998; Pinho et al. 2008). Another molecular marker widely used in population genetics studies are mtDNA sequences. The main disadvantage of this method is that the mtDNA genome is maternally inherited as a single locus and reflects only matrilineal genetic history which is may not be representative of biparental genetic variability (Seeb et al. 1998; Dieringer et al. 2003; Pinho et al. 2008;

Castoe et al. 2010; Mesnick et al. 2011; Groeneveld et al. 2012). The limitations of two previously described methods has led to the development of methods analysing microsatellites - short tandem repeat sequences of nuclear DNA. Microsatellites established themselves as a molecular marker of choice in population genetics studies due to their high variability and ability to resolve population structure (Hauser et al. 2002; Groth et al. 2009; Kennington et al. 2012). However, this approach is prone to several drawbacks: high incidence of homoplasy, null alleles and low genotyping throughput (Narum et al. 2008). Certain species exhibit lower microsatellite occurrence throughout the genome which restricts use of this type of marker for these species (Morin et al. 2004). By contrast, single nucleotide polymorphisms (SNPs) are the most abundant type of variation in DNA sequence among individuals in a population and are widespread throughout the whole genome. SNPs occur in coding and non-coding regions, and fit better into simple mutation models than microsatellites (Bossart et al. 1998; Vignal et al. 2002; Morin et al. 2004). The main drawback of SNP as a genetic marker is that they are usually biallelic, thus more SNPs are required compared to microsatellites, where 5 to 20 alleles can be identified (Anderson et al. 2006). Multiple studies using both markers have shown that one microsatellite marker roughly corresponds 10-15 SNPs (dependent on species studied) (Narum et al. 2008; Glover et al. 2010; Hess et al. 2011; Mesnick et al. 2011). The high abundance of SNPs and potential for automation allow them not only to reach the same resolution level as microsatellites but exceed it (Olsen et al. 2011). Unlike microsatellites, SNPs have been shown to occur in both coding and non-coding regions (Vignal, Milan et al. 2002, Morin, Martien et al. 2009). Therefore, another advantage of SNPs is their ability to detect not only genetic structure (using putatively neutral SNPs) but also to investigate adaptive variation in populations (using SNPs putatively under selection) (Nielsen et al. 2009; Bradbury et al. 2010).

Population structure and environment

Marine environments have long been considered "open" systems, apparently lacking barriers to dispersal, which was assumed to support high levels of gene flow and homogenise genotypes over large geographical distances (Levin 2006). This view has been challenged by more recent studies demonstrating population differentiation and self-recruitment in a number of marine species (Almany et al. 2007; Puebla et al. 2012; Benestan et al. 2015; Pocwierz-Kotus et al. 2015). Many studies have demonstrated that patterns of population divergence

may establish despite pronounced effects of gene flow and they have attributed the mechanism of such a pattern formation to the selective pressure of environmental gradients (Limborg et al. 2012; DeFaveri et al. 2013; Orsini et al. 2013; D'Aloia et al. 2015). Selective mortality of dispersers to new habitats has been associated with a decrease in fitness caused by the adaptations to migrant's original environment and the environment it is attempting to settle in (Marshall et al. 2010). This process has been termed the "phenotype – environment mismatch" and can have important consequences for conservation and sustainable management (Selkoe et al. 2008; Iacchei et al. 2013; Selkoe et al. 2016a).

Recent rapid developments in high-throughput genetic marker discovery approaches, as well as advances in computational biology have made it possible to uncover correlations between the genetic structure of populations with physical and environmental factors creating a new scientific field of landscape genetics (Manel et al. 2003). However, this approach has mostly been applied to terrestrial species (Manel et al. 2003; Storfer et al. 2010). Although marine environments are more susceptible to change and support species dispersal over much greater distances than in terrestrial environments, the landscape genetics approach can be useful when applied to many marine species (Riginos et al. 2001; Galindo et al. 2006; Selkoe et al. 2008; Selkoe et al. 2010; Amaral et al. 2012; Coscia et al. 2012; Limborg et al. 2012). Thus a sub-discipline of landscape genetics - seascape genetics has been established (Selkoe et al. 2008). Microsatellite markers historically have been used for seascape genetics studies, however they are generally neutral markers and therefore can only be used only for inferring environmental barriers to the gene flow that are associated divergence of local populations via genetic drift (Slatkin 1987). Despite that, seascape genetics studies employing microsatellites have been able to identify biological or environmental barriers which consistently reduce the gene flow and due to effects of genetic drift establish patterns of population differentiation over a number of successive generations (Wei et al. 2013; Constable 2014; Hannan 2014; Silva et al. 2015; Benestan et al. 2016). The divergence of markers under selection across habitats can be linked to environmental factors which can provide insights into the genotype-environment interactions (Schmidt et al. 2008). Markers under selection in diversifying environments may lead to the identification of genes associated with a physiological trait (candidate genes), variants of which can increase or decrease fitness of an individual in a specific environment (Nosil et al. 2009). Outlier SNP panels have been used to identify correlations between genetic variation and environmental gradients in a number of marine species with high levels of gene flow. Temperature and

salinity clines between the North Sea and the North Atlantic have been significantly correlated with nine transcriptome-derived outlier SNPs markers in Atlantic herring (Limborg et al. 2012). Adaptations to the local environment have been demonstrated in European hake using outlier SNPs associated with both sea surface temperature and salinity (Milano et al. 2014). Methods of seascape genetics detected local adaptations to differences in salinity, oxygen and temperature between the North Sea and the Baltic Sea in Atlantic cod (Berg et al. 2015).

Jasus edwardsii biology

The red rock lobster *Jasus edwardsii* (Hutton 1875) (Palinuridae) inhabits rocky reefs, between 1 m and 200 m depth, along the coastlines of southern Australia, Tasmania and New Zealand. In New Zealand, its range extends from Three Kings Islands in the north (34° S, 172° E) to the Auckland Islands in the south (51° S, 166° E), and to the Chatham Islands in the east (44° S, 176° W) (Kensler 1967). *Jasus edwardsii* usually shelters by day within holes and crevices on the reef. The larger juveniles and adults are gregarious, with occasionally up to 105 individuals cohabiting in the same shelter (MacDiarmid 1994). The lobsters are keystone predators of sub-tidal reef systems, feed on urchins, crabs and other crustaceans, as well as gastropods, chitons and bivalves, as well as scavenge dead material (Fielder 1965). Immature lobsters have been shown to undertake regular chaotic migration around southern New Zealand (McKoy 1983), which are believed to maintain homogeneity of the population. Seasonal inshore-offshore migrations in *Jasus edwardsii* have been attributed to the molting and reproduction periods, but they generally do not exceed 5 km (Booth 1997).

Moulting occurs once or twice a year, in spring and autumn, with mature females only moulting in autumn to winter and mature males generally moulting in spring (Annala et al. 1988). The increase in carapace length at moulting ranges from 2 to 14 mm (McKoy et al. 1981; Annala et al. 1988). Adult lobsters can reach sizes of over 160 mm carapace length (CL) for females and 200 mm CL for males (MacDiarmid 1991).

The size at onset of maturity is highly variable between locations and generally temperature dependent, ranging from 72 to 171 mm CL (Annala et al. 1980). Females spawn once a year shortly after moulting. Fecundity of mature females varies greatly and depends on the size of the individual female; between 37 to 540 thousand eggs can be produced during the mating period (Annala et al. 1987). Fertilization is external, with the male depositing a fast

dissolving spermatophore on the female's sternum. The spermatophore disintegrates easily which causes the female to extrude her eggs for immediate fertilisation, thus negating the possibility of sperm collection from several males as is possible in other lobster species (Lipcius 1985). Non-fertilized eggs are dislodged during brood grooming and ventilation (Butler 1999). The egg-bearing period usually occurs from April to November (MacDiarmid 1989a).

Jasus edwardsii has been shown to have one of the longest-lived larval durations of all marine species – up to 24 months (Booth et al. 1994), which allows large-scale dispersal and significant gene flow between populations. The *Jasus edwardsii* life cycle is represented in **Fig.1.1.**

Fig. 1.1. *Jasus edwardsii* life cycle. Adapted from Ministry of Fisheries "Rock Lobsters – spiny wanderers".
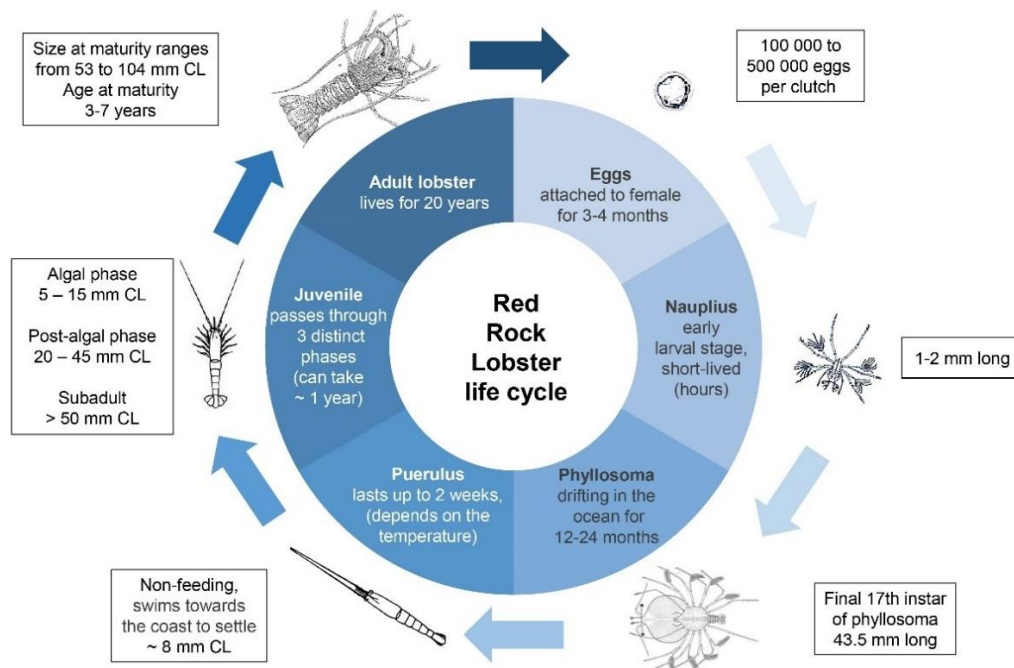
Lobster larval development starts with naupliosoma stage, which lasts only several hours. The naupliosoma molts into a phyllosoma, which is a transparent planktonic zoea. The phyllosoma drifts with offshore currents for 12-24 months before it moves toward the coast. In the inshore waters, the phyllosoma transforms into a puerulus. The puerulus molts into a juvenile a few days after settlement (Booth et al. 1994). Juveniles resemble adult lobsters in body anatomy but are much smaller and transparent.

Fig.1.2. Movements of the red rock lobster during the life cycle. Adapted from (Lipcius et al. 2000)
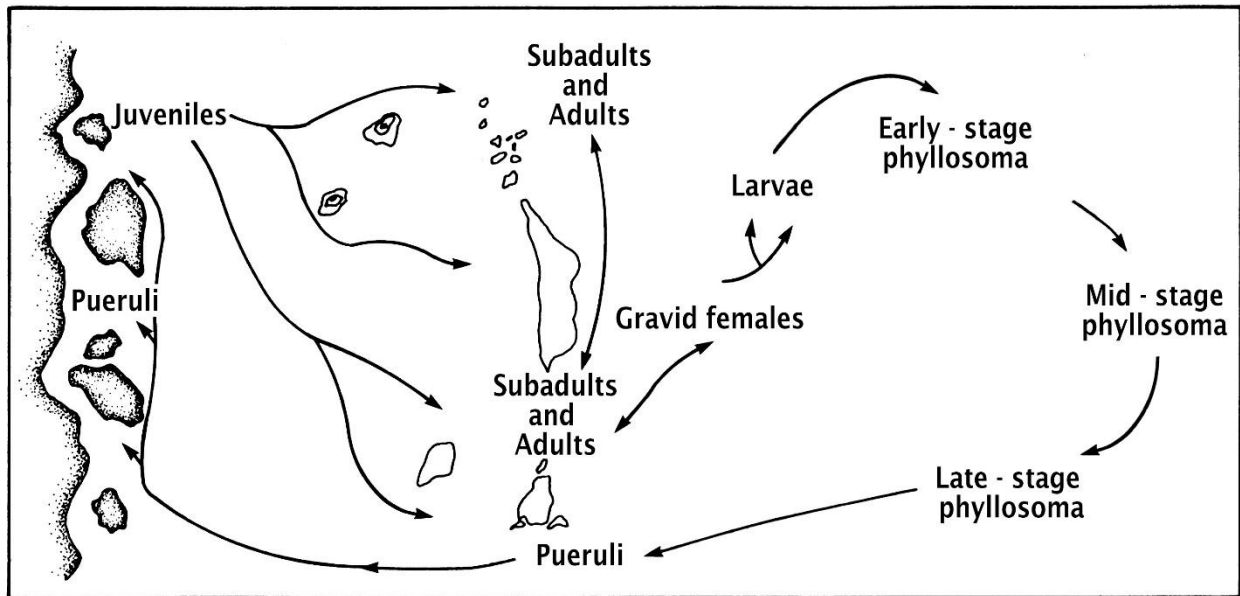
Drifting with currents has been suggested to be the main driver of the early- and mid-stage phyllosoma distribution, however, it does not explain the distribution of the late-stage phyllosomas and pueruli (Chiswell et al. 2005; Chiswell et al. 2008). Pueruli are active swimmers (Jeffs et al. 2001), so once they reach the continental shelf their movements are deliberate, and dispersal is affected by numerous factors: underwater sound, water chemistry, eddychemotaxis, salinity, magnetic fields, and specific behaviour (Cobb 1997; Jeffs et al. 2005). The movements of lobsters during the life cycle are depicted in **Fig.1.2**. The main settlement season in *Jasus edwardsii* larvae occurs in winter and early spring (Booth et al. 1986) when pueruli reach 11.5-11.7 mm CL (Booth 1979). One year after settlement juvenile lobsters reach 31-38 mm CL (McKoy et al. 1981).

Jasus edwardsii fishing industry and stock management

The *Jasus edwardsii* fishery is currently the most economically valuable inshore fishery in New Zealand worth over NZ$230 million to the NZ economy (NZ 2016). New Zealand became one of the first countries to implement a Quota Management System (QMS) in 1986, where the management approach is based on individual tradable quotas (ITQs) for fishers. The QMS is widely regarded as one of the leading fish management tools in the world (Lock

et al. 2007). Each of nine fished quota management areas CRA1-9 (**Fig.1.3**) (CRA 10 is defined administratively and has no commercial landings) was allocated a Total Allowable Commercial Catch (TACC), which is set annually (Booth et al. 2000). Fishing levels are set using the concept of the maximum sustainable yield (MSY), which is represented by the B_{MSY} reference point - the largest average annual catch that can be taken over time without reducing the stock's productive potential (Annala 1996). The *Jasus edwardsii* stock subdivision was determined in 1992 based on several different stock definition approaches. Ihssen et al's. (1981) stock structure definition was used to define the sub-stock term in case of the red rock lobster fishery: "an intraspecific group of randomly mating individuals, with temporal or spatial integrity". Several techniques are available for sub-stock structure assessment, many of which have been applied to the red rock lobsters: genetic differences, morphological differences between pueruli and adults, dispersal patterns, landings and catch per unit effort (CPUE) trends, larval distribution and parasite infestations.

Several early studies were published on genetic differences in *Jasus edwardsii* prior to the introduction of the QMS. Smith et al. (1980) were unable to determine any subpopulation structure based on an assessment of 33 enzyme gene loci (Smith et al. 1980). Booth et al. (1990) studied allele frequencies in the lactate dehydrogenase (*Ldh*) locus and discovered no significant differences between animals from nine sites around New Zealand. Furthermore, a study by Ovenden et al. (1992) using restriction endonuclease analysis of mitochondrial DNA (mtDNA) did not reveal any population substructure. Based on those results, the population was assumed to be comprised of one thoroughly mixed stock, which is the result of a long larval life.

In 2001, Bentley and Starr re-examined established QMS boundaries for the red rock lobster. These authors used a new clustering technique, which addressed the problem of spatial location of areas and eliminated groupings of non-adjacent areas. Three population characteristics were used separately for clustering: size at onset of maturity, trends in catch rates and size frequencies. As a result of this analysis these authors suggested new stock subdivision: CRA1(939, 901-908), CRA2 (909-912), CRA3 (913-918), CRA4 (940-943), CRA5 (919-925), CRA6 (926-928), CRA7 (929-930), CRA8 (931-934) and CRA9 (935-938) (Starr 2001). However, the new stock definition has not yet been implemented.

Fig.1.3. Map indicating the 43 statistical areas and ten quota management areas for *Jasus edwardsii*. Map adapted from NZ Seafood Industry Council report, 2005.
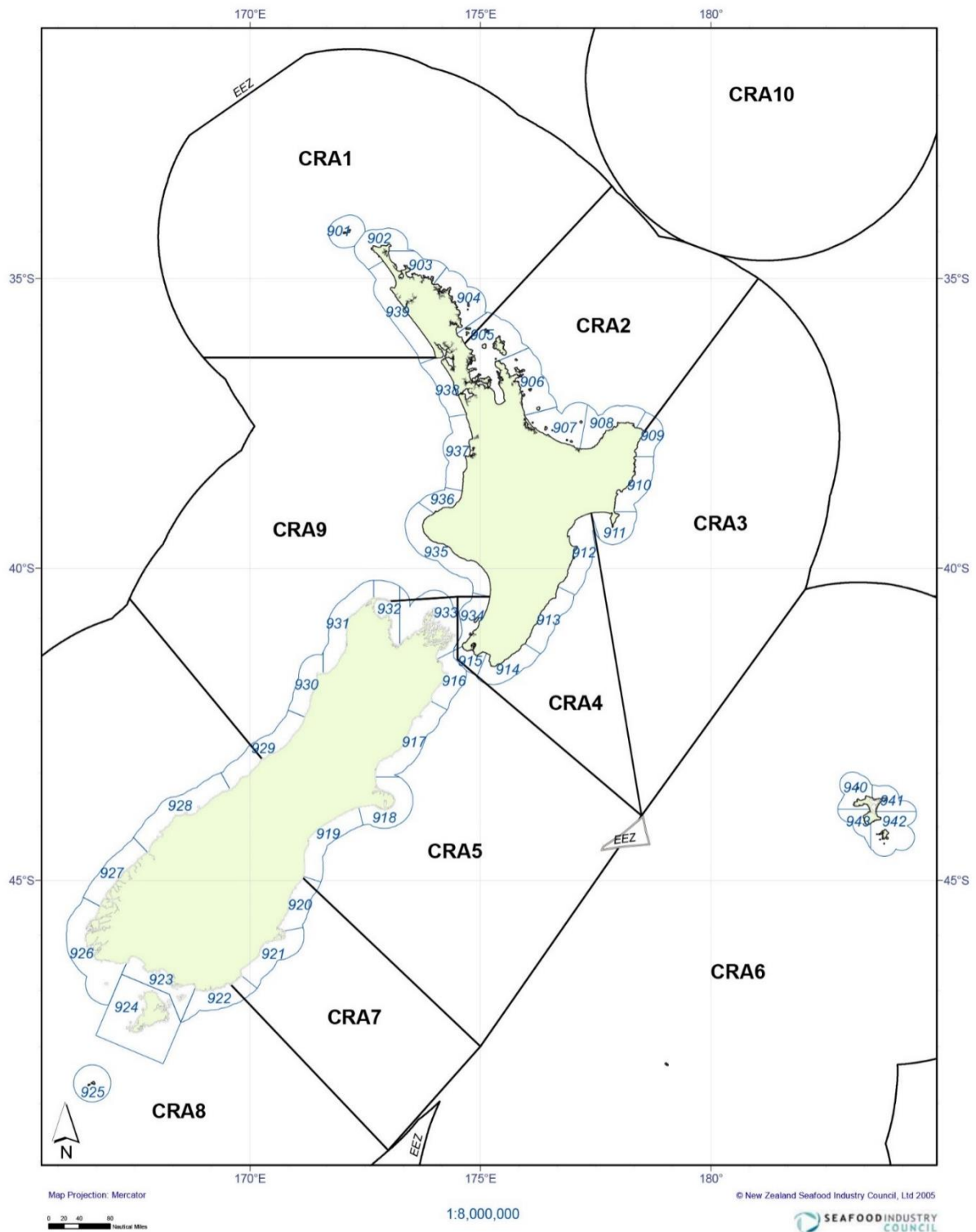
One of the main assumptions underlying the current management of the rock lobster fishery is that the stock is comprised of a single, genetically homogeneous population. However, recently Thomas & Bell (2011) described three genetically distinct subpopulations of *Jasus edwardsii* based on the assessment of seven populations from five CRA zones in New Zealand using eight microsatellite markers. The disregard for genetic structure and patterns of population connectivity can lead to reduced productivity and overexploitation of local populations (Reiss et al. 2009). Applying single stock management strategies to sub-stocks (populations) that differ in their effective population size and levels of reliance on foreign larvae for population replenishment inevitably results in the overexploitation and in extreme cases to extinction of local populations (Hutchinson 2008). Reduction of the larval input from overexploited populations can affect not only demographic stability but reduce genetic diversity (Hauser et al. 2002). A reduction in genetic diversity can negatively impact the adaptability of a species (Reiss et al. 2009) which increasingly becomes a central issue in sustainability of harvested species in the context of global climate change (Reusch et al. 2005). Failure to recognize the genetic structure of the stock already led to overexploitation and resulted in the decline or collapse of several managed fisheries (Hauser et al. 2002; Baum et al. 2003; Hutchinson 2008; Roy et al. 2012). The decline of these fisheries demonstrates the necessity of stock definition in terms of population genetics and evolutionary processes.

Summary and aims of the thesis

Despite a long history of exploitation, there are still many gaps in the knowledge of *Jasus edwardsii*. With the development of next-generation sequencing (NGS) approaches, the development of high-throughput genetic marker panels has become possible for studies on non-model species (Hudson 2008; Williams et al. 2010; Ekblom et al. 2011; Everett et al. 2011). A continuous reduction in cost for generating a large amount of DNA sequence data has resulted in new possibilities for conducting comprehensive studies of large populations (Van Tassell et al. 2008; Davey et al. 2011b; Harrison et al. 2011; Kumar et al. 2012). Sufficiently large genetic marker panels assayed in many populations from distant locations can be combined with geographical and environmental data to determine dispersal patterns in marine species (Weersing et al. 2009; Galindo et al. 2010; White et al. 2010; Selkoe et al. 2011; Liggins et al. 2013). An extensive knowledge of fished species is a powerful tool that can help protect stocks as well as benefit the fisheries through the development of biologically meaningful stock delineation and thus increase sustainability in the long-term.

Thus, the ultimate goal of this PhD project is to provide information to support the long-term sustainability of the red rock lobster fishery. My PhD project will include four projects:

1) To establish a robust molecular and bioinformatic protocol for single nucleotide polymorphism discovery in *Jasus edwardsii*;

2) To develop a genome-wide set of markers and characterise putatively neutral and markers putatively under divergent selection for investigation of patterns of population genetic structure of *Jasus edwardsii*;

3) To analyse population differentiation patterns of the pueruli/juvenile *Jasus edwardsii* life stage to investigate the mechanisms of establishment of adaptive population divergence;

4) To analyse the influence of physical and environmental factors on the patterns of population divergence of *Jasus edwardsii*.

Chapter 2

ddRADseq and bioinformatic optimization for the red rock lobster
Jasus edwardsii
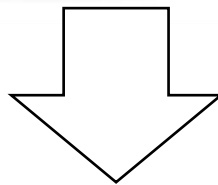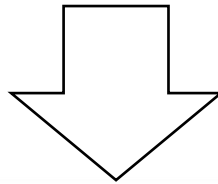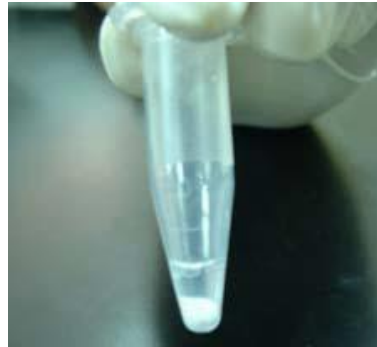
Abstract

The advancement of next generation sequencing and improved reduced representation library (RRL) methodology has facilitated a reduction in the cost of marker discovery and promoted genome-wide *de novo* SNP discovery in many non-model species. In the present study, the double digest restriction site-associated DNA sequencing (ddRADseq) method was employed and optimized for the *Jasus edwardsii* genome. A large genome size with a high level of paralogs led to screening of four enzyme pairs in order to optimise unique fragments with sufficient coverage. Inclusion of the EcoRI-AciI enzyme pair minimized the inclusion of paralogous fragments in the library, while its CpG-methylation sensitivity promoted the representation of genic regions of the genome. In order to compare the effects bioinformatic processing on the population genetic inferences, 120 individuals were sequenced and analysed to compare three pipelines: Stacks, Rad-loci and an in-house pipeline. Comparison of the number of loci/single nucleotide polymorphisms (SNPs), proportion of SNPs under selection, levels of missing data, coverage, error rate as well as population summary statistics (H_O, F_{ST}, F_{IS}) showed that the Rad-loci pipeline was the most optimal trade-off between the SNP discovery rate, proportion of missing data and error rate as population genetic tests from all three pipelines tested were congruent with previously published results. Analysis of technical replicates using PCA and UPGMA clustering revealed high similarity of inter- and intralibrary replicates thus confirming the consistency of wet lab, sequencing and bioinformatic processing and validating the dataset for downstream population genetic analysis.

Introduction

Next Generation Sequencing (NGS) technologies enable massively parallel sequencing at substantially lower costs per base pair compared to non-parallel sequencing methods, such as Sanger sequencing. NGS has substantially advanced the field of population genomics of non-model species (Metzker 2010). The large amounts of data generated by NGS allows for the discovery and characterisation of thousands of Single Nucleotide Polymorphisms (SNPs) (Ekblom et al. 2011). Despite these advancements, whole genome sequencing of species with large or complex genomes remains prohibitively costly (Shendure et al. 2008). Reduced representation libraries (RRLs) have been shown to circumvent these obstacles and allow confident SNP calling in many non-model species (Van Tassell et al. 2008). Genome reduction is achieved through targeting specific regions of the genome; three main strategies are used to obtain RRLs: exome sequencing, RNA sequencing (RNA-seq) (Wang et al. 2017) and restriction site associated DNA sequencing (RADseq) (Davey et al. 2011a). Exome sequencing targets protein-coding genes however its use is restricted by the need for prior knowledge of exon sequences for capture bait design (Andrews et al. 2014). RNA-seq enables analyses of gene expression and can provide invaluable insights into adaptation mechanisms, although expression can vary temporally and be tissue-specific (Wang et al. 2009). RADseq is an effective method for SNP discovery that does not require any prior genomic information and is widely used for genotyping of non-model species (Baird et al. 2008; Davey et al. 2011a; Rowe et al. 2011). Essentially, the RADseq technique yields a library of DNA fragments anchored around a single restriction enzyme and the length of the fragments depends on the mechanical shearing of the DNA. The diversity of restriction enzymes provides a greater level of flexibility compared to the genome reduction methods (RNA sequencing, target capture) and effectively determines the marker density. By using enzymes with longer recognition sequences, less fragments are produced due to the less frequent occurrence of these recognition sequences throughout the genome (Baird et al. 2008). Another way to reduce marker density is to use methylation-sensitive enzymes (Elshire et al. 2011), which will restrict enzymatic fragmentation to non-methylated regions of the genome (Pegadaraju et al. 2013). Peterson et al. (2012) proposed a modified version of RADseq protocol - double-digest RADseq (ddRADseq), that substitutes mechanical shearing of the DNA for an additional restriction enzyme, which reduces library preparation biases of random shearing and results in increases in time and cost efficiency. Furthermore, incorporating a size selection stage into the double-digest library preparation protocol permits

more flexibility to adjust the number of SNPs discovered/number of individuals sequenced per lane (Peterson et al. 2012; Poland et al. 2012).

ddRADseq library preparation relies on an enzymatic fragmentation of genomic DNA with a pair of restriction enzymes. The choice of the enzyme pair is one of the most critical decisions because it influences the distribution and number of the fragments throughout the genome and thus has an impact on the coverage and how representative the library is (Peterson et al. 2012). Thus, optimization of the restriction enzyme combination is an important step in ddRADseq library preparation for any non-model species. A number of software packages are available to conduct a simulated genome digest (Lepais et al. 2014; Wang et al. 2017), although this approach can only be used for species with a reference genome or a genome of a closely related species to that of interest. The need for this optimization step has promoted research specifically focusing on identifying the best performing enzyme pairs for broad taxonomic units including fish, insects (Reiskind et al. 2016) and angiosperms (Yang et al. 2016).

Intrinsic genomic features require the optimization of not only RRL wet lab protocols but also appropriate bioinformatic tools. Genome reduction in non-model species poses a number of challenges during the bioinformatic processing of RADseq data. Variation in coverage can compromise SNP calling due to an inability to distinguish nucleotide polymorphism from sequencing errors (Hohenlohe et al. 2012). False polymorphisms can arise from paralogs or repetitive regions in the genome causing SNPs to be identified when two monomorphic paralogous regions are fixed for alternative variants and erroneously assembled as a single locus (Hohenlohe et al. 2011; Dou et al. 2012). PCR bias for GC-rich fragments can produce PCR duplicates even if the number of amplification cycles is kept to a minimum at the index sequence annealing stage (Davey et al. 2013). Nucleotide variation at restriction enzyme recognition sites can prevent the restriction site from being cleaved and effectively prevent the sequencing of the alternative allele thus producing an "allele dropout" or ADO effect (Gautier et al. 2013). Sequenced fragment length plays an important role in bioinformatic processing as the short length of reads most widely used for sequencing of RADseq libraries impacts the confidence of SNP discovery and genotyping when a reference genome is not available due to the reduced ability to distinguish paralogs. Thus, it is advisable to combine short-read ddRADseq libraries with data obtained from longer-read sequencing platforms (Siccha-Ramirez et al. 2018) and construct a low-quality reference genome (Shafer et al.

2016) or to assemble a pseudo-reference genome from RADseq data (Mandeville et al. 2015), which can be used for mapping of individual reads and genotype inference.

An ever-increasing number of *de novo* and reference – based pipelines demonstrates the multitude of approaches available for SNP discovery: Stacks (Catchen et al. 2011), dDocent (Puritz et al. 2014a), PyRad (Eaton 2014), RaPiD (Willing et al. 2011), RADtools (Baxter et al. 2011), Rainbow (Chong et al. 2012) and AftRAD (Sovic et al. 2015). SNP discovery is further complicated by the number of different parameters and settings within each of the pipelines that also influences loci assembly and consequently biological inferences obtained from the data. Many population genetics studies prefer to rely on in-house pipelines or customised versions and combinations of existing software to accommodate specific challenges of the genome under study (Peterson et al. 2012; Pegadaraju et al. 2013; DaCosta et al. 2014; Candy et al. 2015; Verdu et al. 2016).

The choice of pipeline and appropriate settings has become one of the key challenges in population genomic studies of non-model organisms. In order to understand how pipeline choice and/or the parameters and settings used can affect the conclusions drawn about biological processes within populations it is advisable to assess the summary statistics of multiple bioinformatic pipelines/settings (Shafer et al. 2016; Paris et al. 2017). Metrics that can be used to explore consistency in pipeline comparison include the number of catalog loci, number of SNPs discovered, coverage, error rate, observed heterozygosity (H_O), inbreeding coefficient (F_{IS}), fixation index (F_{ST}), and isolation-by-distance (IBD) if available for comparison from previously published studies (Mastretta-Yanes et al. 2015; Shafer et al. 2016).

After the most optimal wet lab and bioinformatic protocols have been established, another important aspect of processing a large number of individuals is the potential for batch effects – technical or human errors present only in subsets of data due to differences in the processing of batches of samples that can produce false biological signals and need to be controlled for. Sources of batch effects can be both technical (such as change of the equipment used or the quality of reagents) and human (pipetting errors, contamination or mislabelling of the samples) (Mastretta-Yanes et al. 2015). The ddRADseq protocol has a number of inherent sources of error. Enzymatic digestion of DNA can vary between individuals due to variable DNA quality, the effect of which will be amplified by the PCR step of the protocol causing underrepresentation of some fragments in the sample (Davey et

al. 2013). A higher GC content of a fragment will also bias PCR (DaCosta et al. 2014). Amplification success of certain alleles or barcodes can vary, which influences sequencing coverage of corresponding loci or individuals and may produce allele/loci dropout or PCR duplicates (Bonin et al. 2004). A specific source of error in ddRADseq arises from the size selection step. Errors or inconsistencies at this step can introduce fragments out of the target range, sequencing of which will impact the coverage of the target fragments (Peterson et al. 2012). Consistency of the size selection step becomes even more critical if it is performed manually rather than using automated platforms like PippinPrep (Peterson et al. 2012). Manual size selection is performed via a gel cut, thus exposing DNA fragments to the UV light may cause arrest strand polymerisation during PCR thus introducing variation or causing allele or locus dropout (Peterson et al. 2012). In addition, bias can occur during the sequencing process – sequencing errors that occur at variable rates introduce variation across samples and RAD sites (Etter et al. 2012; Loman et al. 2012). When sequencing errors occur in a barcode region it effectively reduces the number of reads obtained from individual tags and can lead to its removal from the dataset if the proportion of missing reads is too high (Hohenlohe et al. 2012). A large number of GC-rich fragments in the library can skew the sequencing coverage as they are usually sequenced at higher depths than fragments with low GC content (Davey et al. 2013). Some of the batch effects can be minimized by randomizing the sample groups across batches although some of the effects cannot be controlled and both errors occurring during wet lab processing and sequencing stages can produce a variable total number of reads and loci per individual, number of reads per locus and per alternative allele for polymorphic loci (Hohenlohe et al. 2012). In order to detect any inconsistencies that may arise during ddRADseq library preparation and sequencing, Mastretta-Yanes et al. (2015) proposed the inclusion of technical replicates in the dataset. DNA obtained from the same individual but assigned a different identification tag should produce the same genotype if all the stages of sample processing were not significantly variable and thus replicate pairs should be expected to cluster together. Correct clustering of replicate pairs can also be employed in the assessment of bioinformatic pipelines/parameters (Mastretta-Yanes et al. 2015).

The relatively large genome size of *Jasus edwardsii*, estimated at 9.8 Gbp (Deiana et al. 1999) (compared to 3.2 Gbp in humans (Venter et al. 2001) or 2.8 Gbp in mice (Church et al. 2009) genomes) is suggestive of high levels of repetitive sequences. The presence of repetitive sequences, mostly transposable elements and short tandem repeats, have been shown to contribute to the large size of the model crustacean genome from *Parhyale*

hawaiensis (3.6 Gbp) (Kao et al. 2016). The genome of another crustacean, the pea aphid, has been estimated to contain approximately 38% transposable elements (The International Aphid Genomics 2010). A large number of genomic repeats has also been noted in the American lobster *Homarus americanus* (Benestan et al. 2015). This led my collaborators (based at La Trobe University, Melbourne, Australia) and I to design pipelines that prioritize paralog detection and filtering. The Rad-loci pipeline designed at La Trobe relies on multiple rounds of clustering of reference loci using the VSearch v.1.1.3 algorithm (Rognes et al. 2016) with the final step being the mapping of all individual reads onto a reference catalog for missing data filtering. I employed an alternative approach to distinguishing paralogs - incorporating existing longer-read *Jasus edwardsii* genomic data generated by the Roche 454 platform (Thomas et al. 2011) and pair-end 300bp Illumina MiSeq reads generated in the pilot sequencing stage of my project. In addition, I included a paralog filtering step with VSearch v.1.1.3 (Rognes et al. 2016).

Therefore, the aim of this chapter is to optimize and validate molecular and bioinformatic protocols for population genetic inferences of the red rock lobster *Jasus edwardsii*.

Specific aims are to:

- To optimize restriction enzyme pairs for the red rock lobster *Jasus edwardsii* genome;
- To compare three *de novo* SNP discovery pipelines;
- To validate the consistency of molecular genetics and bioinformatic methods using inter- and intra – library technical replicates.

Methods

Sampling and DNA extraction

Adult male and female lobsters were sampled in equal proportions where possible by commercial fishermen using baited pots and the number of samples utilized on different stages of the optimisation process is specified in the corresponding sections. A single walking leg was removed from each individual by fishermen and preserved in 96% ethanol until required for DNA extraction. DNA from preserved samples was extracted using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. The DNA concentration was quantified using a Qubit® 2.0 Fluorometer (Life Technologies) and DNA purity was assessed using a NanoDrop 2000. Gel electrophoresis was used to determine DNA integrity. Only relatively intact (molecular weight ≥1000 bp) and pure (260/280 ≥ 1.8) DNA samples were used for ddRADseq library preparation.

ddRADseq protocol optimization and library preparation

A pilot library of six individuals was prepared to test one of the most commonly used enzyme pairs (EcoRI - MspI) and assess its suitability for *Jasus edwardsii* following a modified version of the ddRADseq protocol developed by Peterson et al. (2012). Library preparation started with a double digest of 250 ng of high-quality genomic DNA with frequent (MspI, recognition cite - CCGC) and infrequent cutting (EcoRI, recognition cite – GAATTC) restriction enzymes and ligation of sequencing adapters containing in-line barcodes. Following ligation, double size selection of barcoded fragments was performed using Agencourt AMPure XP magnetic beads (Beckman Coulter) (DeAngelis et al. 1995) and fragments larger than 1000 bp and smaller than 200 bp were removed. In the next step, Illumina TruSeq LT compatible index sequences were attached to barcoded fragments via PCR. After post-PCR clean-up using Agencourt AMPure XP magnetic beads (Beckman Coulter) DNA concentrations were quantified with a Qubit® 2.0 Fluorometer (Life Technologies) and standardized. Libraries were pooled, with each individual contributing 25 ng of DNA, and visualized using agarose gels to screen for the occurrence of paralogs. Fragments with molecular weights between 400, 500 and 600 bp were over-represented in the EcoRI – MspI pilot library, which warranted further investigation for a more optimal enzyme pair. From the size of the over-represented fragments they were likely a result of digestion by the frequently cutting enzyme, MspI. Therefore, three more frequent cutters producing

fragments with sticky ends (5`-GC overhang) appropriate for barcoding and indexing were tested: HpaII, HinP1I and AciI following the same library preparation procedure described above. When the most optimal enzyme pairs were identified, a pilot library of 12 individuals was prepared and a size selection step was performed with fragments between 400 and 500 bp. The pilot library was sequenced on the Illumina MiSeq platform with V2 paired-end 250 bp kit.

After enzyme pair selection, the size selection window and sequencing parameters (sequencing length, paired-end vs. single-end, and data output) were optimized for all additional libraries following the protocol described above. Libraries were pooled at 25 ng of DNA per individual and the size selection step was performed via an agarose gel cut with fragments between 400 and 500 bp targeted. DNA from the excised gel was extracted using the Wizard SV Gel and PCR Clean-Up System (Promega). Final library concentration was quantified with a Qubit 2.0 fluorometer and concentrated, if required, using an Eppendorf Concentrator 5301. The fragment size distribution and concentration of libraries were confirmed using an Agilent ScreenTape assay and qPCR. Samples were sequenced on the Illumina HiSeq2500 platform with single end 100 bp HT chemistry kits at the Australian Genome Research Facility (AGRF) in Melbourne.

For optimization of bioinformatic processing, two ddRADseq libraries were created comprised of 140 lobsters originating from NZ and five lobsters from Tasmania, Australia (**Fig.2.1**). After bioinformatic processing and optimization steps were completed, large scale library preparation and sequencing commenced. Intra- and inter- library technical replicates were introduced as a control for possible batch-effects between libraries and differences in the processing of individuals during library prep and sequencing (Mastretta-Yanes et al. 2015). Thirty adult lobster samples originating from Tasmania, Australia were included to serve as a reference group for comparison with previously published results. In total, 734 lobsters were sequenced, which included 62 unique individuals that were replicated at least twice, resulting in a total of 130 replicates (**Fig.2.2**).

Fig.2.1. Sampling locations for the pilot HiSeq sequencing libraries (bioinformatic optimization steps) are indicated by the blue boxes with the total number of individuals included in the library in the boxes.
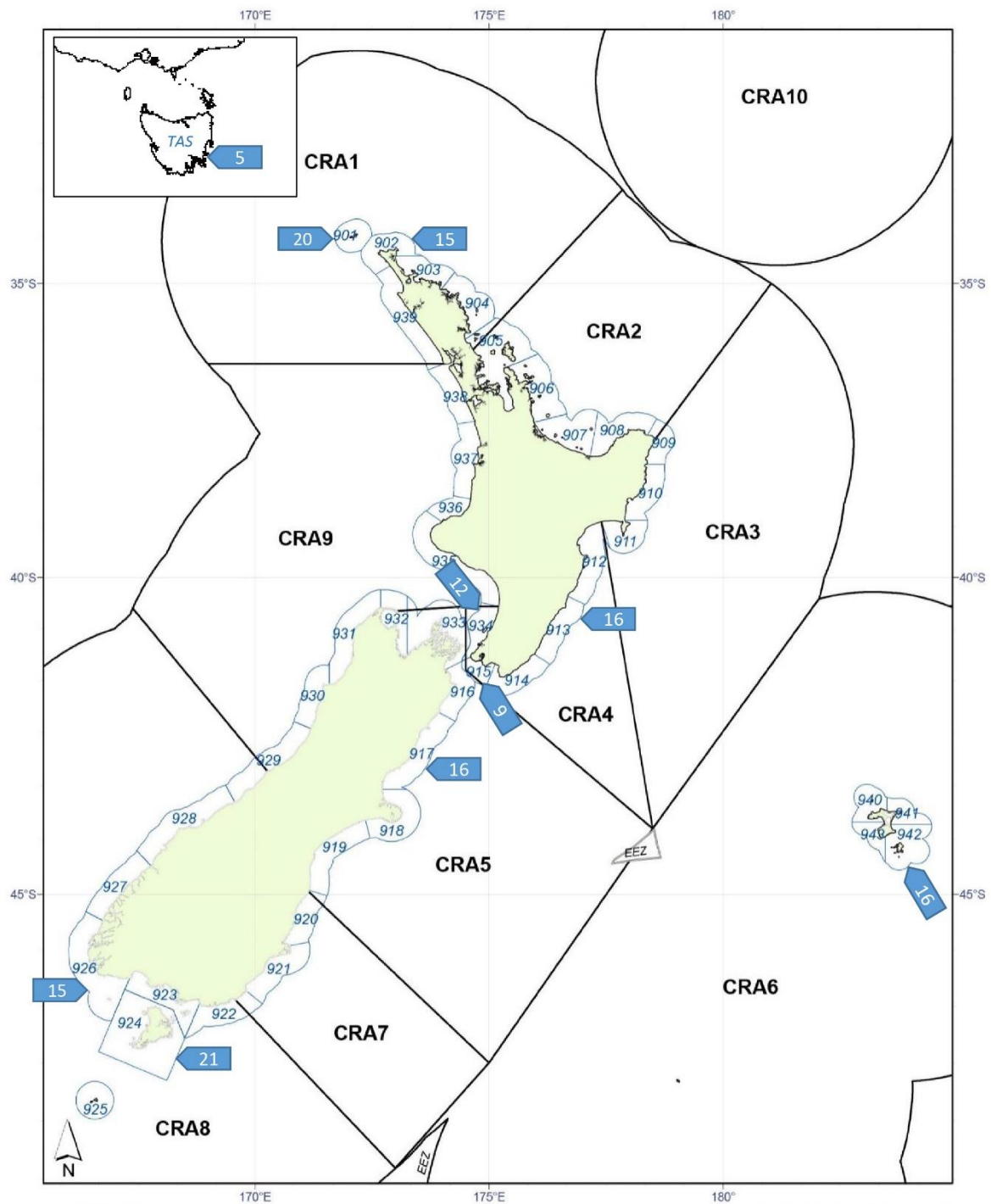
Fig.2.2. Sampling locations of replicates are indicated by the blue boxes with the total number of replicates per location given in the boxes.
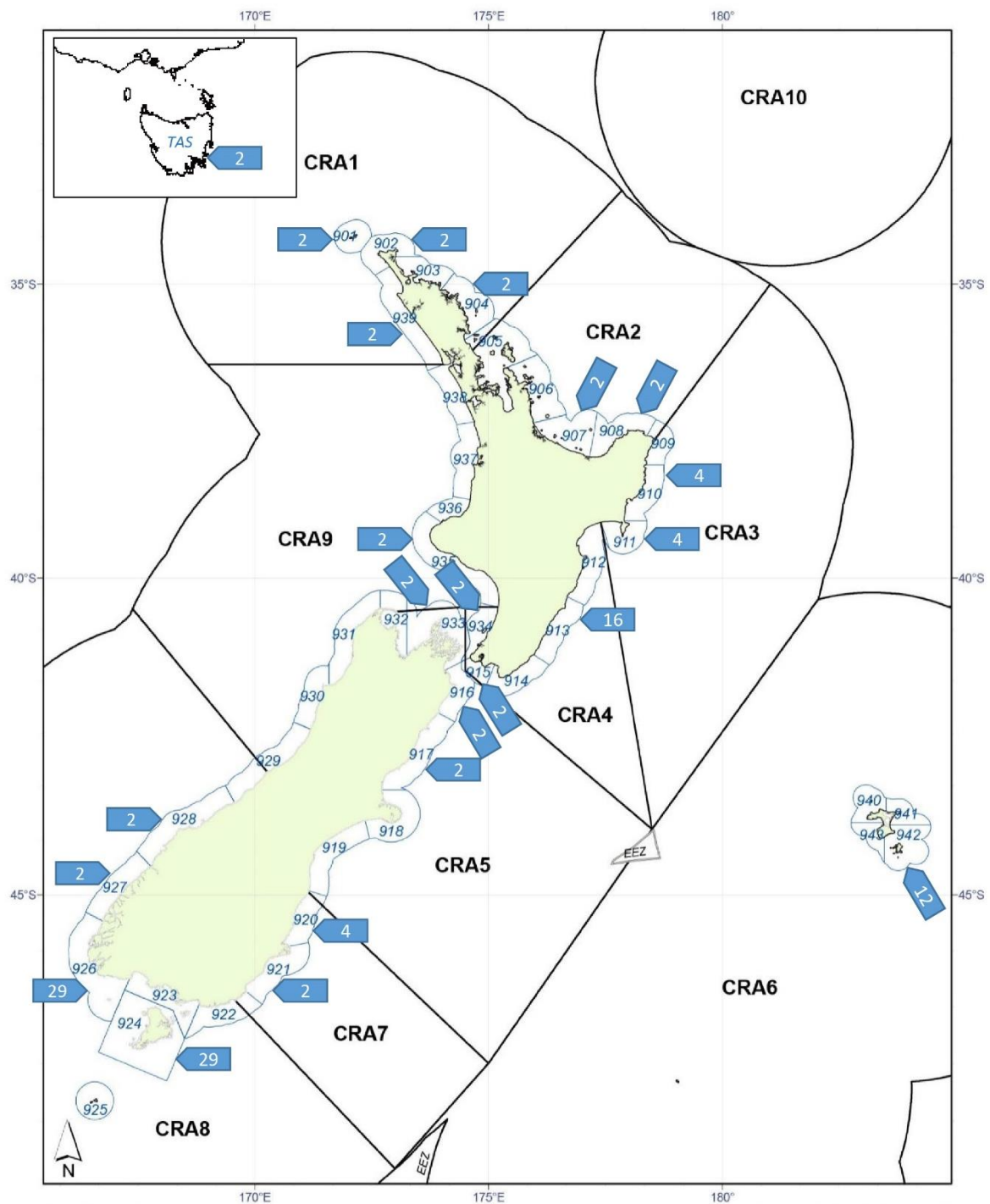
Bioinformatic analysis

Preliminary processing of the ddRADseq data

The quality of sequences obtained was assessed using the FastQC v.0.10.1 quality control tool (Andrews 2010). Reads were then trimmed to 95 bp using TRIMMOMATIC v 0.3 (Bolger et al. 2014). Trimmed reads were filtered for possible contamination using Kraken v.3.5.0 (Wood et al. 2014) and were further demultiplexed using Stacks v.1.37 module *"*process_radtags" (Catchen et al. 2011).

Three bioinformatic pipelines were assessed here: Stacks, Rad-loci and an in-house created pipeline. Stacks has been chosen for comparison as it is considered to be a benchmark software for *de novo* SNP discovery . The Rad-loci pipeline and in house pipeline were included in the analysis as these SNP discovery pipelines focus on paralog detection.

Stacks

The first pipeline used for SNP discovery was Stacks v.1.37 (Catchen et al. 2011). The *de novo* module of Stacks (denovo_map.pl) identifies unique fragments anchored at the restriction enzyme cut site and forms a reference catalog of fragments (a list of the most uniformly occurring sequences throughout the dataset) with sufficient coverage. Individual sequences are aligned to the matching reference sequence and form a "stack" of the sequences of the same length which allows confident genotype calling using a maximum likelihood statistical model. A number of combinations of main parameters were tested in order to identify the most optimal for *Jasus edwardsii* by following the guide to Stacks parameter choice (Paris et al. 2017): *M* (number of mismatches allowed between alleles to form a locus) ranged from 2 to 4, *n* (number of mismatches allowed between loci to accept the locus into the catalog) ranged from 1 to 4, *m* (minimum depth of coverage) ranged from 2 to 6.

The population module of Stacks was executed with the following parameters: *p* (minimum number of populations a locus must be present in) was set to 2 as the dataset was comprised of individuals from Australia and NZ, and *r* (minimum percentage of individuals in a population a locus must be present in) was set to 0.8 in order to retain loci with a small

amount of missing data. The number of assembled loci, SNPs discovered and characterised as putatively under selection (outlier) and putatively neutral, as well as number of individuals that retained at least 80% of loci discovered, were assessed and the most optimal combination of parameters was determined following guidelines of Paris et al (2017) (to maximize the number of extra polymorphic loci identified across 80% of the population, and use $n = M \pm 1$); this identified the following settings: $M = 2$, $n = 1$, $m = 4$.

Rad-loci pipeline

The Rad-loci pipeline developed at La Trobe University (https://github.com/molecularbiodiversity/rad-loci) (Villacorta-Rath et al. 2016) was also used for reference catalog compilation. This pipeline involves the clustering of sequences, which was performed using VSearch v.1.1.3 (Rognes et al. 2016), with the minimal occurrence of a unique sequence set equal to the number of individuals sequenced, which promotes the selection of loci with lower levels of missing data. To merge similar reads that may have occurred from sequencing errors, the clustering identity was set to 94%, meaning a maximum 7 bp mismatch for 95 bp fragments. The clusters obtained were filtered based on the number of members in the cluster: clusters with less than 2 members were removed as they were considered uninformative, clusters with more than 16 members were also removed to avoid multi-mapping. In the next step, sequences were aligned onto the prepared catalog to estimate the number of alleles for each cluster and the catalog was re-filtered to retain clusters with 2 to 16 alleles. Individual sequences were mapped onto the re-filtered catalog once again and loci with more than 40% missing data per locus were removed. Here "locus" is used to describe a unique reference sequence 95 bp long between the EcoRI and AciI cut sites with a cluster of mappable sequence reads.

In-house pipeline

Lastly, I evaluated an in-house pipeline that was developed for the reference catalog construction. The main objective of this pipeline was to use already available sequencing information in the lieu of a reference genome being available for *Jasus edwardsii*. During the initial stages of optimization of the ddRADseq protocol a large number of repeats in the genome of red rock lobster were discovered (see Restriction enzyme screening for

optimization for *Jasus edwardsii* species). Additional sequencing data, especially longer sequences can aid in resolving paralogs in the genome and the construction of a "pseudo-reference" is encouraged when no other reference genome options are available (Shafer et al. 2016; Siccha-Ramirez et al. 2018). Thus, alignment to previously available Roche 454 sequencing information (Thomas et al. 2013) as well as paired end 250 bp Illumina MiSeq sequences generated during the pilot sequencing stage were incorporated into the bioinformatic processing of the short reads (100 bp) generated by the Illumina HiSeq platform.

The 454 and MiSeq reads were quality screened using FastQC v.0.10.1 (Andrews 2010) and "soft" trimmed (the remainder of the read was trimmed after three nucleotides in a row had a Phred score below 30) using TRIMMOMATIC v 0.3 (Bolger et al. 2014). Pair-end reads were merged using PEAR v 0.9.6 (Zhang et al. 2014) with minimum overlap set to 10 bp and minimal length of assembled sequence set to 50 bp. Filtered and prepared 454 and MiSeq sequences, along with reads from two ddRADseq libraries were assembled using Gossamer v.1.3.0 (Conway et al. 2012) with a k-mer size 62 ,which was identified as most optimal via comparison of assembled contigs. VSearch v.1.1.3 (Rognes et al. 2016) was employed for filtering paralogs (retained sequences with less than 95% similarity) and coverage (minimum of 10x) thus forming a reference catalog.

SNP calling and characterization

The in-house and Rad-loci pipeline only create a catalog of reference loci and therefore mapping of the individual reads onto the catalog and SNP calling relies on a third-party software. For both of these pipelines demultiplexed, quality and contamination filtered individual reads were mapped onto the reference catalog using the Burrows-Wheeler Aligner v 0.7.13 (Li et al. 2010). Reference catalog loci with a minimum coverage of 10x were used in an effort to minimize the effect of allele dropout (ADO). SNP calling was performed by two independent programs: Genome Analysis Toolkit (GATK) v 3.5.0 (McKenna et al. 2010) and VarScan v 2.4.1 (Koboldt et al. 2009). Only SNPs consistently called by both programs were selected using SelectVariants module of GATK software. SNPs in concordance identified by GATK and VarScan were further filtered using VCFtools v 0.1.14 (Danecek et al. 2011).

For all three pipelines non-biallelic loci were removed with the *--min-alleles* 2 *--max-alleles* 2 options and only one SNP per locus was retained using the *--thin* 95 option. Filtering of rare alleles was performed using the minor allele frequency filtering option *--maf* set to 0.1. The Linkage Disequilibrium (LD) filter was used to minimize the occurrence of linked loci (option *--min-r2*) with $r^2 > 0.8$ threshold in addition to retaining only one SNP per locus. Individuals and loci with more than 20% missing data were also removed from any further analysis. In addition to commonly used metrics for pipeline comparison (number of loci, number of SNPs), a filtered set of SNPs was further divided into two panels according to calculated F_{ST} value compared to neutral distribution (Excoffier et al. 2010) – neural loci and loci putatively under selection (further referred to as outlier loci). One of the most commonly used methods is based on global F_{ST}, which is employed to identify genetic markers with higher genetic distance than expected in a neutral distribution. A number of demographical scenarios have been shown to influence these estimations and result in type I (false positive) and type II (false negative) errors (Excoffier et al. 2009).

Dependent on the level of selection, mutation rate and hierarchical structure of the populations, different implementations of simulation methods may perform better than others (Narum et al. 2011) and in order to obtain the most conservative estimation, markers under selection were identified with several F_{ST} outlier tests. BAYESCAN v2.1 (Perez-Figueroa et al. 2010) was used with 10 000 iterations and a burn-in of 200 000 steps. A Prior model of 10 000 was used in order minimize the number of false positives (Lotterhos et al. 2015). Arlequin v3.5.2.2 (Excoffier et al. 2010) runs were implemented using 10 000 simulations and 1000 demes. 500 000 simulations and a false discovery rate of 0.1 were used when running Lositan (Antao et al. 2008). A Q-value of 0.05 was used in BAYESCAN runs and the p-value was set at 0.05 in Arlequin and Lositan runs. Only SNPs consistently identified by all three programs were used for further analysis with all three pipelines. In the process of outlier marker detection the mean neutral level of allelic variation was calculated and allelic variation of each locus was compared to the mean value. There are two types of markers under selection, divergent (allelic variation is significantly higher than the mean) and balancing (allelic variation is significantly lower than the mean). A marker under divergent selection indicates that it's variants in the gene can cause a fitness advantage in the particular environment and hence they are selected for and this lineage can diverge from the source population (Nosil et al. 2009). Balancing selection acts against an increase in allele frequency of the advantageous allele via maintaining multiple alleles of the locus/gene under selection

(Charlesworth 2006). Markers under balancing selection usually exhibit lower levels of genetic differentiation when compared to neutral markers and for this reason they are rarely used to examine population differentiation (Nayfa et al. 2016). One of the mechanisms proposed to explain this phenomenon is that balancing selection increases within-deme diversity relative to total diversity thus producing lower F_{ST}. Alternatively, balancing selection may increase the effective migration rate because it gives selective advantage to the novel alleles and increases their chances to spread within a population when compared to the neutral alleles (Schierup et al. 2000). Furthermore, the outlier panel was divided into loci under balancing and divergent selection, only loci under divergent selection were retained for analysis.

Comparison of pipelines through Australia – New Zealand population structure delineation

Descriptive statistics as well as population genetics statistics were calculated for all of the pipelines employed. Descriptive statistics included the number of reference catalog loci, total number of SNPs and the number of neutral and outlier SNPs. In addition, the number of individuals passing the threshold of a maximum of 20% missing data was taken into account when comparing pipelines. Mean SNP error rate was calculated using the R function developed by Mastretta-Yanes et al. (2015). Population genetics statistics included observed heterozygosities (H_O), inbreeding coefficients (F_{IS}), and fixation indices (F_{ST}). All population genetics analyses were performed for neutral and outlier panels using R package diveRsity (Keenan et al. 2013). Pairwise F_{ST} was calculated between New Zealand and Australian populations of *Jasus edwardsii* and compared to a previously reported population differentiation (Thomas 2012; Morgan et al. 2013; Villacorta-Rath et al. 2016).

Technical replicates analysis

Analysis of technical replicates was performed to control for possible batch effects and validate the use of SNPs discovered for population genetics analysis in Chapter 3 and seascape genetics in Chapter 5 of this thesis. Spatial distribution of replicates was assessed using a Principal Component Analysis (PCA) performed in R environment with adegenet v.1.4-1 (Jombart et al. 2011). In addition, hierarchical clustering was performed using the function *hclust* from package stats-package v3.4.1 (Team 2017).

Results

Restriction enzyme screening for Jasus edwardsii optimization

The first enzyme pair tested was EcoRI (infrequent cutter) and MspI (frequent cutter), which are described as one of the most commonly used enzyme pairs by Peterson et al (2012). Genomic DNA was digested with EcoRI – MspI and visualized via gel electrophoresis. The resulting gel showed an even distribution of fragment sizes, which is required for the ddRADseq protocol (**Fig.2.3**.). However, in the concluding stage of pilot library preparation (six individuals total), when samples were pooled for the final size selection step, gel electrophoresis revealed multiple concentrated bands across the target size region: 400, 500 and 600 bp (**Fig.2.4 A**). These bands indicated a large number of fragments of the same size that were presumed to originate from paralogs in the genome or preferential amplification and therefore was unsuitable for continuing with the sequencing process. The position of the bands on both sides of the target size range made it very difficult to make a consistent gel cut and warranted screening for a more optimal enzyme pair.

Fig.2.3. Electrophoresis gel images of *Jasus edwardsii* genomic DNA digestion with a range of restriction enzymes.
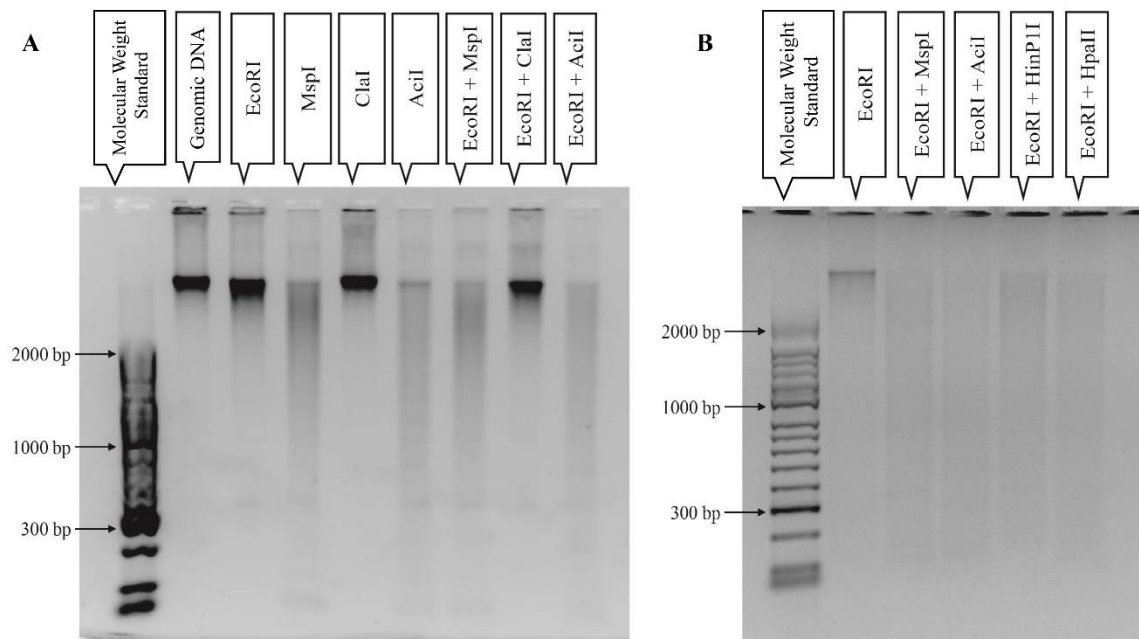
Three more enzyme pairs were compatible with the barcode adapter's sticky ends and were tested: EcoRI – HpaII, EcoRI – HinP1I and EcoRI – AciI. Genomic DNA digests (**Fig.2.3**)

showed an even distribution of fragments and pilot library preparation was tested with all three enzyme pairs. EcoRI – HpaII and EcoRI – HinP1I libraries (**Fig.2.4** panels B and C respectively) contained several "paralog" bands with the most prominent band at 500 bp, which rendered them unusable for library preparation. The EcoRI – AciI library (**Fig.2.4** panel D) had only one band on the lower end of the target size range, which could be relatively easily excluded during the size selection gel cut. As a result, the EcoRI – AciI enzyme pair was selected for all subsequent library preparations.

Fig.2.4. Electrophoresis gel images of *Jasus edwardsii* ddRADseq libraries prepared with 4 pairs of enzymes: A – EcoRI – MspI, B – EcoRI – HinP1I, C – EcoRI – HpaII and D – EcoRI – AciI.
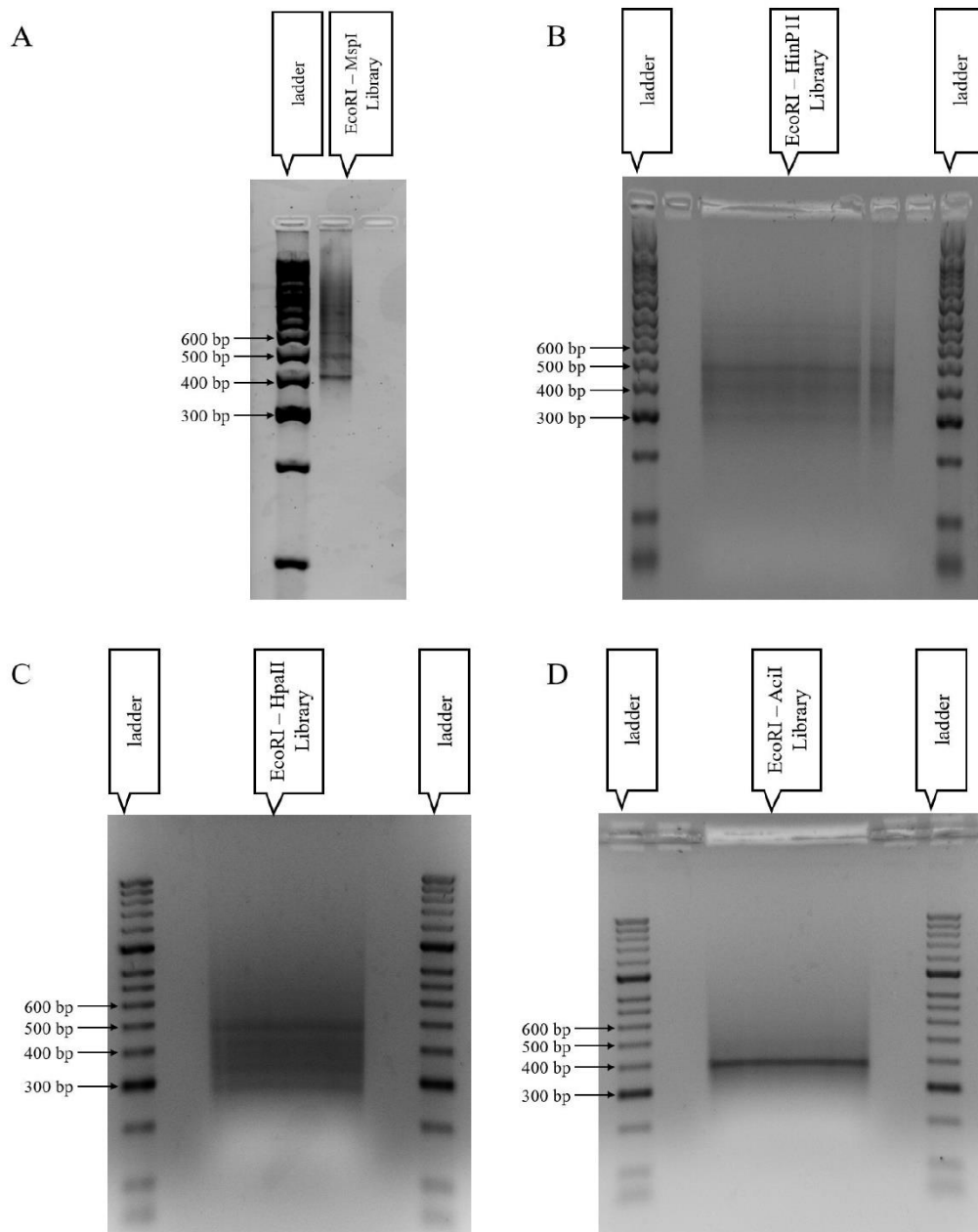
Pipeline comparison

For optimization of bioinformatic processing, 145 individuals were sequenced over two ddRADseq libraries. After quality filtering and the removal of technical replicates from the dataset it consisted of 120 individuals originating from New Zealand and five individuals originating from Australia. Population differentiation inferences for pipeline comparison were performed using neutral and outlier SNPs separately.

Analysis of descriptive statistics for the three pipelines showed relatively similar loci and SNP discovery rates between Rad-loci and in-house pipelines, while the number of loci and SNPs identified by Stacks was more modest (**Table 2.1**). Characterisation of SNPs revealed that 3.2 and 3.5 % of all SNPs were identified as being under diversifying selection for the Rad-loci and in-house pipelines respectively. Only 2.8% of the total SNPs were characterised as being under diversifying selection among the markers discovered by the Stacks pipeline.

Table 2.1. Descriptive statistics for the three pipelines used: the number of sequences in the reference catalog after filtering (Catalog loci), the number of identified polymorphisms after filtering (Total SNPs) and the numbers of neutral and outlier SNPs identified.

Pipeline	Catalog loci	Total SNPs	Outlier SNPs	Neutral SNPs
Stacks	1343	1343	81	964
Rad-loci	3813	1787	122	1258
In-house	4143	2086	145	1473

After mapping of the individual reads onto the catalog, the Stacks pipeline produced a lower level of coverage compared to the other two pipelines (**Table 2.2**) and 10 individuals had to be removed from further analysis (8% of the dataset) as they were missing more than 20% data, while both the Rad-loci and in-house pipeline had only one individual with a high level of missing data (0.8% of the dataset), which was the same for both pipelines. Mean depth of coverage was highest for the Rad-loci pipeline – 34.8x and the minimum depth of coverage was 18.4x. The in-house pipeline generated data with a mean coverage of 32.91x and the minimum depth of coverage was 18.7x. The Stacks pipeline produced the lowest mean and

minimum depth of coverage – 28.02x and 14.8x respectively. Despite the lower depth of coverage, the Stacks pipeline dataset was characterised by the lowest error rate – 0.0378 when compared to the Rad-loci (0.0409) and in-house (0.0463) pipelines (**Table 2.2**).

Table 2.2. Mean coverage, proportion of individuals with more than 20% missing data and mean SNP error rate per pipeline. Coverage provides an average number of reads mapped onto the reference catalog for SNP calling.

Pipeline	Coverage	% missing data	SNP error rate
Stacks	28.02 ± 7.6	8	0.0378 ± 0.034
Rad-loci	34.80 ± 8.3	0.8	0.0409 ± 0.023
In-house	32.91 ± 7.1	0.8	0.0463 ± 0.013

Pipeline comparison based on the population genetic inference

An initial population genetics analysis was performed in order to test the results for consistency with previously identified population structure between *Jasus edwardsii* populations in Australia and New Zealand (Morgan et al. 2013; Thomas et al. 2013; Villacorta-Rath et al. 2016). F-statistic analyses were performed using both neutral and outlier SNP panels (**Table 2.3**). All pairwise F_{ST} values were significant (p-value = 0.0001) for all three pipelines and for both types of marker. Values obtained using the neutral panels of all three pipelines showed weak differentiation between countries with F_{ST} ranging from 0.013 to 0.019. In contrast, panels of outlier markers indicated higher levels of population differentiation between Australia and New Zealand and ranged from 0.186 to 0.258.

Table 2.3. F_{ST}-statistic calculated for neutral and outlier panels for all three pipelines. Statistically significant values (p = 0.0001) are shown in bold.

		Stacks				Rad-loci				In-house	
		NZ	AUS			NZ	AUS			NZ	AUS
Neutral panel	NZ	0			NZ	0			NZ	0	
	AUS	**0.013**	0		AUS	**0.016**	0		AUS	**0.019**	0
		NZ	AUS			NZ	AUS			NZ	AUS
Outlier panel	NZ	0			NZ	0			NZ	0	
	AUS	**0.186**	0		AUS	**0.258**	0		AUS	**0.217**	0

The inbreeding coefficient was negative and significantly different from zero across both panels for all three pipelines and varied between pipelines with the lowest values being produced by the Rad-loci pipeline and the highest by in-house pipeline (**Table 2.4**). However, no significant difference was found between the levels of observed heterozygosity for the pipelines and SNP panels used.

Table 2.4. Descriptive statistics calculated for neutral and outlier SNP panels.

	Neutral panel		Outlier panel	
	H_O (±SD)	F_{IS}	H_O (±SD)	F_{IS}
Stacks				
AUS	0.370 (±0.238)	-0.246	0.336 (±0.285)	-0.219
NZ	0.398 (±0.289)	-0.279	0.375 (±0.337)	-0.231
Rad-loci				
AUS	0.470 (±0.208)	-0.406	0.352 (±0.215)	-0.439
NZ	0.498 (±0.189)	-0.466	0.365 (±0.137)	-0.397
In-house				
AUS	0.338 (±0.238)	-0.135	0.320 (±0.231)	-0.173
NZ	0.398 (±0.229)	-0.185	0.350 (±0.237)	-0.175

Technical replicate analysis

To assess the consistency of wet lab practices and bioinformatic analysis, 62 adult lobsters originating from NZ were replicated between 1 to 7 times and sequenced across 11 adult rock lobster ddRADseq libraries. Two samples were introduced as inter-library replicates, and the rest of the samples replicated were intra-library replicates. Two individuals had more than 20% missing data and were removed from the dataset with their corresponding replicate pair. The total count after all filtering stages was 130 individuals with most individuals replicated only once. Genetic distance between technical replicates should be minimal regardless of whether the SNP is under selection or not because replicates are essentially genetically identical, thus the analysis of this dataset was performed with full set of 2764 markers.

The mean SNP error rate calculated for the technical replicates included in the full-size dataset was 0.0493 ± 0.018. Principal Component Analysis of technical replicates showed that technical replicates of the same individual were consistently placed within the shortest distance of each other, regardless of it being intra- or inter- library replicate (**Fig.2.5,** panel A). Hierarchical clustering using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method resulted in technical replicates being clustered together (**Fig.2.5,** panel B).

Fig.2.5. Technical replicate analysis. Panel A depicts Principal Component analysis on pairs of replicates, panel B – UPGMA clustering of the replicates.
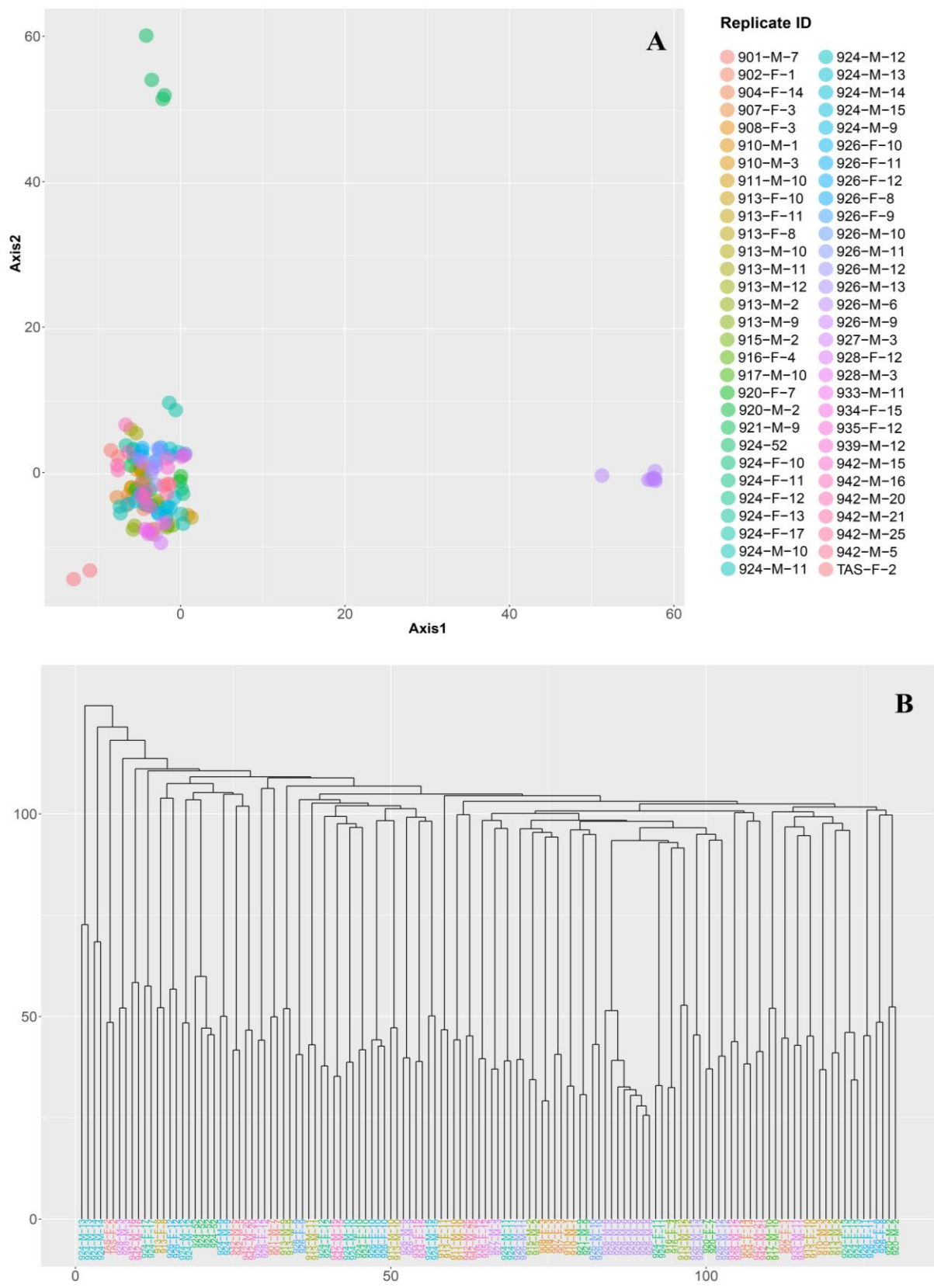
Discussion

The ddRADseq approach to genome reduction combines flexibility in customisation for a wide range of genomes and study goals with a low cost of generating high-resolution population genomic data (Peterson et al. 2012; Puritz et al. 2014b). The *Jasus edwardsii* genome poses two main challenges: its large genome size of 9.8 Gbp (Deiana et al. 1999) and high levels of repetitive sequences. To achieve the optimal representation of unique fragments without sacrificing the coverage needed for confident *de novo* genotype calling four pairs of restriction enzymes were tested. Pair EcoRI-AciI was chosen as it allowed the exclusion of repetitive fragments and the use of a CpG-methylation sensitive enzyme (AciI) reduces the number of restriction site associated fragments by targeting genic regions (Elshire et al. 2011). The large amounts of sequencing data generated by the ddRADseq approach requires careful consideration for bioinformatic processing of the data as it can influence population genetic inferences (Mastretta-Yanes et al. 2015; Shafer et al. 2016; Paris et al. 2017). To assess the influence of bioinformatic processing protocols on the population differentiation patterns of 120 lobsters I used three pipelines: Stacks, Rad-loci and an in-house pipeline. Summary statistics (the number of loci and markers discovered and proportion of markers under selection), data filtering metrics (missing data, coverage, error rate) and population summary statistics (H_O, F_{ST}, F_{IS}) were used to compare the pipelines. In addition, population differentiation patterns produced by the pipelines were compared to previously published results (Morgan et al. 2013; Thomas et al. 2013; Villacorta-Rath et al. 2016). The Rad-loci pipeline produced lower levels of missing data and a lower error rate without sacrificing the SNP discovery rate and was used for all further data analysis later in this thesis. Finally, technical replicates introduced into the full-size dataset were analysed for consistency using PCA and UPGMA clustering approaches and showed high levels of reproducibility and adequacy of the dataset for downstream population genetic analyses.

Optimization of the ddRADseq wetlab protocol

The ddRADseq protocol was designed to allow sufficient flexibility to accommodate virtually any genome. The main protocol adjustments are the choice of restriction enzyme and a size range of fragments selected for sequencing (Peterson et al. 2012). The red rock lobster genome size has been estimated to be just under 10 Gbp (Deiana et al. 1999) and large genomes often have a high number of repetitive sequences and high paralogy levels have

been observed in lobsters (Benestan et al. 2015), thus the ddRADseq protocol for the red rock lobster genome required optimization of the standard protocol in order to obtain a sufficient number of unique loci for population genetic inference without sacrificing the quality of the SNP genotyping. After testing four pairs of enzymes for the *Jasus edwardsii* genome EcoRI – AciI was considered the most optimal. AciI restriction enzyme is a CpG methylation sensitive enzyme and although the use of methylation sensitive enzymes have been cautioned against due to potential to target both genetic and epigenetic polymorphisms (Jiang et al. 2016), methylation-sensitive enzymes have been successfully employed in a number of reduced representation sequencing projects (Pegadaraju et al. 2013; Araneda et al. 2016; Schield et al. 2016). The distribution of methylated sites has been shown to be non-random and can form "gene deserts" in the genome (Zilberman et al. 2007; Jones 2012). Repetitive DNA fragments, which are often produced by transposable elements and retroviruses, are heavily methylated, thus, the use of methylation-sensitive enzyme minimises the number of non-unique DNA fragments in the reduced representation library (Elshire et al. 2011). Additionally, targeting the hypo-methylated regions of the genome allows more genic regions to be included in the library, which can be very useful in testing whether the study species might be under selective pressure (Pegadaraju et al. 2013; Araneda et al. 2016).

Optimization of bioinformatic processing of the sequencing data

The in-house pipeline produced the largest number of reference catalog loci, which in turn yielded a higher number of discovered SNPs, while the Stacks pipeline loci catalog consisted of only 1343 sequences and the same number of SNPs (due to one SNP per loci filtering parameter). The Rad-loci pipeline had an intermediate loci and SNP discovery rate. Although maximising the number of markers discovered is often desirable and has been used as a main guide in the pipeline choice (Callicrate et al. 2014; Pante et al. 2014), consideration of other output statistics has been recommended (Mastretta-Yanes et al. 2015; Shafer et al. 2016). Interestingly, the proportion of outlier SNPs detected was relatively consistent between all three pipelines. Another important metric is coverage, as sufficient coverage can counteract to some extent the effects of sequencing errors, assembly errors and missing data (Bentley et al. 2008). Generally, *de novo* sequencing projects require higher coverage than genome assembly studies (Buerkle et al. 2013). Also, coverage cut-off is often dependent on the proportion of the genome sampled and sample size, and thus is variable between studies

(Pool et al. 2010). If maximal representation of all loci in all individuals is desired a 60x coverage has been estimated for *de novo* sequencing although this level of coverage is rarely needed (Davey et al. 2011b). A depth of coverage of 20x has been advised when accurate genotype calling is required (Andrews et al. 2014), although 7x coverage has been reported as sufficient for population genetics inferences in the non-model plant species *Berberis alpine* after careful tailoring of the bioinformatic processing of the data (Mastretta-Yanes et al. 2015). Mean coverage levels for all the pipelines tested here exceed the coverage cut-off of 20x, with highest coverage obtained with Rad-loci pipeline. The Stacks pipeline produced the lowest level of coverage out of three pipelines tested, which is consistent with the highest number of individuals being removed due to high levels of missing data. The sequencing process is random and because not all the loci will be sequenced for all the individuals missing data can create an uncertainty as to whether or not all the allelic variance has been sequenced (Bentley et al. 2008). Thus both sequencing errors and missing data can significantly influence population genetic inferences, especially demographic processes and selection, and therefore lower levels of missing data are desirable (Pool et al. 2010).

SNP error rate gives an indication of the level of 'noise' in the genetic signal and is a very important metric in the pipeline comparison for population genetic analyses. Mastretta-Yanes et al. (2015) reported error rates ranging from 0.0243 (for the set of parameters authors identified as most optimal) to 0.0578 (for a high coverage parameter set). Error rates obtained for all three pipelines in the present study ranged from 0.0378 to 0.0463 and are comparable to the error rates estimated by Mastretta-Yanes et al. (2015). In addition, Mastretta-Yanes et al. (2015) reported that Stacks parameter combinations characterized by a smaller error rate produced higher pairwise F_{ST} estimates. In my study the error rate was highest for the in-house pipeline and lowest for the Stacks pipeline, although error rates of the Stacks and Rad-loci pipeline weren't significantly different. F_{ST} estimates, however, were not consistent with the error rates calculated: the highest level of differentiation was identified by the outlier markers for the Rad-loci pipeline (F_{ST} = 0.258, p-value = 0.0001) and the lowest by the Stacks pipeline (F_{ST} = 0.186, p-value = 0.0001). This discrepancy may originate from the number of SNPs under selection discovered by the pipelines: 122 (Rad-loci) vs 81 (Stacks) because the resolving power of biallelic markers like SNPs relies on a large number of markers being identified (Mesak et al. 2014). The higher error rate produced by the in-house pipeline is consistent with a slightly lower F_{ST} value (F_{ST} = 0.217, p-value = 0.0001).

Pairwise F_{ST} values obtained between Australia and New Zealand were also compared to previously published research. Thomas & Bell (2013) originally reported genetic differentiation across the Tasman Sea using microsatellite markers with an F_{ST} value of 0.018 (p-value = 0.010). The three pipelines tested identified comparable levels of differentiation of neutral markers: 0.013 (Stacks), 0.016 (Rad-loci) and 0.019 (in-house). These values are also similar to the value obtained in the study by Morgan et al. (2013) conducted using nine microsatellite markers (including eight microsatellite markers developed by Thomas & Bell (2013) with one extra marker discovered) and reported significant differentiation (F_{ST}=0.0305, p-value < 0.05). In an earlier study using Single Nucleotide Polymorphisms using outlier/neutral SNP panels to investigate AUS-NZ population differentiation, a total of 88 individuals comprised of four ddRADseq libraries and were sequenced using Illumina MiSeq next-generation sequencing platform with v2 chemistry and 250 bp paired-end kits (Villacorta-Rath et al. 2016). Data was analysed using the Rad-loci pipeline and yielded 535 neutral and 121 outlier SNPs and reported a pairwise F_{ST} value of 0.022 for neutral markers and a F_{ST} value of 0.134 for outlier markers, with both being significant (Villacorta-Rath et al. 2016). Although the same pipeline was used (Rad-loci), the F_{ST} value obtained in present study with outlier markers demonstrated a much higher level of differentiation (0.134 in the previous study vs 0.258 reported here with Rad-loci pipeline and congruently higher levels of differentiation detected by other pipelines: Stacks – 0.186, in-house – 0.217). Multiple studies have shown that the resolving power of SNPs is dependent on the number of markers used (Vignal et al. 2002; Smith et al. 2007; Narum et al. 2008; Helyar et al. 2011), but the total number of SNPs under selection discovered was 127, which was very similar to the earlier study that used 121 SNPs (Villacorta-Rath et al. 2016). The increase in resolving power of the markers might be explained by a more stringent selection process (markers consistent with Arlequin, BayeScan and Lositan) in order to minimize Type I and II errors, while in the previous study markers were identified using only FDist2 method implemented in Lositan. This hypothesis is supported by the similar levels of AUS-NZ divergence detected with neutral markers (0.02 in the previous study vs. neutral marker F_{ST} ranging 0.13 to 0.19 obtained with different pipelines in the present study).

The consistency of population genetic inference between the pipelines I analysed and with previously published data supports the use of a wide range of bioinformatic processing protocols. However, for large-scale population structure analysis the use of all three pipelines would be very time and resource costly. Based on the summary statistics obtained here the

Rad-loci pipeline has been chosen for the analysis of the full-size dataset as a trade-off between the number of markers discovered and minimising the SNP error rate and levels of missing data.

Validation of consistency of the wet lab and bioinformatic processing via technical replicates

As a result of the work I conducted, a vast amount of sequencing data was obtained (up to 3 million reads per individual) with 734 adult lobsters being sequenced across 11 ddRADseq libraries. Consistency of the size selection between the libraries is one of the most vulnerable stages in ddRADseq library preparation and can have major consequences on population genetic inference. Inconsistent size composition of restriction associated fragments between libraries can reduce sample size as more individuals have to be filtered out as having high levels of missing data, which in turn can increase the lower bound on the detectable allele frequency (Lynch 2009). In addition to technique variation introduced during wet lab processing and sequencing of the samples in batches, bioinformatic analysis poses another great concern for this type of study (Shafer et al. 2016). *De novo* assembly and reference catalog construction is a pivotal step in SNP discovery and its quality has been shown to vary with different parameters (Catchen et al. 2013; Paris et al. 2017). The absence of a reference genome prevents the assessment of catalog loci assembly in *Jasus edwardsii* and the use of technical replicates has been described as a reliable alternative (Mastretta-Yanes et al. 2015). A total of 64 intra-library and 11 inter-library replicates were sequenced across 11 ddRADseq libraries of adult red rock lobsters. Principal Component Analysis and UPGMA hierarchical clustering of technical replicates showed corresponding replicates clustered together, which allows me to conclude that the molecular and bioinformatic procedures were performed consistently across multiple libraries and provides quality assurance that the population structure inferred from the dataset should be biologically meaningful.

Mean SNP error rate was also estimated using technical replicates for the full-size dataset and although it is slightly higher than the error rate estimated for the pilot pipeline comparison dataset (0.04931 ± 0.018 vs 0.0409 ± 0.023) it is below error rate estimated by Mastretta-Yanes et al. (2015) for a high-coverage Stacks parameter combination.

In conclusion, wet lab protocols have been optimized for the relatively large genome of *Jasus edwardsii*, which has the additional challenge of high levels of repetition in its genome. Bioinformatic processing has been optimized by testing three de novo restriction site

assembly and SNP discovery pipelines: Stacks, Rad-loci and in-house pipelines and the Rad-loci pipeline has been chosen as the most optimal trade-off between marker discovery rate, level of missing data and SNP error rate. Finally, the consistency and validity of the whole genotyping routine, starting from the ddRADseq library preparation to SNP identification has been assessed using technical replicates. PCA and UPGMA clustering of technical replicates confirmed reproducibility and validity of the protocols used thus controlling for the effects of the data generation process on the biological inferences.

Chapter 3

Neutral and markers under selection reveal contrasting patterns of differentiation of red rock lobster *Jasus edwardsii* in New Zealand

Abstract

Determining genetic structure and population connectivity is important for understanding evolutionary and ecological processes and for the conservation and sustainable management of harvested species. This study aimed to investigate the effects of gene flow and selection on the genetic structure and population connectivity of the species with one of the longest known pelagic larval duration reported – the red rock lobster *Jasus edwardsii*. For this study, 644 adult lobsters, collected from 23 locations from across most of the species' Australian and New Zealand distribution range, were genotyped at 2336 SNP markers using a ddRADseq approach. Markers were characterized into putatively neutral (1648 SNPs) and those under divergent selection (147 SNPs). Previously published population differentiation between NZ and AUS has been detected with both types of markers, however, no significant population structuring within NZ was identified using neutral markers in the present study, which indicated a high level of gene flow and connectivity between populations (F_{ST} ranged 0.0001 to 0.0008, P < 0.001). Contrastingly, there was evidence of relatively high selective pressure identified through the population genetic analysis of outlier markers identified through the analysis of outlier markers; three main regions were apparent within NZ: North-East, North-West and South sub-populations as parts of a larger NZ metapopulation (F_{ST} ranged 0.025 to 0.049, P < 0.001). Interestingly, the North-West and South sub-populations appear to have adaptively diverged in response to the extremes of an environmental cline stretched along the west coast of NZ. The results of this study suggest that local selection may be an important mechanism influencing population structure of red rock lobster despite high levels of gene flow and demographic connectivity. Results obtained using outlier markers suggest that differential selection across NZ may promote the reproduction of genotypes adapted to local condtions, and neutral markers indicate that larvae of successful breeders are thoroughly mixed during the two-year-long pelagic larval duration.

Introduction

Understanding population connectivity, patterns of differentiation and gene flow are critical for the conservation (Burgess et al. 2014) and sustainable management of harvested species (Kough et al. 2013). The adult life stages of many marine animals exhibit relatively restricted patterns of movement (Barnes et al. 1999). For these species planktonic larval dispersal is often the central mechanisms promoting population connectivity, often over large distances (Kritzer et al. 2010). Planktonic larval duration (PLD) can vary considerably between species – from hours to years. PLD has traditionally been assumed to correlate with dispersal distance due to the large numbers of dispersers, current flow patterns and a lack of apparent barriers to larval transport (Shanks et al. 2003). However, recent studies of larval connectivity have challenged this assumption by describing patterns of self-recruitment and local larval retention (Almany et al. 2007; Puebla et al. 2012; Benestan et al. 2015; Pocwierz-Kotus et al. 2015). The identification of populations that rely heavily on local recruits is one of the critical considerations for conservation and sustainable harvest because they are under higher extinction threat from overfishing and habitat destruction. Non-self-recruiting populations rely on an influx of recruits and some populations may act as a source of larvae subsidizing other populations ("sinks") within a larger metapopulation. Thus, population connectivity is crucial for generating and maintaining metapopulation structure and demographical stability (Weersing et al. 2009; Lowe et al. 2010; Lowe et al. 2017). Larval dispersal also maintains gene flow and circulates novel alleles between populations, which provides the potential for population resilience to changing environments and human exploitation (Jump et al. 2009). Genetic diversity can provide valuable insights into the overall state and wellbeing of a population (Johannesson et al. 2006; Worm et al. 2006). A sufficient level of adaptive genetic divergence is required in a population to be able to adapt to the changing environment (Jump et al. 2005; Jump et al. 2009). The diversity of neutral markers is mainly a product of random processes, such as genetic drift (Luikart et al. 2003). In contrast, adaptive markers (markers undergoing adaptive selection in the population) are the result of selective and stochastic processes (Galindo et al. 2009). Therefore, studying adaptive genetic diversity is essential to understanding the long-term survival potential of a population (Eckert et al. 2008; Gebremedhin et al. 2009).

Reproductively effective migration is a mechanism of transfer of novel alleles between spatially distributed populations (Slatkin 1987). However, if a barrier to gene flow arises effects of the genetic drift become more pronounced over time and can result in a significant

loss of genetic variation, especially in populations with small effective size (Hellberg et al. 2002). Genetic drift, as well as gene flow, are processes influencing the genetic diversity of neutral markers that have been employed for decades to estimate levels of genetic differentiation (Bruford et al. 1993; Streiff et al. 2001; Hauser et al. 2002; Ellegren 2004; Chistiakov et al. 2006).

Another major force that can significantly influence genetic diversity of populations is selection, which in combination with gene flow produces adaptive loci. Gene flow effectively increases, and to an extent homogenizes, the gene pool upon which selective force can be applied. Selection leads to an increase in the allele frequencies of advantageous or adaptive genes but at the same time is counteracted by the effect of gene flow (Kirk et al. 2011). The population differentiation pattern arising from the combination of these two processes will depend on the strength of any selection pressure and the effectiveness of the gene flow.

In a natural population interplay of all three mentioned processes (genetic drift, genetic flow and selection) can result in different patterns of differentiation in adaptive versus neutral genes, which have been demonstrated by multiple studies (Freamo et al. 2011; Matala et al. 2014), often with markers experiencing selection having higher resolving power than neutral markers (Milano et al. 2014; Villacorta-Rath et al. 2016).

Jasus edwardsii or the red rock lobster (Hutton 1875) is a species of spiny lobster belonging to the Palinuridae family. It is distributed throughout the coastal waters of southern Australia and New Zealand and has a complex life cycle: adults inhabit rocky reefs where they spawn, and the first development stage - naupliosoma molts into phyllosoma within several hours of being released. The phyllosoma drift with offshore currents for 12-24 months before they return to inshore waters, where it transforms into puerulus. Pueruli have been shown to be active swimmers (Jeffs et al. 2001), so once they reach continental shelf their movements are deliberate, and dispersal is affected by numerous factors: underwater sound, water chemistry, eddychemotaxis, salinity, magnetic fields, and specific behavior (Cobb 1997; Jeffs et al. 2005). The puerulus molts into a juvenile within a few days after settlement, and the juvenile grows into an adult (Booth et al. 1994). Once settled, rock lobster movements generally do not exceed 5 km (Booth 1997) so most of the demographic (and genetic) connectivity is carried out by the planktonic life stage.

Tracking the fate of planktonic larvae is notoriously difficult, mainly due to their small size compared to the extent of the marine environment (Becker et al. 2007). As a result, physical modeling, geochemical tracers, and genetic markers are the main approaches available to

estimate larval dispersal and connectivity of red rock lobster (Levin 2006). Oceanographic modelling *of Jasus edwardsii* settlement by Chiswell & Booth (2008) identified four main areas of dispersal within NZ. According to their model, populations in the far North receive most of their settlers from the West coast of NZ and supply larvae to the populations along the East coast. In turn, the East coast region entrains a large amount of locally produced settlers owing to the Wairarapa eddy and contributes moderately to the larval pool in the South of NZ. The Chatham Islands act mostly as a "sink" for larvae originating along the West coast of South Island. Lastly, high levels of self-recruitment were predicted to occur in the South Island.

Drifting with currents was suggested to be the main driver of the early- and mid-stage phyllosoma distribution, however, it does not explain the distribution of the late-stage phyllosomas and pueruli, which is one of the limitations of the model (Chiswell et al. 2005; Chiswell et al. 2008). Furthermore, the model by Chiswell & Booth (2008) assumed no mortality, which is another important limitation because selective mortality of the migrants to the new environment can reduce effective population connectivity (Marshall et al. 2010).

A number of genetic tools have been used in order to estimate *Jasus edwardsii* genetic population connectivity. Smith et al. (1980) described red rock lobster populations as thoroughly mixed throughout their range after analysis of 21 protein encoded by 33 enzyme gene loci (Smith et al. 1980). No significant genetic differentiation in the lactate dehydrogenase (*Ldh*) locus have been reported (Booth et al. 1990). Analysis of nucleotide sequence polymorphism of mtDNA demonstrated no population differentiation in Australia and New Zealand (Ovenden et al. 1992). These earlier studies all suggested that populations were well-mixed. However, this assumption was challenged by Thomas & Bell (2013) who described significant differentiation between Australian and New Zealand lobster stocks (F_{ST} = 0.018, P = 0.010) using a set of eight microsatellite markers. Additionally, they discovered, however weak, genetic signal of three genetically distinct subpopulations of *Jasus edwardsii* within New Zealand: Northern NZ, central NZ and Southern NZ (Thomas et al. 2013). A study by Morgan et al. (2013) using the same markers with one additional marker confirmed significant differentiation between populations of red rock lobster across the Tasman Sea, but little evidence for structuring within Australia.

Microsatellites have been a molecular marker of choice for over a decade in population genetics studies, however, the recent development of high-throughput sequencing, as well as advances in computational biology, have made extensive genetic marker discovery possible and led to wider use of Single Nucleotide Polymorphism markers (SNPs) in studies of non-

model species (Helyar et al. 2011). Several studies have compared the performance of STR, and selectively neutral SNP markers and the level of resolution appears to be dependent on genetic attributes of the organism studied and allelic richness of microsatellites used. For example, 481 SNPs provided a better resolution than 32 microsatellite loci in a study of killifish (*Kryptolebias marmoratus)* (Mesak et al. 2014), but Narum et al. (Narum et al. 2008) found that 13 STRs and 37 SNP markers in Chinook salmon *(Oncorhynchus tshawytscha)* showed similar results on a large geographic scale (>2700 km), however microsatellites provided better resolution at a finer spatial scale (~3 km). Depending on the species studied, on average, one microsatellite marker has been reported to have the same resolving power of roughly 10-15 SNPs (Narum et al. 2008; Glover et al. 2010; Hess et al. 2011; Mesnick et al. 2011). However, a small number of markers (which is common in microsatellite studies) may not be able to accurately represent the genome-wide genetic diversity of natural populations (Väli et al. 2008). Development of large panels of SNP markers has become more feasible with the advances in next-generation sequencing and marker discovery in non-model species, combined with a high number of SNPs occurring throughout the genome. High-density selectively neutral SNP panels have been shown not only to reach the same level of resolution as microsatellites but have a high potential to exceed it (Olsen et al. 2011; Mesak et al. 2014). SNP markers under selection (also referred to as adaptive or outlier markers) can also provide additional information about genetic diversity, population differentiation, local adaptation and survival/persistence potential (Milano et al. 2014; Pilot et al. 2014; Villacorta-Rath et al. 2016). Markers under selection are identified by comparing the level of allelic variation of each marker to the mean neutral allelic variation. A significantly higher level of allelic variation than can be expected from genetic drift alone indicates that a marker is under divergent selection - allelic variants of this locus provide a fitness advantage and are selected for, causing the genetic divergence of this lineage (Nosil et al. 2009). Significantly lower levels of allelic variation compared to mean neutral marker variation usually characterises markers under balancing selection (see **Chapter 2,** *SNP calling and characterization*) therefore for this study I focused on markers under selection.

Therefore, the aim of this chapter is to develop two panels of SNP markers: a panel of putatively neutral markers and a panel of markers putatively under diversifying selection to assess population genetic structure of the red rock lobster *Jasus edwardsii* in New Zealand.

Specific aims include:

- To investigate population structure of the red rock lobster *Jasus edwardsii* using selectively neutral SNP markers
- To investigate the influence of selection on the population structure of red rock lobster *Jasus edwardsii* using SNP markers under divergent selection
- To compare the population differentiation patterns identified by different types of marker – microsatellites vs SNPs
- To compare inferred genetic connectivity patterns with demographic connectivity predictions based on oceanographic models (Chiswell et al. 2008; Chiswell et al. 2011)

Methods

Sampling and DNA extraction

A total of 660 adult lobster samples were used in this study. Lobsters were collected between July 2015 and March 2016 from 23 locations. Twenty sampling sites were located around the main islands of New Zealand, 2 sites were near offshore islands and 1 sampling location was in Tasmania, Australia (see **Fig.3.1, Table 3.1**). Thirty adult lobster samples originating from Tasmania, Australia were included to serve as a reference group for comparison with previously published results (Morgan et al. 2013; Thomas et al. 2013; Villacorta-Rath et al. 2016). Adult male and female lobsters were sampled in equal proportions where possible by commercial fishermen using baited pots. A single walking leg was removed by fishermen and preserved in 96% ethanol until required for DNA extraction. The number of samples per location varied between 14 and 35. DNA from preserved samples was extracted using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. The DNA concentration was quantified using a Qubit® 2.0 Fluorometer (Life Technologies) and DNA purity was assessed using a NanoDrop 2000. Gel electrophoresis was used to determine DNA integrity. Only relatively intact (molecular weight ≥ 1000 bp) and pure ($260/280 \geq 1.8$) DNA were used for ddRADseq library preparation.

Table 3.1. Lobster sampling locations with the total number sequenced per location (N_{SEQ}) and number of lobsters after filtering (N_{FILT}).

Region	Sampling cite ID	Latitude	Longitude	Date	N_{SEQ}	N_{FILT}
North Island	901	34° 7'58.85"S	172° 4'25.16"E	Sep-15	31	30
North Island	902	34°24'16.35"S	172°46'56.00"E	Sep-15	31	30
North Island	904	35°41'43.86"S	174°33'50.77"E	Sep-15	31	30
North Island	907	37° 6'43.30"S	176°52'6.00"E	Aug-15	31	30
North Island	908	37°36'17.15"S	177°55'20.62"E	Aug-15	15	14
North Island	910	38°39'3.84"S	178°10'9.56"E	Aug-15	18	16
North Island	911	39°16'21.10"S	178°46'22.44"E	Sep-15	31	29
North Island	913	41°12'40.00"S	177°27'54.72"E	Jul-15	39	30
North Island	915	41° 6'40.00"S	174°47'40.00"E	Jul-15	20	19
South Island	916	41°43'18.11"S	174°16'54.39"E	Oct-15	31	30
South Island	917	42°26'12.39"S	173°42'4.72"E	Oct-15	31	30
South Island	920	45°33'3.59"S	170°45'10.19"E	Oct-15	32	30
South Island	921	45°53'35.52"S	170°41'13.80"E	Sep-15	31	30
South Island	924	47°17'47.26"S	167°35'34.44"E	Sep-15	50	32
South Island	926	46° 0'2.06"S	166°35'34.32"E	Sep-15	48	29
South Island	927	45° 4'36.90"S	167° 6'27.86"E	Sep-15	31	30
South Island	928	44°34'59.70"S	167°48'25.28"E	Sep-15	31	29
South Island	933	41°21'0.24"S	174° 5'58.84"E	Oct-15	31	29
North Island	934	41° 6'40.00"S	174°47'40.00"E	Jul-15	20	19
South Island	935	39°16'38.87"S	173°43'18.71"E	Aug-15	31	30
South Island	939	35°42'8.44"S	173°27'44.46"E	Aug-15	31	30
Chatham Islands	942	44°20'55.74"S	176°12'30.60"W	Aug-15	45	35
Tasmania	TAS	43°39'27.78"S	147°24'10.01"E	Mar-16	44	33

Fig. 3.1. Map of sampling locations. The numbers in blue rectangles reflect the number of adult lobsters collected at each location. Map adapted from NZ Seafood Industry Council report, 2005.
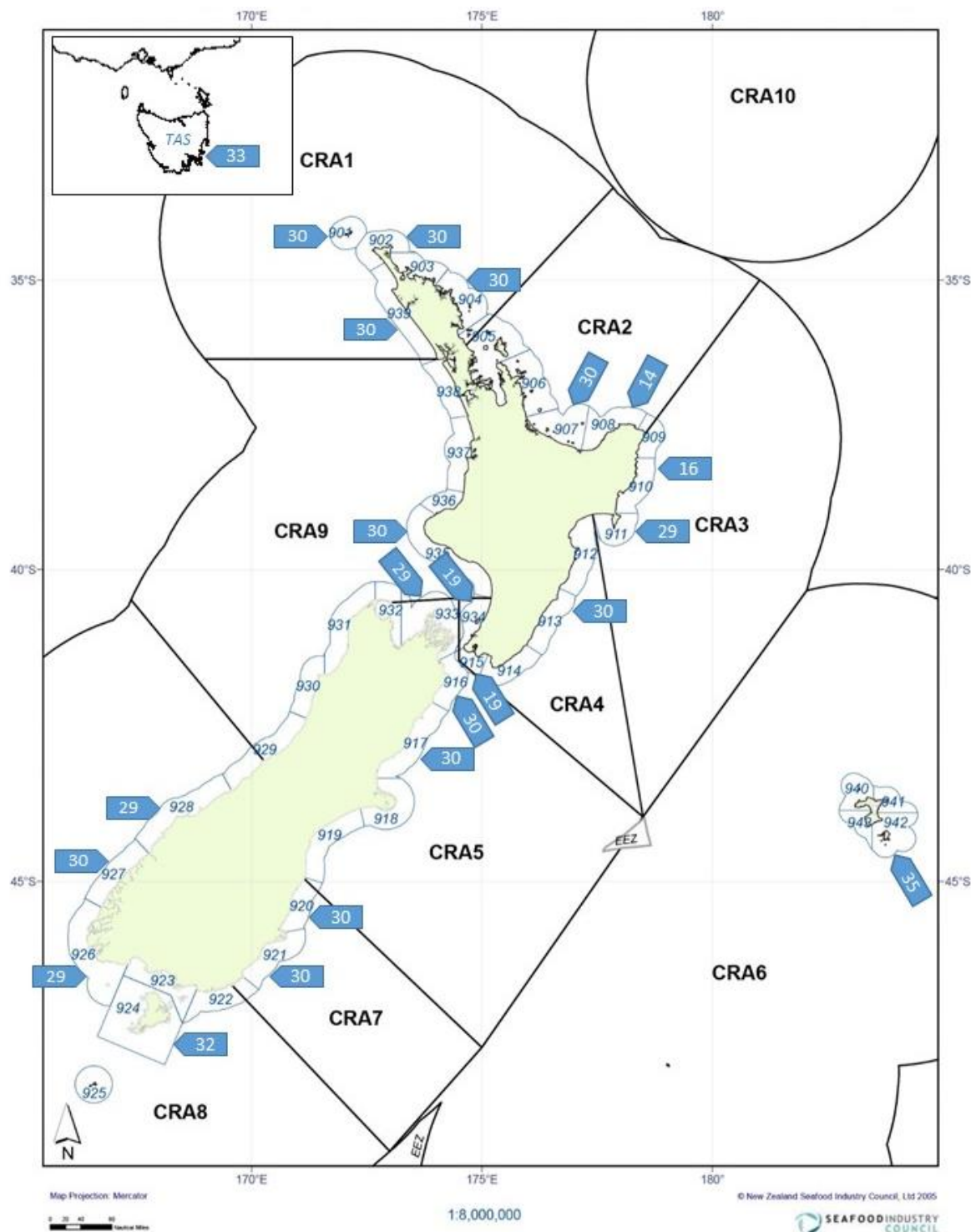
Library preparation and sequencing

All libraries were prepared following a modified version of the ddRADseq protocol developed by Peterson et al. (2012). Library preparation started with a double digest of 250 ng of high-quality genomic DNA with frequent (AciI, recognition cite - CCGC) and infrequent (EcoRI, recognition cite – GAATTC) cutting restriction enzymes and ligation of sequencing adapters containing in-line barcodes. Following ligation, double size selection of barcoded fragments was performed using Agencourt AMPure XP magnetic beads (Beckman Coulter) (DeAngelis, Wang et al. 1995), fragments larger than 1000 bp and smaller than 200 bp were removed. Illumina TruSeq LT compatible index sequences were then attached to barcoded fragments via PCR. After post-PCR clean-up using Agencourt AMPure XP magnetic beads (Beckman Coulter) DNA concentrations were quantified with a Qubit® 2.0 Fluorometer (Life Technologies) and standardized. Libraries were pooled with each individual contributing 25 ng of DNA, pooled library size selection was performed via an agarose gel cut with fragments between 400 and 500 bp being targeted. DNA from the excised gel were extracted using a Wizard SV Gel and PCR Clean-Up System (Promega). Final library concentrations were quantified with a Qubit 2.0 fluorometer and concentrated if needed using an Eppendorf Concentrator 5301. Size selection and concentration of libraries were confirmed using Agilent ScreenTape assay and qPCR. A total of nine adult lobster libraries were sequenced on the Illumina HiSeq2500 platform with single end 100 bp HT chemistry kits at the Australian Genome Research Facility in Melbourne.

Bioinformatic analysis

Preliminary processing of the ddRADseq data

The quality of sequences obtained was assessed using the FastQC v.0.10.1 quality control tool (Andrews 2010). Reads were then trimmed to 95 bp using TRIMMOMATIC v 0.3 (Bolger et al. 2014). Trimmed reads were filtered for possible contamination using Kraken v.3.5.0 (Wood et al. 2014) and were further demultiplexed using Stacks v.1.37 module *"process_radtags"* (Catchen et al. 2011).

A reference catalog (a list of the most uniformly occurring sequences throughout the dataset) was built using a Rad-loci pipeline. The Rad-loci pipeline developed at La Trobe University (https://github.com/molecularbiodiversity/rad-loci) (Villacorta-Rath et al. 2016) was also

used for reference catalog compilation. This pipeline involves the clustering of sequences, which was performed using VSearch v.1.1.3 (Rognes et al. 2016), with the minimal occurrence of a unique sequence set equal to the number of individuals sequenced, which promotes the selection of loci with lower levels of missing data. To merge similar reads that may have occurred from sequencing errors, the clustering identity was set to 94%, meaning a maximum 7 bp mismatch for 95 bp fragments. The clusters obtained were filtered based on the number of members in the cluster: clusters with less than 2 members were removed as they were considered uninformative, clusters with more than 16 members were also removed to avoid multi-mapping. In the next step, sequences were aligned onto the prepared catalog to estimate the number of alleles for each cluster and the catalog was re-filtered to retain clusters with 2 to 16 alleles. Individual sequences were mapped onto the re-filtered catalog once again and loci with more than 40% missing data per locus were removed. Here "locus" is used to describe a unique reference sequence 95 bp long between the EcoRI and AciI cut sites with a cluster of mappable sequence reads.

SNP calling and characterization

The in-house and Rad-loci pipeline only create a catalog of reference loci and therefore mapping of the individual reads onto the catalog and SNP calling relies on a third-party software. For both of these pipelines demultiplexed, quality and contamination filtered individual reads were mapped onto the reference catalog using the Burrows-Wheeler Aligner v 0.7.13 (Li et al. 2010). Reference catalog loci with a minimum coverage of 10x were used in an effort to minimize the effect of allele dropout (ADO). SNP calling was performed by two independent programs: Genome Analysis Toolkit (GATK) v 3.5.0 (McKenna et al. 2010) and VarScan v 2.4.1 (Koboldt et al. 2009). Only SNPs consistently called by both programs were selected using SelectVariants module of GATK software. SNPs in concordance identified by GATK and VarScan were further filtered using VCFtools v 0.1.14 (Danecek et al. 2011).

For all three pipelines non-biallelic loci were removed with the *--min-alleles* 2 *--max-alleles* 2 options and only one SNP per locus was retained using the *--thin* 95 option. Filtering of rare alleles was performed using the minor allele frequency filtering option *--maf* set to 0.1. The Linkage Disequilibrium (LD) filter was used to minimize the occurrence of linked loci (option *--min-r2*) with $r^2 > 0.8$ threshold in addition to retaining only one SNP per locus.

Individuals and loci with more than 20% missing data were also removed from any further analysis. In addition to commonly used metrics for pipeline comparison (number of loci, number of SNPs), a filtered set of SNPs was further divided into two panels according to calculated F_{ST} value compared to neutral distribution (Excoffier et al. 2010) – neural loci and loci putatively under selection (further referred to as outlier loci). One of the most commonly used methods is based on global F_{ST}, which is employed to identify genetic markers with higher genetic distance than expected in a neutral distribution. A number of demographical scenarios have been shown to influence these estimations and result in type I (false positive) and type II (false negative) errors (Excoffier et al. 2009).

Dependent on the level of selection, mutation rate and hierarchical structure of the populations, different implementations of simulation methods may perform better than others (Narum et al. 2011) and in order to obtain the most conservative estimation, markers under selection were identified with several F_{ST} outlier tests. BAYESCAN v2.1 (Perez-Figueroa et al. 2010) was used with 10 000 iterations and a burn-in of 200 000 steps. A Prior model of 10 000 was used in order minimize the number of false positives (Lotterhos et al. 2015). Arlequin v3.5.2.2 (Excoffier et al. 2010) runs were implemented using 10 000 simulations and 1000 demes. 500 000 simulations and a false discovery rate of 0.1 were used when running Lositan (Antao et al. 2008). A Q-value of 0.05 was used in BAYESCAN runs and the p-value was set at 0.05 in Arlequin and Lositan runs. Only SNPs consistently identified by all three programs were used for further analysis with all three pipelines. Furthermore, the outlier panel was divided into loci under balancing and divergent selection, only loci under divergent selection were retained for analysis.

Population genetics analysis

Preliminary analysis was performed treating each sampling location as a separate population. All population genetics analyses were performed for neutral, neutral+outlier and outlier panels. F-statistics, observed (H_o) and expected heterozygosity (H_E) were calculated using Arlequin v3.5.2.2 (Excoffier et al. 2010) and the R package diveRsity (Keenan et al. 2013). A two-sample Wilcoxon test was used to detect whether mean H_o differed significantly from mean H_E for both SNP panels. Population structure was analysed using a Bayesian model-based clustering algorithm implemented in STRUCTURE (Pritchard et al. 2003). In addition, a non model-based multivariate analysis, Discriminant Analysis of Principal Components (DAPC), was implemented in the R package adegent (Jombart et al. 2011).

DAPC utilizes two statistical approaches: Principal Component Analysis (PCA) and Discriminant Analysis (DA). Initially, multi-genotype SNP data is transformed via PCA into a set of linearly uncorrelated variables or principal components (PC), which further undergoes dimensionality reduction using DA. Discriminant Analysis is a multivariate method capable of distinguishing between-group and within-group genetic variation. DA produces discriminant functions that maximize the genetic differentiation between groups and minimize the within-group variation. In a final step, the probability of membership in each of the inferred clusters is calculated for each individual based on the retained discriminant functions. These membership probabilities are a measure of "genetic proximity" of the individual to the obtained clusters (Jombart et al. 2010; Jombart et al. 2011; Roullier et al. 2013). Correlations were estimated between pairwise oceanographic distances and pairwise genetic distances using the Mantel test function from the R library *adegenet* (Jombart et al. 2011), and the significance of the test was assessed using 10 000 permutations. oceanographic distance was calculated as the sum of the shortest possible coastal distances between each pair of populations in R using the geo*sphere* package.

Results

Genotyping results

A total of 734 adult lobster samples (including technical replicates) were sequenced across 9 ddRADseq libraries and 15 samples were removed from the dataset as they had more than 20% missing data. One unique individual per two or more replicates used for technical validation in Chapter 2 was included in the final dataset for population genetic analysis (**Table 3.2**).

Table 3.2. Final counts of individuals in the datasets analysed.

	Number of individuals
Total sequenced	734
Excluded	15
Replicates	62
Population analysis	644

The average number of sequence reads per individual across all 11 libraries was close to 3 million with 4 to 5 % being removed as reads resulting from contamination, and less than 3% removed due to low quality or ambiguous barcodes/RAD-tags. After filtering, the reference catalogue retained more than 9500 loci. From the reads mapped onto the catalogue 2336 SNPs were consistently identified by GATK and VarScan. Comprehensive statistics of all data processing and filtering stages are presented in **Table 3.3**.

Table 3.3. Stages of sequencing data filtering.

Sequence filtering			Average number of reads per individual	2969221
			Average percentage of contamination-filtered reads	3.9
			Average percentage of low quality and ambiguous barcode reads	2.5
Reference	Catalogue	filtering	Total loci	23169
			Filtered loci	9651
	SNP filtering		Called with VarScan	159139
			VarScan filtered	3325
			Called with GATK	211553
			GATK filtered	7597
			VarScan/GATK concordance	2336
			Neutral	1648
			Under balancing selection	561
			Under divergent selection	127

The adult lobster SNPs were ranked according to single loci F_{ST} and separated into putatively neutral markers (73.4%), and those putatively under selection (outliers), which were further divided into markers under balancing selection (18.4%) and under diversifying selection (8.2%). In the case of population genetics analysis, F_{ST} ranking of SNPs has been shown to influence the pattern of differentiation exhibited by the subpopulations (Moen et al. 2008; Therkildsen et al. 2013; Barth et al. 2017), hence markers were divided into two panels. 1648 SNPs were consistently identified by Arlequin v3.5.2.2, BAYESCAN v2.1 and Lositan as

neutral and 127 SNPs were identified as being under divergent selection (referred to as 'outlier' markers) (**Fig. 3.2**).

Fig.3.2. Number of SNPs identified as neutrals (panel A) and outliers (panel B) using three genome scan methods.
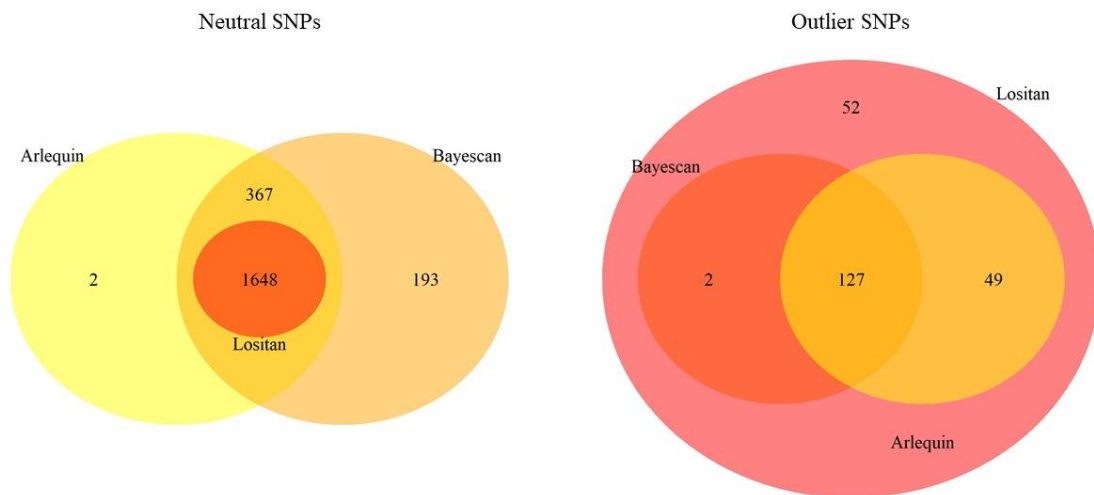
Preliminary population genetics analysis

F-statistics

Strong differentiation between Australia and New Zealand was detected by all three SNP panels (neutral, neutral+outlier and outlier) with the outlier panel showing the highest levels of differentiation and neutral panel showing the lowest. F_{ST} values between NZ populations and Tasmania ranged from 0.013 – 0.02 for the neutral SNP panel and 0.096 – 0.132 for the outlier panel and showed intermediate values for neutral+outlier, ranging 0.052 - 0.07 (**Fig.3.3**). While all SNP panels failed to detect strong differentiation between sampling locations within NZ, the outlier and to slightly lesser extent neutral+outlier panels showed consistent levels of low but significant differentiation between several regions of NZ. The western part of the North Island (statistical areas 935, 939, 901 and 902 – this region will be here after referred to as "NW" region) exhibited the lowest differentiation from Tasmanian samples (0.1 average). The NW region also showed low but significant differentiation from all the other NZ sampling locations. The NW region was differentiated to the same extent from statistical areas located on the East coast of North Island (904, 907, 908, 910, 911, 913,

915), Chatham Is. (942) and the East coast of northern South Island (933, 934, 916, 917), hereafter referred to as "NE" region. Statistical areas 920, 921, 924, 926, 927 and 928 (most southern areas of South Island, hereafter "S" region) also had significant, however low, levels of differentiation from the NE and NW regions.

Fig.3.3. Matrices of pairwise F_{ST} values obtained with the neutral SNP panel (A), neutral + outlier SNPs (B) and outlier SNP panel (C).
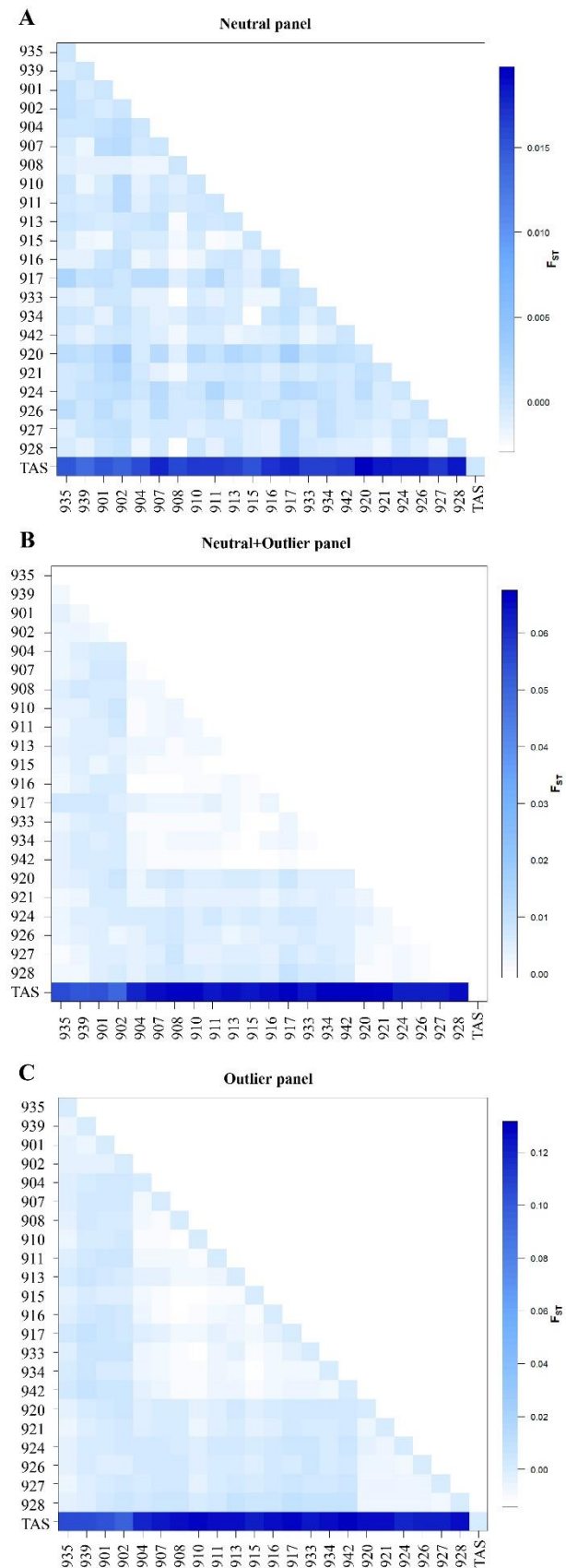
Bayesian clustering analysis with STRUCTURE

Genetic differentiation between the regions identified above was further supported by Bayesian clustering analysis. For all three panels K=3 was determined as most optimal number of clusters by the Evanno method (Evanno et al. 2005) implemented in STRUCTURE HARVESTER. The neutral SNP panel showed strong AUS/NZ clustering and lack of differentiation within NZ (**Fig.3.4** panel A). Clustering produced with the outlier panel found a TAS cluster (red), a NE cluster (mostly green) and a NW + S cluster (mostly blue) (**Fig.3.4** panel C), with slightly higher admixture of characteristic for TAS genotypes at multiple loci in populations in the NE cluster compared to the rest of the NZ sampling sites. A very similar pattern of genotypes at multiple loci was produced by a combination of neutral and outlier SNP markers: clear differentiation of Tasmanian samples and two clusters within NZ (NE and NW+S) (**Fig.3.4** panel B).

Fig.3.4. Bayesian clustering analysis of genetic differentiation among 644 individuals. Panel A are the results obtained with the neutral SNP panel, panel B – neutral and outlier SNPs combined, panel C – the outlier SNP panel.
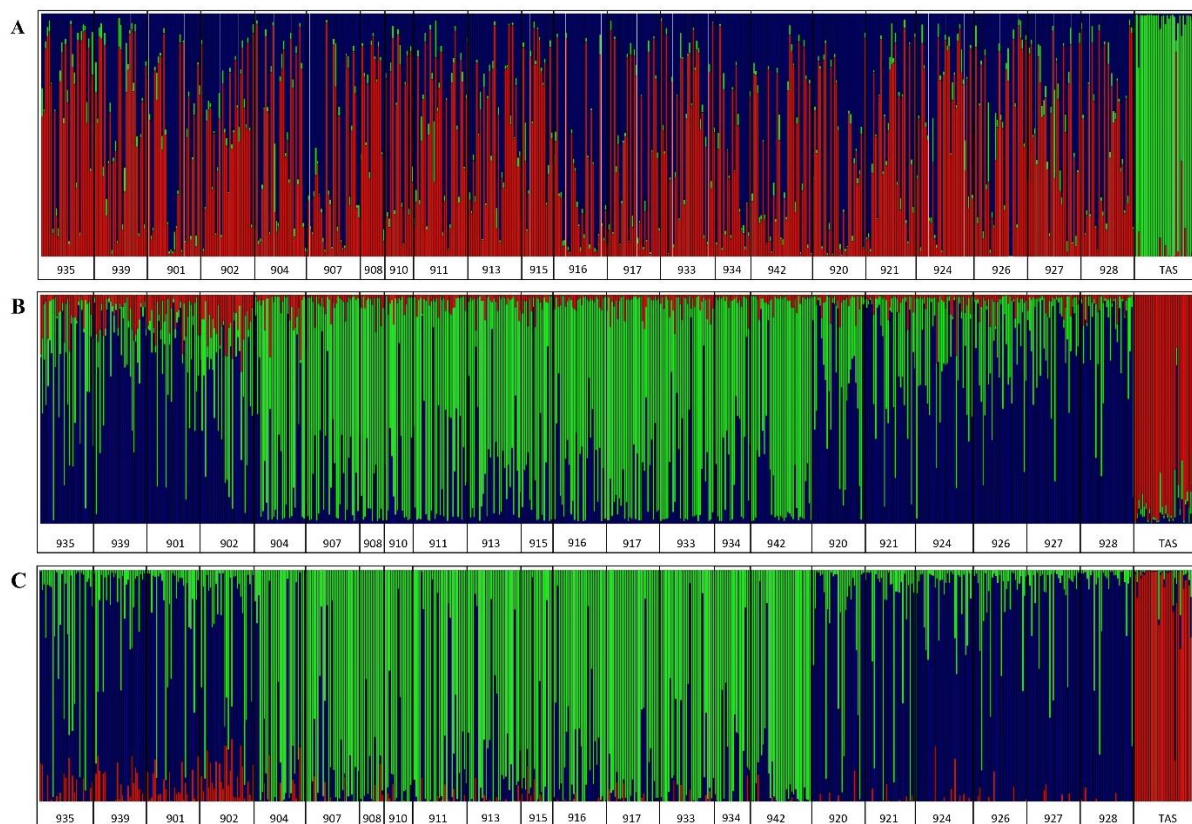
Discriminant Analysis of Principal Components

DAPC analysis using the neutral and outlier panel produced four clusters and neutral+outlier panel produced three clusters with 1 of the clusters being comprised solely of samples from Australia in all the panels (**Fig.3.5**). However, the neutral panel had low admixture of TAS samples in the three other clusters representing mostly samples from NZ (**Fig.3.5,** panel A), while neutral+outlier panel produced admixture of 0.3% of TAS samples in the Cluster 3 (**Fig.3.5,** panel B). No admixture of samples from Australia was detected using the outlier panel. In addition, each of the NZ-only clusters identified with the outlier panel predominantly contained samples from one of the previously identified regions: 89% of samples in Cluster 2 were from NW region (shades of orange), 66% of samples in Cluster 3 were from the S region (shades of blue) and 81% of Cluster 4 contained samples from the NE region (shades of green) (**Fig.3.5,** panel C). In contrast, the DAPC for the neutral markers produced clusters containing similar proportions of all 3 regions of NZ (shades of orange – NW region, shades of green – NE region, shades of blue – S region) with admixture of TAS samples (**Fig.3.5,** panel A). Clustering obtained with neutral and outlier markers combined produced intermediate clustering pattern: only two NZ identified but they represent the regions previously identified: 80% of Cluster 1 are samples from NE region and Cluster 3 consists mainly of samples from NW region (34%) and from S region (52%) (**Fig.3.5,** panel B).

Fig.3.5. Discriminant Analysis of Principal Components (DAPC) of genetic differentiation among 23 sampling locations. Panel A clustering is produced from the neutral SNP panel, panel B is produced from neutral and outlier SNPs combined. Panel C clustering is produced from the outlier SNP panel.

CLUSTER 2

939 18%
901 29%
935 9%
926 5%
924 2%
917 1%
915 2%
904 1%
902 33%

CLUSTER 3

928 12%
935 8%
939 7%
901 3%
902 1%
904 3%
907 2%
911 3%
913 2%
915 1%
916 2%
933 2%
942 1%
920 11%
921 10%
924 11%
926 10%
927 12%

CLUSTER 1

TAS 100%

1

CLUSTER 4

901 1%
924 2%
926 1%
935 2%
902 2%
904 8%
920 1%
921 2%
942 11%
907 8%
933 8%
908 5%
917 10%
910 5%
916 9%
911 7%
915-34 10%
913 8%

C
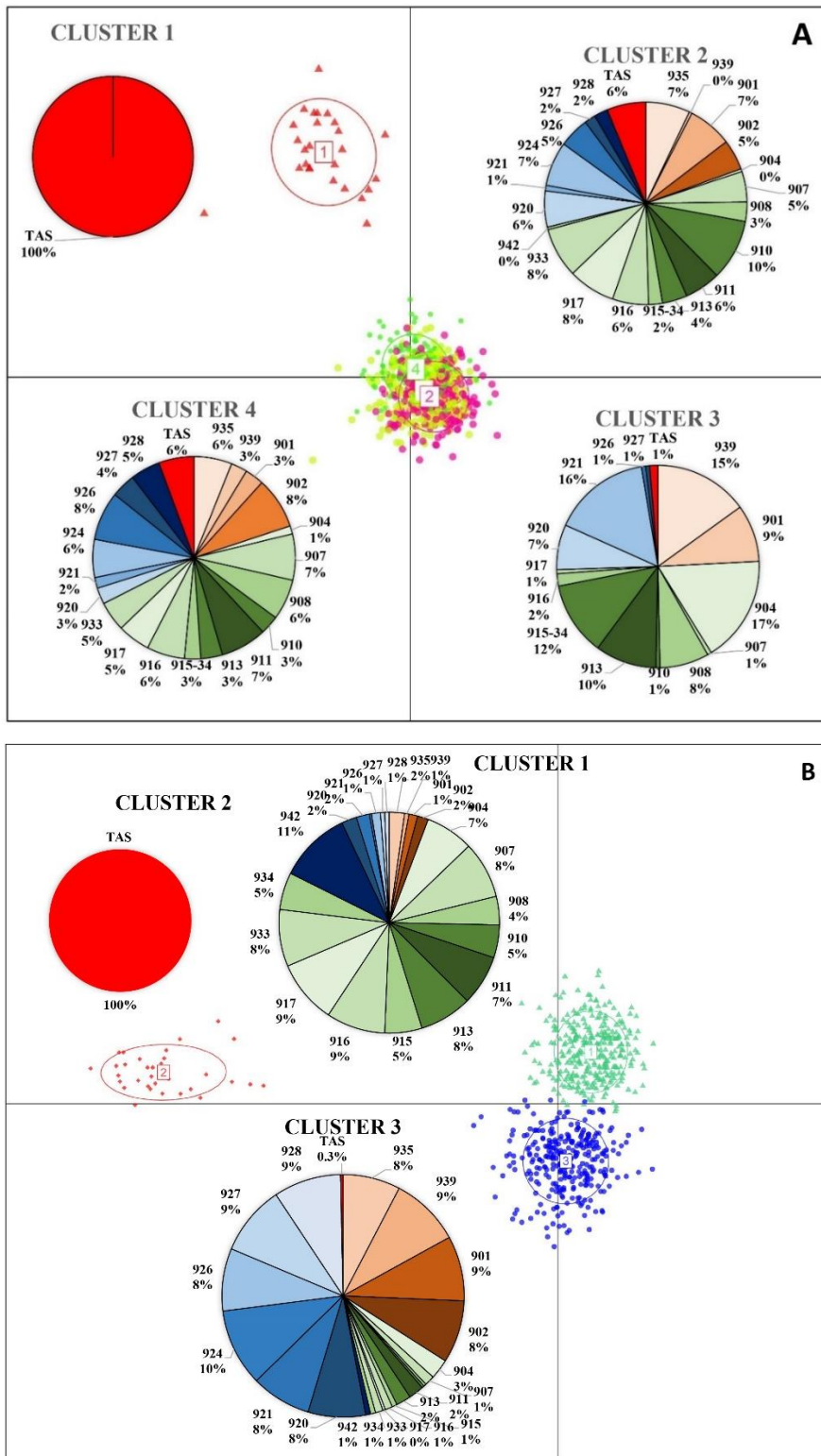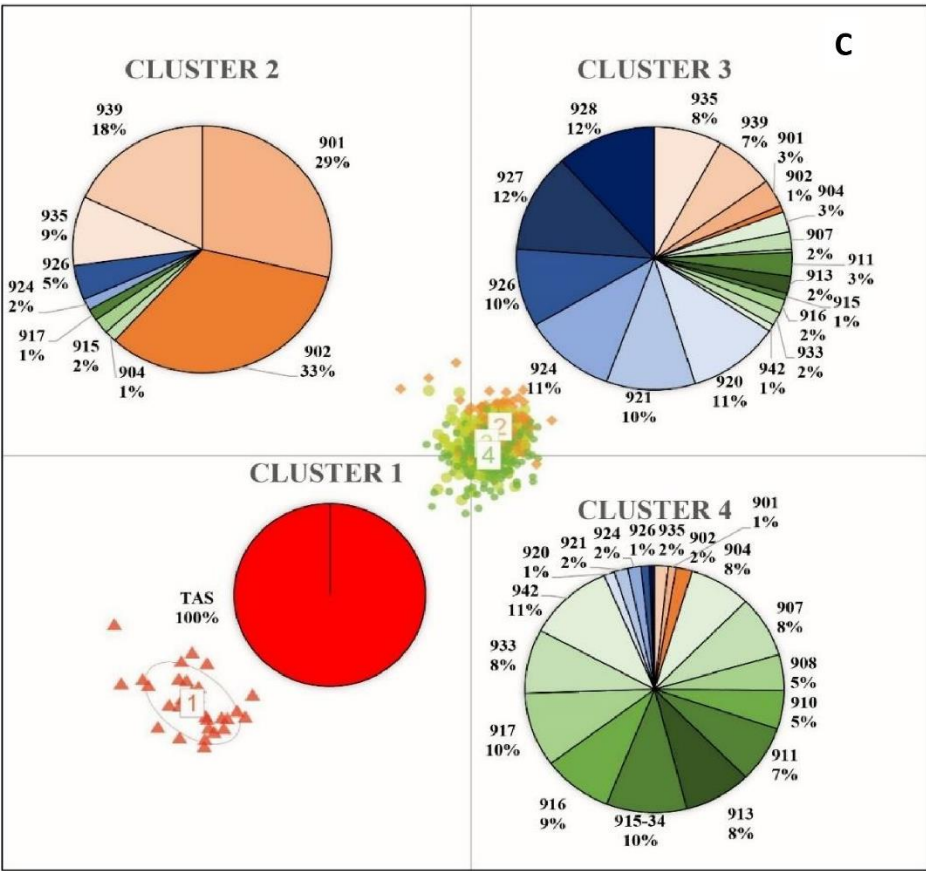
Isolation by distance

The correlation between genetic divergence and oceanographic distances was slightly negative (R = -0.092), but not significant (p-value = 0.774) when tested using neutral SNP markers (**Fig. 3.6. Panel A**). In contrast, a strong (R = 0.519) highly significant (p-value = 0.001) correlation was detected using outlier markers, indicating strong isolation by distance (**Fig. 3.6. Panel C**). Similarly to other types of analysis, neutral and outlier markers combined show intermediate trend between the two: R = 0.073, p-value = 0.257 (**Fig. 3.6. Panel B**). This pattern suggests that the selective force acting on *Jasus edwardsii* across its distribution range is not a distance between the populations *per se* (as shown by the lack of IBD pattern for the neutral markers), but rather an environmental factor or combination of factors that vary with distance across NZ and produce a strong IBD pattern seen when testing with outlier SNP markers.

Fig. 3.6. Correlation between pairwise genetic (F_{ST}) and oceanographic distances. Linear regression lines (blue) are fitted with a 95% confidence interval (grey). Panel A produced from neutral markers, p-value = 0.774. Panel B produced from neutral and outlier markers combined, p-value = 0.257. Panel C produced from the outlier markers, p-value = 0.001.

B

R = 0.073

C

R = 0.519
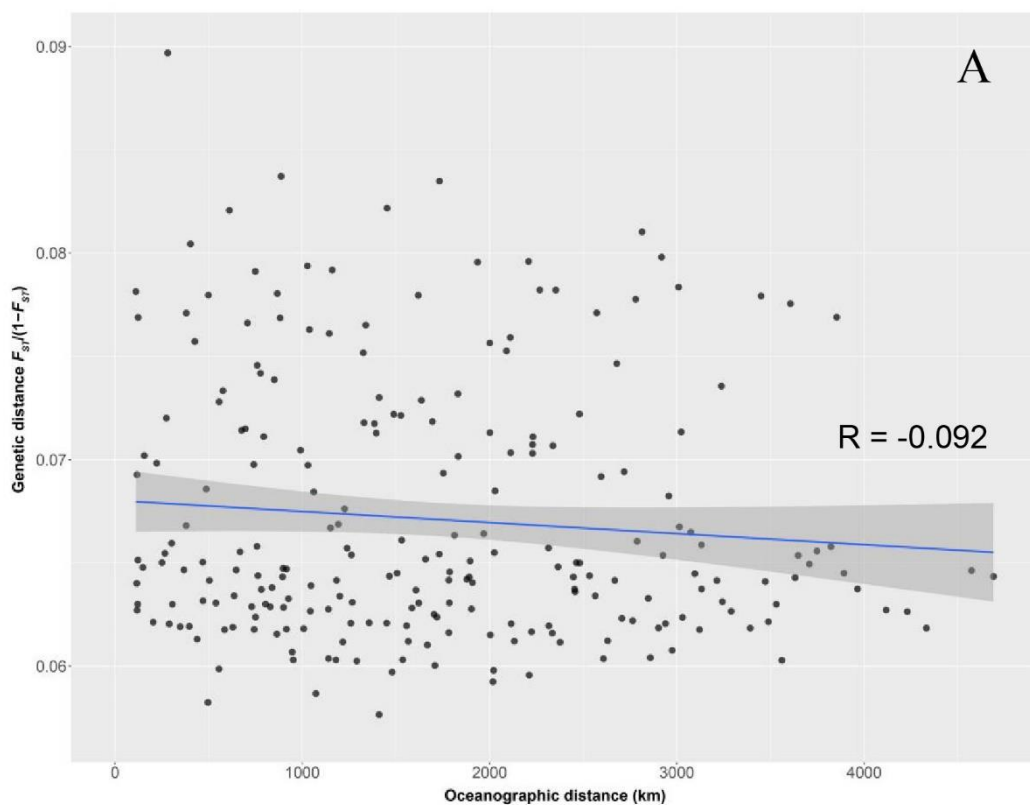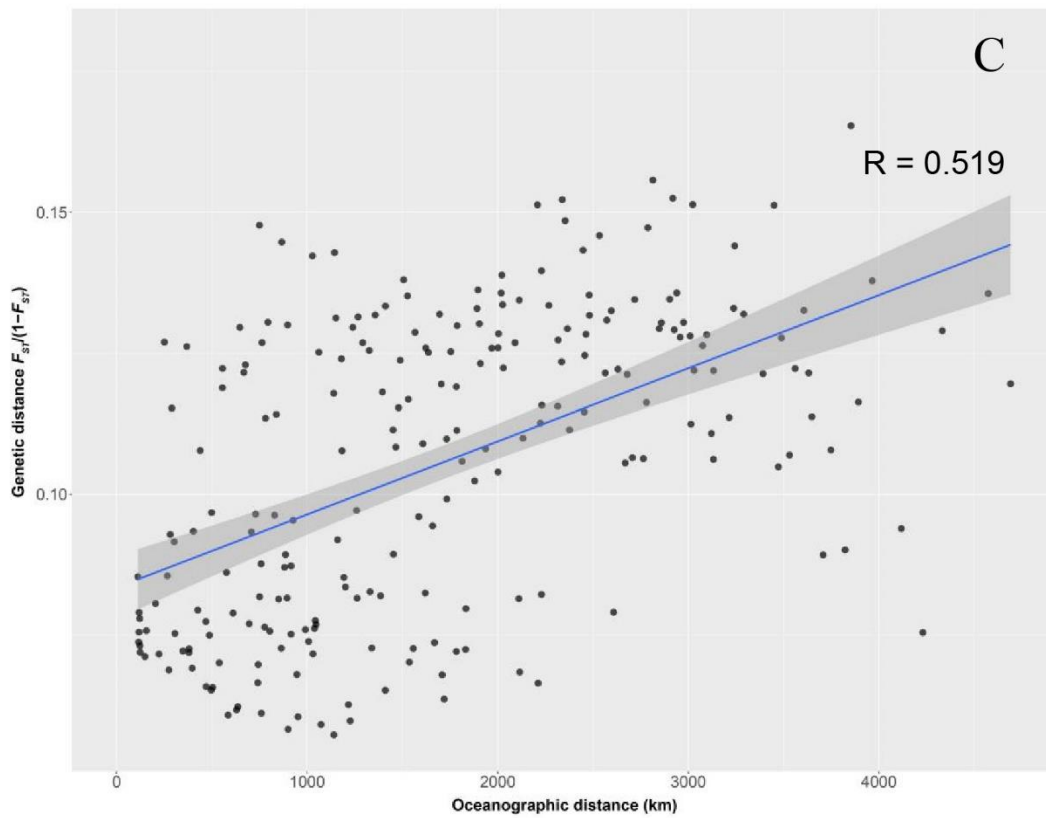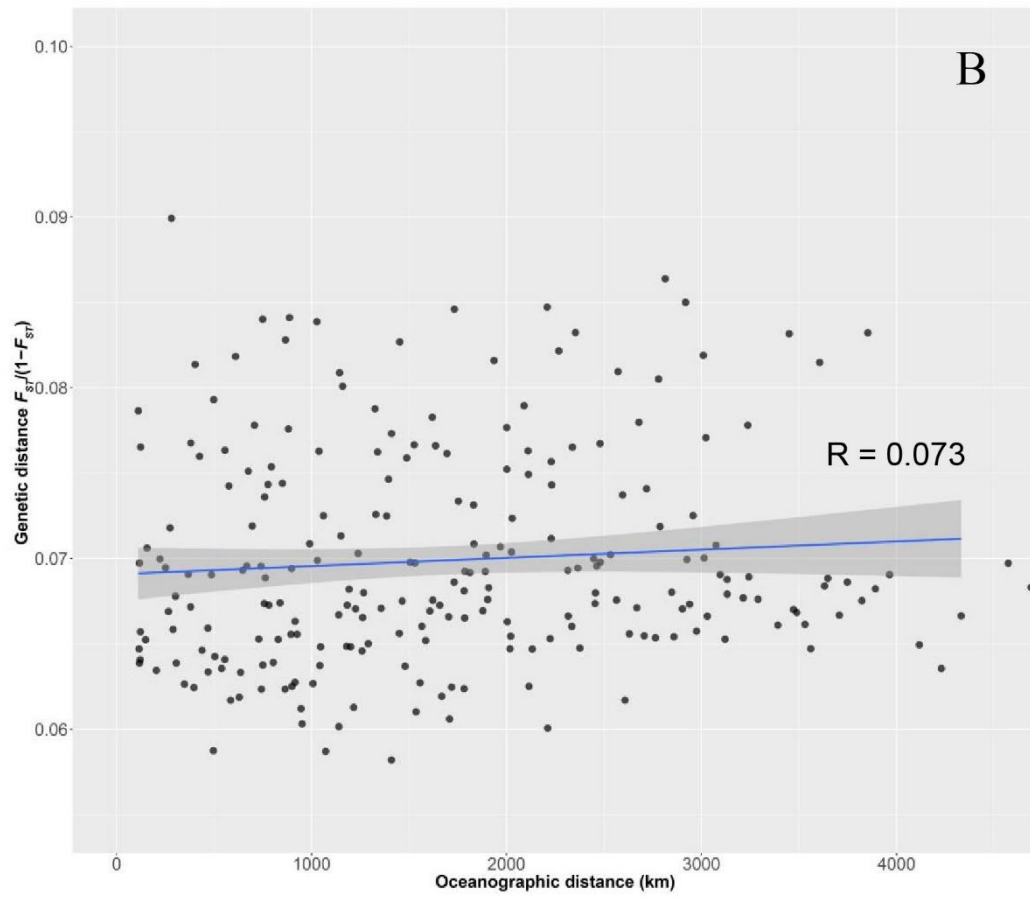
Analysis of metapopulations identified

A full analysis (F-statistics, Bayesian clustering analysis and DAPC) was repeated on NZ samples only and population differentiation was assessed between previously identified clusters of samples:

- NW region (stat areas 935, 939, 901 and 902) – colored red;
- NE region (stat areas 904, 907, 908, 910, 911, 913, 915, 942, 916, 917, 933 and 934) – colored blue;
- S region (stat areas 920, 921, 924, 926, 927, 928) – colored green.

Fig. 3.7. Map of metapopulations identified. Statistical areas belonging to the NW region shown in red, belonging t the NE region in blue and belonging to the S region in green. Map adapted from NZ Seafood Industry Council report, 2005.
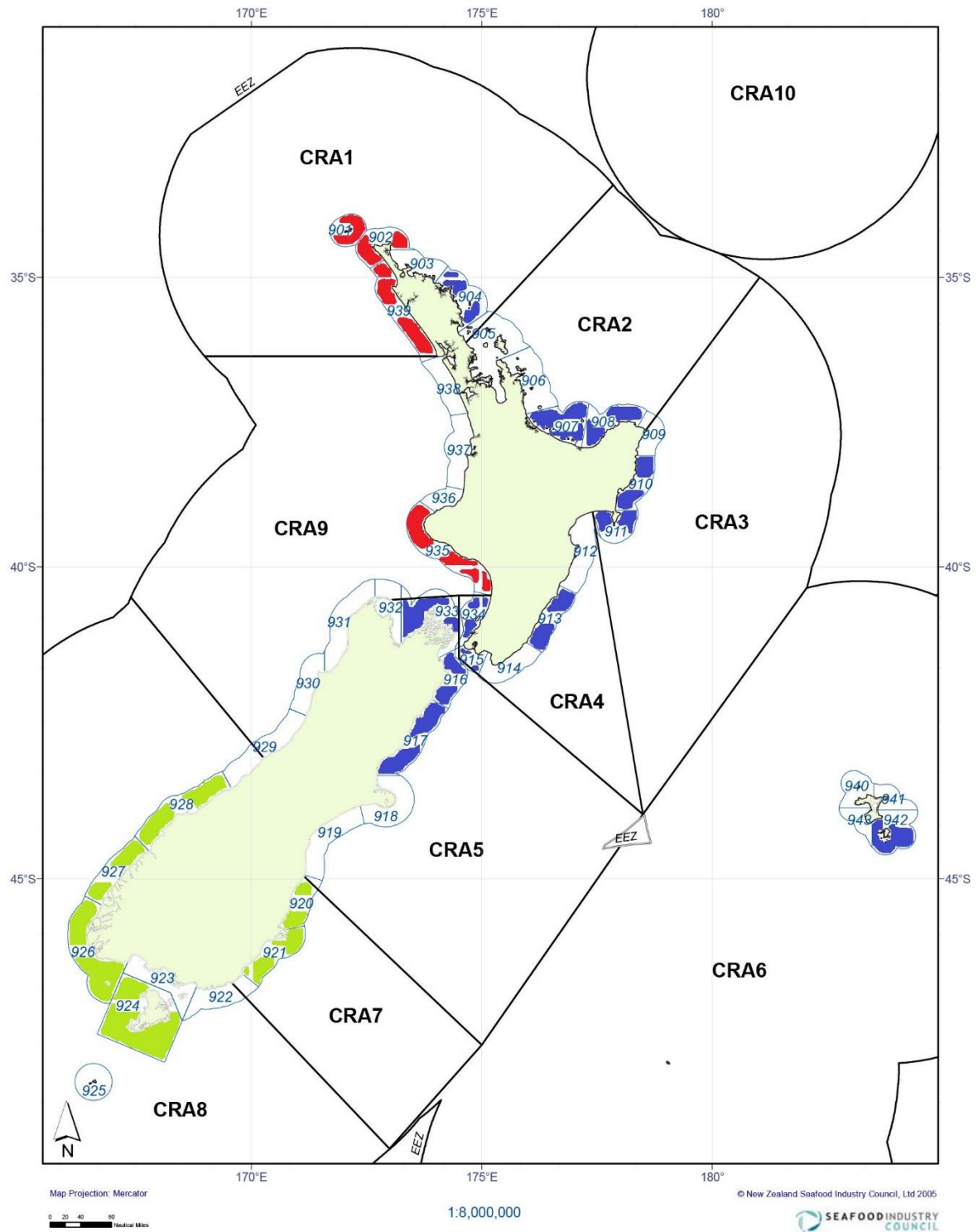
F-statistics

Pairwise F_{ST} values calculated with all three SNP panels were highly significant (**Table 3.4**). F_{ST} values produced with the outlier SNP panels confirmed genetic differentiation between the previously observed regions (NW, NE and S) and indicated a lower level of division between the NW and S regions compared to separation of the NE from both the NW and S regions. Despite being significant, F_{ST} values obtained using the neutral markers were very low and most likely do not reflect ecologically relevant population differentiation. F_{ST} values obtained using the combined set of markers produced significant results with values being intermediate between purely neutral or purely outlier panels of SNPs.

Table 3.4. Pairwise F_{ST} values calculated after pooling of the statistical areas into subpopulations. Results are presented for neutral, outlier and neutral +outlier panels separately. Values in bold are statistically significant at p<0.0001.

		NW	NE	S
Neutral panel	NW	0		
	NE	**0.0001**	0	
	S	**0.0008**	**0.0006**	0
		NW	NE	S
Outlier panel	NW	0		
	NE	**0.0488**	0	
	S	**0.0253**	**0.049**	0
		NW	NE	S
Neutral + Outlier panel	NW	0		
	NE	**0.004**	0	
	S	**0.003**	**0.004**	0

Inbreeding coefficients F_{IS} obtained using the neutral and neutral+outlier SNP panels were very similar and ranged from -0.02 to - 0.28 (**Table 3.5**). The outlier panel produced negative values of F_{IS}, which ranged from -0.028 to -0.036.

Table 3.5. Descriptive statistics calculated after pooling of the statistical areas into subpopulations. Results are presented for neutral, outlier and neutral+outlier panels separately.

Region	Neutral panel			Outlier panel			Neutral + Outlier panel		
	H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}
NW	0.405 (±0.188)	0.367 (±0.104)	- 0.027	0.394 (±0.125)	0.379 (±0.097)	- 0.036	0.367 (±0.138)	0.356 (±0.103)	- 0.028
NE	0.403 (±0.186)	0.368 (±0.103)	- 0.02	0.352 (±0.150)	0.342 (±0.128)	- 0.029	0.362 (±0.136)	0.354 (±0.104)	- 0.021
S	0.404 (±0.189)	0.367 (±0.105)	- 0.023	0.352 (±0.142)	0.343 (±0.121)	- 0.028	0.362 (±0.138)	0.354 (±0.105)	- 0.023

Discriminant Analysis of Principal Components

Analysis of the neutral panel identified two clusters, however both of them contained individuals from all three regions in similar proportions (**Fig.3.7**, panel A). DAPC analysis using the combined set of SNPs also produced two clusters, however, Cluster 1 is mostly represented by the NE samples and Cluster 2 is comprised of NW and S samples (**Fig.3.7**, panel B). Analysis performed using outlier panel produced very similar results to those observed previously by Thomas and Bell (2013): clustering individuals into three groups representing NW (Cluster 1), NE (Cluster 3) and S regions (Cluster 2), with admixture from individuals from other regions (**Fig.3.7**, panel C), although the boundary between NW and NE region appears to lie further North than in microsatellite study.

Fig.3.8. Discriminant Analysis of Principal Components (DAPC) of genetic differentiation among 3 regions. Panel A represents clustering produced by the neutral SNP panel, panel B – clustering produced by the neutral and outlier SNPs combined, panel C – produced by the outlier SNP panel.
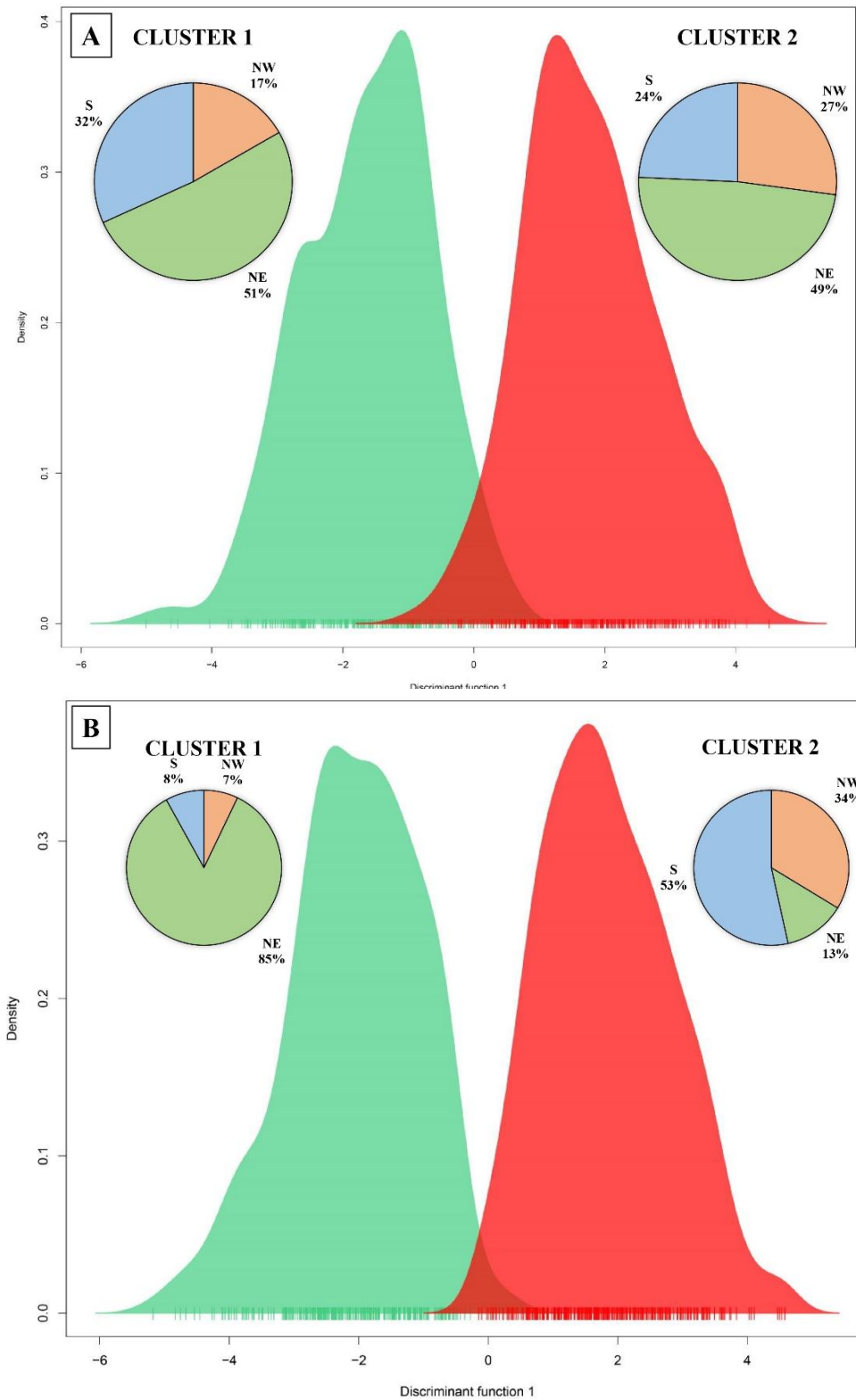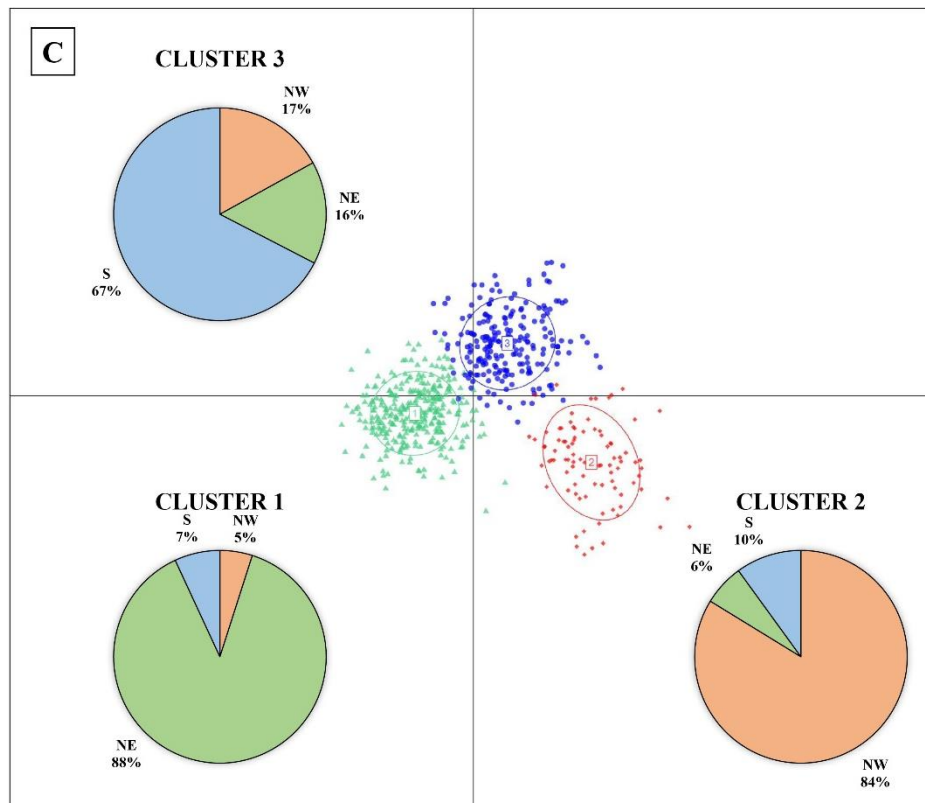
Results of the DAPC analysis using outlier markers represented as compoplot (a bar plot of the membership assignment probability of individuals to the identified clusters) (**Fig.3.9**) suggests the presence of a cline along the West coast of NZ with NW and S regions being the extremes and statistical areas 935 and 939 being intermediates. The proportion of individuals belonging to the "green", NW cluster is the highest and decreases with proximity to the "red", S cluster. At the same time, proportion of individuals belonging to the S cluster, which is lowest in statistical area 902 increases further along the cline running along the NZ West coast.

Fig.3.9. DAPC results obtained with outlier markers viewed via compoplot graphic function of the adegenet package showing the probabilities of assignment of individuals to the clusters identified.
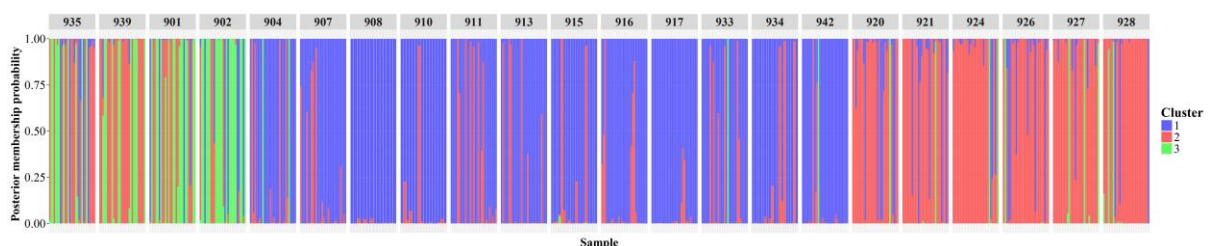
Bayesian clustering analysis with STRUCTURE

Bayesian clustering was highly consistent with the previous DAPC results and the F-statistics. The Evanno method identified the highest likelihood of K=2 for neutral and neutral+outlier panels and K=3 for the outlier panel. The outlier markers identified NZ lobsters as being comprised of 3 subpopulations: NW (mostly green), NE (mostly blue) and S (mostly red) (**Fig.3.10,** panel C). Interestingly, the lower levels of differentiation between the NW and S regions reflected in the F_{ST} values appear to be a result of gene flow from the North to South Island along the West coast combined with an intermediate pattern of selective pressure between the NW and S regions, which can be seen in the higher admixture of genotypes belonging to the red set of markers in stat area 935 than in the rest of NW region. Analysis of neutral panel failed to detect any population structure (**Fig.3.10,** panel A), while combined set of markers was able to distinguish between NE (mostly green) and NW+S regions (mostly red) (**Fig.3.10,** panel B).

Fig.3.10. Bayesian clustering analysis (STRUCTURE v.2.3.4) of genetic differentiation between 3 regions. Panel A represents clustering produced by the neutral SNPs, panel B is produced by neutral and outlier SNPs and panel C is produced by outlier SNPs.
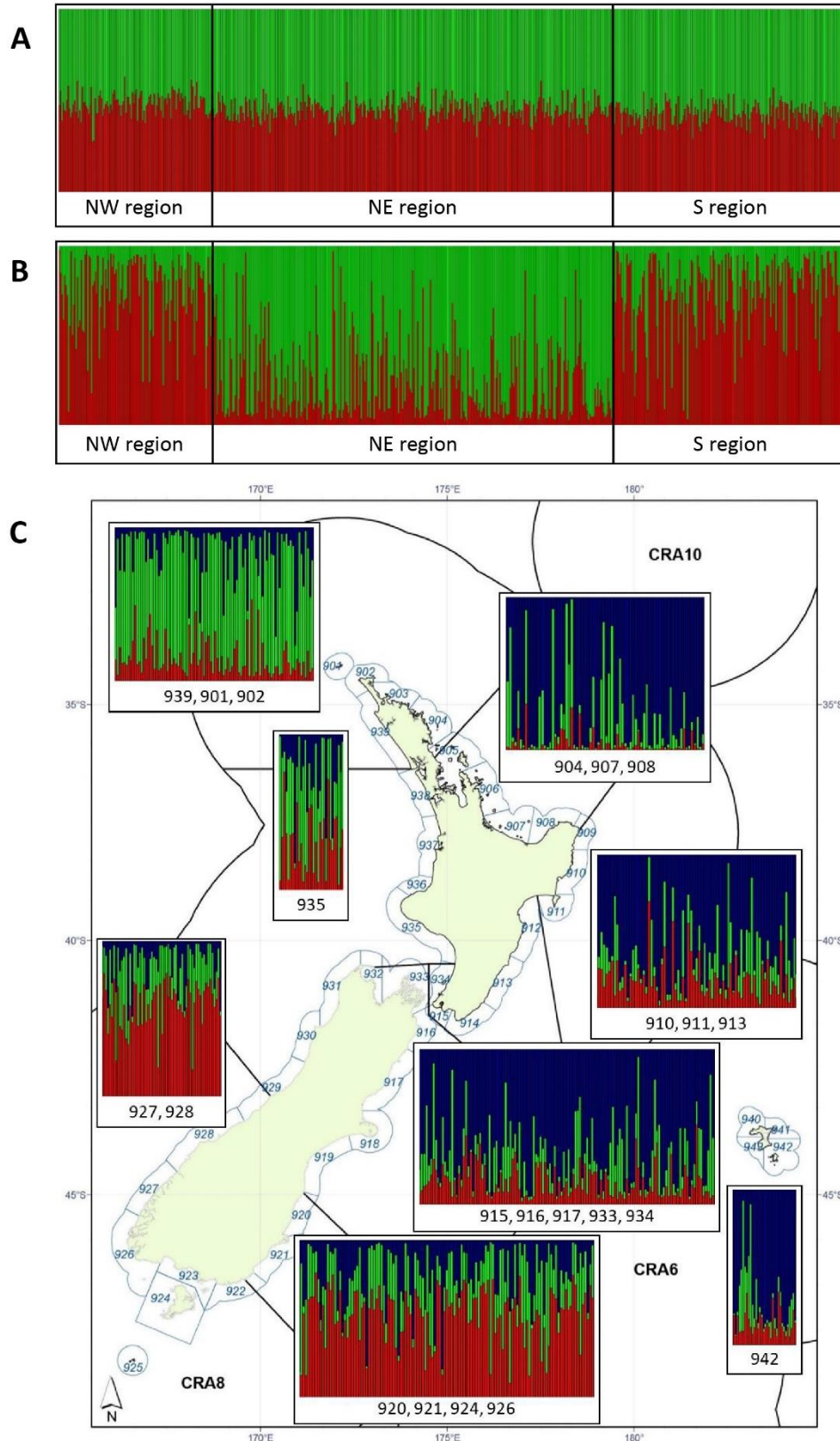
Discussion

In this chapter I aimed to assess population differentiation patterns of *Jasus edwardsii* using neutral and outlier Single Nucleotide Polymorphism markers. Results from the analysis of the neutral SNP data, supported by earlier oceanographic models (Chiswell et al. 2008; Chiswell et al. 2011), suggest effective mixing of larvae throughout the NZ distribution range of *J.edwardsii*. Despite high population connectivity and substantial levels of gene flow, outlier SNP markers indicated a significant level of local selection capable of counteracting the effect of migration, potentially via selective mortality of less adapted genotypes. Analysis of outlier markers consistently identified three genetically different regions within NZ: North-East, North-West and South sub-populations. Here, I discuss evidence for the contribution of local selection to the population structure of red rock lobster despite high levels of demographic connectivity.

Population structure inferred from neutrals and outlier markers

Neutral SNP markers failed to detect any population sub-division within NZ using a number of different analyses, which reflects unimpeded gene flow that homogenized the allele frequencies at neutrally evolving genetic loci (Hellberg et al. 2002). In contrast, markers under selection identified three broad regions within NZ: North – West of the North Island, North – East of North and South Islands including Chatham Islands, and the most Southern part of the South Island. Analysis of the combined set of markers consistently produced an intermediate pattern of differentiation of purely neutral and outlier sets of markers and this effect of reduction in resolving power of purely outlier SNPs have been previously reported in Atlantic Cod (*Gadus morhua L.)* (Berg et al. 2015). However, study of Atlantic Cod demonstrates that adaptation to the selective pressure of the environment appears to drive neutral population structure via mechanisms of "isolation by adaptation" or "isolation by environment". In case of red rock lobster selective pressure is likely to be not sufficient to affect genetic diversity of the neutral markers. Patterns of differentiation seen in outlier markers has been reported to originate from local selection (Seeb et al. 2011), which could be related to a single, or a combination, of environmental factors. Patterns of weak selection reflected in neutral markers compared to strong differentiation at markers under selection have been observed in many marine organisms with high levels of gene flow and connectivity between populations: Eulachon (Candy et al. 2015), cod (Pocwierz-Kotus et al. 2015; Barth

et al. 2017), salmon (Freamo et al. 2011; Seeb et al. 2011; Ozerov et al. 2013), herring (Lamichhaney et al. 2012), Pacific lamprey (Hess et al. 2013), Australian sea mullet (Krück et al. 2013) and oysters (Nayfa et al. 2016). The availability of a reference genome for cod allowed some of the selective markers to be mapped to a large region ~5 Mb of chromosomal rearrangement, which is characterized by a low recombination rate and is suspected to contain genes critical for survival at low salinities. Genomic islands of divergence occurring in the regions of reduced recombination is one of the main mechanisms of adaptive divergence of local populations and although it is likely to be responsible for population differentiation of *Jasus edwardsii* this cannot be confirmed until a reference genome is available.

In addition to the existence of three genetically distinct regions within NZ, multiple tests (F_{ST}, DAPC and Structure) actually suggest a cline of adaptive variation along the West coast of NZ. The proportion of genotypes adapted to the Northern-most environmental conditions decreases while admixture of Southern-most genotypes increases along a North-South gradient; this warrants further investigation into the selective forces affecting local populations.

Interestingly, when tested for isolation by distance, neutral and outlier markers produced contrasting correlation patterns. A slightly negative correlation between genetic differentiation and oceanographic distance was detected with the neutral markers, which supports the hypothesis of high gene flow throughout NZ, most likely as a result of a combination of long PLD and effective transport via oceanic currents. However, genetic differentiation of the outlier markers strongly correlated with oceanographic distance. Patterns of isolation by distance investigated using neutral and outlier SNPs combined shows some similarity to the previously reported pattern of IBD investigated using eight microsatellite markers. Thomas & Bell (2013) reported weak but significant positive correlation between F_{ST} and oceanographic distance: R = 0.16, p-value = 0.037. Here I report R = 0.076, p-value = 0.257. Re-analysis of microsatellite data with a higher number of iterations (500 000 here vs 50 000 in Thomas & Bell (2013)) identified one of the microsatellite as being putatively under selection (**Appendix 1**). The inclusion of a neutral marker linked to the genomic region under selection can lead to the skewed estimation of population differentiation and connectivity (Selkoe et al. 2006) and might have contributed or enhanced the observed pattern of IBD. Oceanographic distance *per se* cannot be viewed as a selective factor, it can influence the population differentiation via impeding gene flow, thus,

IBD pattern of differentiation of markers under selection can only be attributed to a single or combination of environmental factors that vary with oceanographic distance. This conclusion is further supported by the evidence of the cline of genetic differentiation along the NZ West coast, which likely spans several environmental gradients. My results suggest that despite the long PLD of *Jasus edwardsii* supporting high levels of gene flow, the effective migration/reproductive success of migrants to the new environment is reduced, most likely due to a phenotype – environment mismatch causing selective mortality.

Comparison of population structure inferred with SNP markers to earlier studies

Several earlier studies have studied genetic differentiation in *Jasus edwardsii*, with the earlier studies reaching the same conclusion of a single panmictic stock existing across the range of the species (Smith 1980; Booth et al. 1990; Ovenden et al. 1992). However, panmixia within the NZ stock was contradicted by a more recent study describing finer-scale population structure within New Zealand by Thomas and Bell (Thomas et al. 2013). Eight microsatellite markers produced a weak genetic signal delineating three sub-populations within NZ: a central NZ cluster (comprised of Wellington, Kaikoura, South West coast and Chatham islands), a northern NZ cluster (Hauraki Gulf) and a southern NZ cluster (represented by Stewart Island population).

Population structure produced by local selection should not influence the putatively microsatellite neutral markers, however, a phenomenon known as Isolation by Adaptation (IBA) at neutral loci can reconcile the seemingly opposing results of earlier microsatellites data compared to my SNP data. Essentially, selective pressure can translate into divergence at non-linked neutral loci, which has been reported before in stick insects (Nosil et al. 2008) and stickleback (ÓLafsdÓTtir et al. 2007).

Several mechanisms of how selective pressure might translate into divergence at neutral loci have been described. One hypothesis is genetic hitchhiking – a phenomenon where the frequency of neutral loci can be increased by a selective force if they are linked to a locus that selection is directed at (Bierne 2010; Roesti et al. 2014). The hitchhiking effect on neutral loci reduces with distance to that gene/locus under selection, however this effect can extend to a considerable distance; simulations have shown that strong selection can promote divergence at neutral loci as far as the length of the chromosome (Charlesworth et al. 1997; Kim et al. 2002). The hitchhiking effect strongly depends on the recombination

frequency:selection strength ratio due to the chance of the bond between neutral and selected loci being broken by recombination (Barton 2000; Andolfatto 2001; Kim et al. 2002).

Another mechanism that might explain the discrepancy between earlier microsatellite data and the current SNP results is where a microsatellite marker is generated in a region already under selection in the population, then this neutral marker is likely to show a substantial level of divergence (Nosil et al. 2009; Gagnaire et al. 2015). Supporting this scenario is research pertaining to "genomic island growth", which predicts that regions of genetic differentiation will cluster in proximity of one another rather than being randomly distributed throughout the genome (Gavrilets 2004; Barton 2005; Kirkpatrick et al. 2006). In the majority of cases "genomic islands of divergence" are associated with centromeric, telomeric or rearranged regions of the chromosome where the recombination rate is reduced due to lower chance of forming a chiasma (Turner et al. 2005; Cruickshank et al. 2014). Nonetheless, higher divergence at neutral loci linked to the islands of increased divergence have been documented in *Drosophila,* although islands outside of the regions of reduced recombination rate tend to be much smaller (McGaugh et al. 2012).

Divergent selection can affect completely non-linked neutral loci and promote their divergent differentiation by reducing effective population size and increasing the effect of genetic drift in sub-populations (Nosil et al. 2005; Nosil et al. 2009). Local selection reduces gene flow by causing a lower amount of effective migration due to migrants being selected against if they do not possess the advantageous alleles for survival in the new environment (Charlesworth et al. 1997). In the case of *Jasus edwardsii*, divergent local adaptation at both extremes of New Zealand might create a "genetic barrier" of settlement/post-settlement mortality for the migrants from the central regions of New Zealand.

Comparison of genetic population structure with oceanographic models

The three main sub-populations identified with the microsatellite markers are the largest larval producers and two of them are thought to experience high levels of self-recruitment based on the oceanographic model developed by Chiswell and Booth (2008). Identification of self-recruiting populations is critical for conservation and sustainable harvesting of species due to their greater vulnerability to overfishing and habitat destruction of those populations (Strathmann et al. 2002). Populations relying on local recruits might require greater conservation efforts and are more likely to exhibit higher levels of differentiation due to local

adaptation (Strathmann et al. 2002). The highest level of self-recruitment calculated from Table 3 from Chiswell and Booth (2008) is for South region (CRA7 and CRA8, which corresponds to stat areas 920, 921, 924, 926, 927, 928 in this study) and is estimated to be as high as 94%. The North-West region (CRA1 and possibly CRA9, which are represented by stat areas 939, 901 and 902 in this study) is estimated to be able to self-recruit more than half of the larvae produced there (54.5%). A much broader North-East region that includes CRA2, CRA3, CRA4, CRA5 and CRA6 (corresponding to data from stat areas 904, 907, 908, 910, 911, 913, 915, 916, 917, 933, 934 and 942) are predicted to be able to retain 30.8% of the larvae produced despite two major eddies occurring in the areas (East cape and Wairarapa eddies) that have been shown to be important in larval entrainment (Chiswell et al. 2010). Such a low level of estimated self – recruitment might be due to large amounts of larvae being entrained elsewhere, mainly in CRA8 and CRA1. However, this prediction did not receive any support from either the earlier microsatellite or my SNP data. The microsatellite study showed extremely low admixture of NW genotypes in the S region, similarly the Hauraki Gulf sampling site (although technically located in CRA2) showed substantial level of divergence from the NW region. Outlier SNP markers also show low levels of admixture of NW genotypes in the S region. Genetic markers indicate lower levels of entrainment of larvae outside of their "source" region than is predicted from oceanographic model which might be accounted for by selective mortality at settlement. For example, larvae originating in the NW region and attempting to settle in two regions located at the extremities of the environmental gradient might not have the genotypic/phenotypic potential to survive to adulthood there (Hare et al. 2005). According to the settlement model, larvae from the far South (CRA8) self-recruit and seed CRA9, CRA7 and CRA5, which is only partially supported by the population genetic data with both types of markers. The genetic makeup of populations in stat areas 916, 917 and 933 is virtually identical to the neighboring CRA4 area and more distant CRA3, CRA2 and CRA6 areas. Chiswell and Booth (2008) identified the Chatham Islands as a non-self-recruiting non-local larval sink that also does not contribute migrants to the mainland stock. This assumption was confirmed by the genetic variation study using microsatellite markers and the results presented here are in concordance with this as the genetic make-up of the Chatham Island population was virtually identical to the mainland populations of the NE region. However, the earlier model predicted connectivity of CRA1 (northernmost part of North Island) and CRA6 (Chatham Island), but investigation of genetic diversity did not support the assumption of significant genotype influx in this direction by both microsatellite and SNP studies. The genetic makeup of the Chatham Islands

population has been shown to be genetically homogenic with central NZ populations (CRA4 and 5). Chiswell and Booth's (2008) oceanographic model predicted local and non-local sinks of larvae in the south region of South Island, however the microsatellite study revealed a high level of self – recruitment with a low input of foreign alleles that might be indicative of reduced larval transport and increased mortality at settlement in the region for a period of time that resulted in a level of differentiation observed.

In the case of *Jasus edwardsii*, population stratification was consistent with oceanographic models and was mainly attributed to the direction and strength of the currents (Chiswell et al. 2008), however the extent of the influence of environmental gradients within New Zealand on genetic population structure requires further investigation. Results presented in this chapter raise another important question – on what life stage of *Jasus edwardsii* is selective pressure exerted and what role does it play in the maintenance of genetic diversity of the stock?

Chapter 4

Environmental selection at settlement produces adaptively divergent recruits of the red rock lobster *Jasus edwardsii*

Abstract

Anthropogenic changes to the marine environment test the adaptability of marine animals. Thus, study of the interplay of dispersal and local adaptive divergence is crucial for understanding patterns of population connectivity and the adaptive potential of a species. Genetic differentiation analyses of juveniles of a species can provide insight into possible mechanisms that promote adaptive divergence. This is increasingly being recognised as critical for species conservation and management in the face of climate change and its effect on the distribution range of marine animals. The primary aim of this study was to investigate the levels of connectivity and adaptive divergence of juveniles of the red rock lobster *Jasus edwardsii*. Panels of putatively neutral and putatively under selection markers were developed for population differentiation analysis. The divergence of markers under selection indicated two genetically different regions (North-East and South), which was very similar to those for adult lobsters. A low level of divergence for the neutral markers signaled effective dispersal of the larvae, and in combination with patterns of local adaptation, suggests a phenotype – environment mismatch, which results in post-settlement mortality of non-adapted genotypes. Additionally, the similarity between adaptive population differentiation of adult lobsters and late juvenile/early juveniles is indicative of local selection acting on earlier stages of *Jasus edwardsii* development than were sampled here.

Introduction

Many marine species have a bipartite life cycle with larvae undergoing dispersal and maintaining population connectivity across vast distances, while the main function of the adult life stage is survival to maturity and reproduction. Population adaptability and resilience rely heavily on recruitment being sufficient to support population persistence. Dispersal, settlement and eventually recruitment are affected by a number of factors summarised in **Chapter 1, Fig.1.1**. The success of dispersal and the effectiveness of recruitment depend on both pre- and post-settlement barriers to connectivity (Sanford et al. 2011). Pre-settlement barriers include physical (direction of currents, isoclines, eddies), as well as biological barriers (predation or starvation). Larval mortality increases with a longer pelagic larval duration and thus can affect the connectivity between remote populations even when physical barriers are lacking (Pineda et al. 2009). Larvae surviving the dispersal phase face new challenges in the new environment they find themselves: physiological costs of dispersal and possible phenotype – environment mismatches (Marshall et al. 2010).

Indirect costs of dispersal are especially pronounced in non-feeding larvae – depleted energy levels during dispersal may reduce the fitness of newly-settled migrants and reduce their chances of survival (Shima et al. 2010). However, larval experience and latent effects of dispersal can affect feeding larvae. For example, when food is concentrated in coastal waters larvae travelling with oceanic currents have less access to food sources and survival of local settlers is higher (Pechenik 2006). Another post-settlement connectivity barrier is potential phenotype – environment mismatches. Heterogenous environments promote genotypic and phenotypic adaptations of populations. Local selection produces genotypes (and phenotypes) "optimised" for the present conditions, however, it also reduces fitness if the "optimised" genotype is dispersed to a habitat with different conditions (DeWitt et al. 2004). Thus phenotype – environment mismatches encourage self-recruitment via increased chances of survival of local genotypes and decreased the fitness of settlers from an alternative environment (Nosil et al. 2005). These two mechanisms of reduced connectivity can be distinguished by patterns of divergence of genetic markers. Post-settlement mortality caused by the larval experience and deferred costs of dispersal is directly dependent on the dispersal distance and should translate into a clear Isolation-by-Distance (IBD) divergence pattern of neutral and outlier markers. Phenotype – environment mismatch, on the other hand, affects genetic markers under selection and should produce patterns of divergence in only outlier markers (Marshall et al. 2010). If the selective force is strong enough to overcome the

homogenising effect of the gene flow, divergence of neutral markers can occur under the influence of genetic drift. Thus, investigation of population differentiation patterns with both neutral markers and markers under selection can provide insight into the relationship between selective environmental pressure and recruitment (Nordeide et al. 2011).

Understanding larval connectivity and subsequent recruitment to a local population is essential for species conservation and the sustainable management of harvested resources. Local selective pressure can influence settlement and recruitment in the short-term and affect the genetic makeup and divergence of the populations, and should be taken into account in the development of conservation and management units (Conover et al. 2006; Fogarty et al. 2007). The identification of adaptively diverged populations and the translocation of adapted individuals into new populations have been shown to support genetic and demographic rescue in Trinidadian guppies (Fitzpatrick et al. 2016). Population genetics of juvenile life stages is often studied in order to elucidate the mechanisms of establishment of discrete adult stocks (Palsbøll et al. 2007). Investigation of differentiation of Southern Ocean krill furcilia larval stages, juveniles, and adults with cytochrome b SNPs allowed the identification of multiple sources of recruitment into the genetically homogenous stock and revealed cohort effects producing genetic divergence between the recruitment periods (Batta-Lona et al. 2011).

The importance of identifying and preserving locally adapted populations in order to ensure the genetic diversity of stock has been recognised for commercially harvested species (Hilborn et al. 2003). For example, a study of genetic divergence of juvenile sockeye salmon *Oncorhynchus nerka* using 45 SNPs, which included two loci under positive selection, highlighted the contribution of local adaptation to the spawning rather than nursery habitat in the formation of adult salmon population differentiation (McGlauflin et al. 2011).

Often a mix of adult and juvenile stages is assessed in order to identify processes common for both life stages. A number of studies on Atlantic cod *Gadus mohrhus L.* have identified population differentiation of adults and juveniles at pantophysin and haemoglobin genes, which is proposed to result from the selective pressure of temperature (Pogson et al. 2003; Westgaard et al. 2007; Fevolden et al. 2012; Andersen et al. 2015). Investigation of patterns of local adaptation has gained particular importance for fisheries management due to climate change-induced range shifts in the distribution of marine species (Last et al. 2011). Distribution shifts can prevent potential extinction of the species, but it requires an effective mechanism of dispersal to habitats for expansion (e Costa et al. 2014). Thus, studying

dispersal connectivity and adaptation capacity, along with mechanisms of local adaptation of fished species are critical for long-term sustainable management.

An overview of the *Jasus edwardsii* life cycle is given in **Chapter 1, Fig. 1.1**. Mating usually occurs in winter, after female lobsters have finished moulting. Fertilization is external, with male *Jasus edwardsii* lobsters depositing a spermatophore on the sternum of the female (MacDiarmid 1989a). After brooding the eggs for 3 to 6 months, females travel offshore to release the hatched larvae - usually in spring and summer (**Chapter 1**, **Fig. 1.2**). The first larval stage is called the naupliosoma and is very short-lived – 30 to 60 minutes, the larvae then moults into a first phyllosoma instar. *Jasus edwardsii* phyllosoma development can last for 12 to 24 months and includes 17 instar stages (Booth et al. 1994; Kittaka et al. 2005). Early stage phyllosoma can be found over several hundred kilometers offshore, while stages 5 to 11 can be found as far as 1300 km from the shore (Booth et al. 1994; Jeffs et al. 2001). Late-stage phyllosoma begin to travel back to shore in order to undergo metamorphosis into pueruli near the continental shelf (~200 km offshore) (Jeffs et al. 2001). Pueruli continue to travel closer to the shore and are mostly found within 100 km of the coast (Booth et al. 1994). Drifting with currents was suggested to be the main driver of the early- and mid-stage phyllosoma distribution, however it does not explain the distribution of the late-stage phyllosomas and pueruli (Chiswell et al. 2005; Chiswell et al. 2008). Pueruli have been shown to be active swimmers (Jeffs et al. 2001), so once they reach the continental shelf their movements are deliberate, and dispersal is affected by numerous factors: underwater sound, water chemistry, eddychemotaxis, salinity, magnetic fields, and specific behavior (Cobb 1997; Jeffs et al. 2005).

Pueruli morphology is very similar to the morphology of the juveniles, however pueruli are almost completely transparent. Pueruli prefer to settle in the small crevices in the shallow rocky shore (Booth et al. 1986). Settlement occurs throughout the year but the main settlement seasons are in the Austral summer (January - February) and winter (June - July) (Booth et al. 1994). The duration of the puerulus stage after the settlement is dependent on the time of settlement and seasonal temperature and can vary between one and three weeks (Booth et al. 1992). Pueruli metamorphose into the first benthic instar of the juvenile lobster with a CL of 5 – 7 mm and their colour intensifies with time. Three ecologically distinct phases of juvenile lobster development can be distinguished: algal, post-algal and subadult. Early benthic juveniles prefer to shelter in algal clumps but once they grow larger they switch

to rock crevices (Butler et al. 2000). Subadults become more gregarious and locate shelters occupied by conspecifics using chemotaxis (Lipcius et al. 2000).

Allozyme variation has been studied in red rock lobster across the Tasman sea, and Smith et al. (1980) were unable to determine any genetic variation between subpopulations based on the assessment of 33 enzyme gene loci. Booth et al. (1990) studied allele frequencies in lactate dehydrogenase (*Ldh*) locus and discovered no significant differences between animals from nine sites spread around New Zealand. However, the assumption of homogeneity of red rock lobster stock across its range has been challenged by several studies that revealed significant differentiation between Australian and New Zealand populations and provided some evidence for possible self-recruitment within New Zealand (Thomas 2012; Morgan et al. 2013; Villacorta-Rath et al. 2016). Investigation of the adult lobster population differentiation with neutral markers revealed unimpeded gene flow and high levels of connectivity between populations. However, analysis of the outlier markers indicated presence of a selective pressure counteracting the effects of gene flow and thus forming three genetically distinct regions with clines of admixture of diverged genotypes between them (**Chapter 3**). However, low levels of admixture of non-optimal genotypes in the locally adapted populations allow for homogenisation of the allele frequencies in the next generation, which explains the pattern of virtually no population differentiation at neutral loci. From settlement into adulthood, lobsters are subject to the local selection that reduces frequencies of non-adapted alleles. Red rock lobster phyllosoma are able to feed during their two-year-long offshore development phase and have been shown to accumulate enough resources for successful return to the shore and settlement of the non-feeding puerulus, thus minimizing reliance on the maternal effects and energetic costs of a New Zealand wide dispersal (Jeffs et al. 2004; Jeffs et al. 2005). However, any mismatch between phenotype (and underlying it genotype) and environment produced by the dispersal of larvae of locally adapted animals is likely to cause higher mortality of the settlers, preventing them from reproducing in the new habitat and effectively reducing population connectivity (Marshall et al. 2010). The genetic divergence patterns of markers under selection identified previously in adult rock lobsters indicate that selection has occurred locally, most likely during the settlement process or immediately after. The mismatch between the environment and migrant phenotypes (and underlying genotypes) that causes selective mortality of less adapted genotypes have been described as the main mechanism of establishment of local adaptation (Nosil et al. 2005; Hedrick 2006; Marshall et al. 2010). The first *Jasus edwardsii* stage to actively migrate to the

new environment in order to find appropriate settlement ground are the puerulus, which makes this life stage a likely candidate as a target of selection. This hypothesis can be tested by examining the patterns of divergence of neutral and outlier markers.

Thus, the aim of this study was to investigate the effects of gene flow and local selection on the early stages of *Jasus edwardsii* settlement.

Specific aims:

- To develop and validate panels of putatively neutral and under selection SNP markers for pueruli/juvenile *Jasus edwardsii;*
- To determine the mode of adaptive divergence between pueruli/juvenile populations via comparing patterns of neutral SNPs and SNPs putatively under selection;
- To test whether the selective pressure is exerted on pueruli/juvenile rather than sub-adult and adult life stages of *Jasus edwardsii* by comparing population differentiation patterns between these stages.

Methods

Sampling and DNA extraction

A total of 120 juvenile lobsters were collected from 4 locations (see **Fig.4.1**, **Table 4.1**) over the winter (June - August) settlement period using pueruli 'crevice' collectors (Booth et al. 1986) and were preserved whole in 96% ethanol. The pueruli stage of metamorphosis is relatively short and varies in length due to temperature, and in the 2016 settlement season most of the lobsters collected had already gone through the pueruli – juvenile metamorphosis. The occurrence of metamorphosis in spiny lobsters is judged based on loss of transparency, which obscures the hepatopancreas that is clearly visible in H-phase pueruli with an initial minuscule accumulation of colour (Ventura et al. 2015). Preservation in 96% ethanol likely affected the colour of transparent pueruli rendering them white and interfering with the visibility of the hepatopancreas, which made it virtually impossible to distinguish between the last pre-metamorphosis pueruli stage and the earliest post-metamorphosis juvenile stage (**Fig.4.2,** left panel), thus all the lobsters collected during 2016 settlement period will be referred to as juveniles. Unfortunately, settlement rates at the time of sampling weren't sufficient to obtain 30 samples in the same stage of development for all four locations, so a mixture of stages was used. Three of the broad age categories in the **Fig.4.2** were used in equal proportions where possible, and analysis of Principal Components was used to detect any batch effects between three age groups.

Table 4.1. Pueruli/juvenile lobster sampling locations.

Region	Sampling location	Latitude	Longitude	Date	N_{SEQ}	N_{FILT}
North Is (NE)	WHA	38°34'26.10"S	178°13'54.11"E	Jun - Aug 16	32	30
North Is (NE)	NAP	39°28'14.62"S	176°54'54.18"E	Jun - Aug 16	32	30
North Is (NE)	CAS	40°54'34.29"S	176°13'18.73"E	Jun - Aug 16	32	30
South Is (S)	JAC	43°57'35.86"S	168°37'29.08"E	Jun - Aug 16	32	30

Fig. 4.1. Map of sampling locations: WHA - Whangara, NAP – Napier, CAS – Castlepoint, JAC – Jackson Bay. Numbers in green rectangles represent the number of juvenile lobsters used. Adapted from NZ Seafood Industry Council report, 2005.
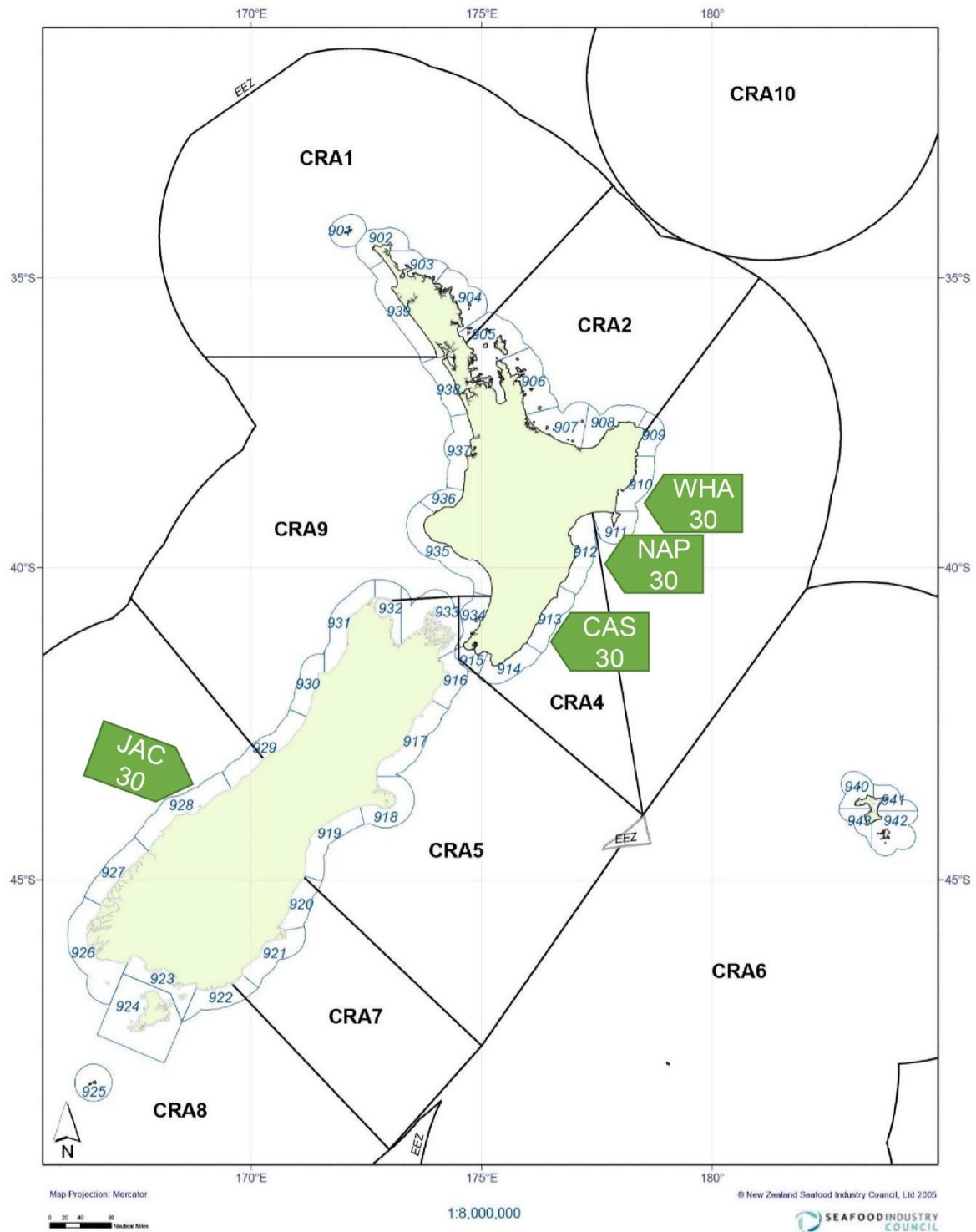
Fig.4.2. A range of pueruli – juvenile lobsters illustrating the accumulation of colour with time since pueruli – juvenile metamorphosis. The left panel shows pre-metamorphosis pueruli/early post-metamorphic juvenile; the central panel shows an intensification of exoskeleton colour post-metamorphosis and the right panel illustrates a fully formed juvenile with developed exoskeleton color.

Library preparation and sequencing

DNA from preserved samples (antennae) was extracted and processed following optimized ddRADseq protocol first developed by Peterson et al. (2012) and described in detail in **Chapter 3**.

After processing, two juvenile lobster libraries were sequenced on the Illumina HiSeq2500 platform with single end 100 bp HT chemistry kits at the Australian Genome Research Facility in Melbourne.

Bioinformatic analysis

Preliminary processing of the ddRADseq data

The quality of sequences obtained was assessed using the FastQC v.0.10.1 quality control tool (Andrews 2010). Reads were then trimmed to 95 bp using TRIMMOMATIC v 0.3 (Bolger et al. 2014). Trimmed reads were filtered for possible contamination using Kraken v.3.5.0 (Wood et al. 2014) and were further demultiplexed using Stacks v.1.37 module *"process_radtags"* (Catchen et al. 2011).

Sequencing data from adult lobster libraries and pueruli/juvenile libraries were pooled and a reference catalogue built using the Rad-loci pipeline. The Rad-loci pipeline developed at La Trobe University (https://github.com/molecularbiodiversity/rad-loci) (Villacorta-Rath et al. 2016) was also used for reference catalog compilation. This pipeline involves the clustering of sequences, which was performed using VSearch v.1.1.3 (Rognes et al. 2016), with the minimal occurrence of a unique sequence set equal to the number of individuals sequenced, which promotes the selection of loci with lower levels of missing data. To merge similar reads that may have occurred from sequencing errors, the clustering identity was set to 94%, meaning a maximum 7 bp mismatch for 95 bp fragments. The clusters obtained were filtered based on the number of members in the cluster: clusters with less than 2 members were removed as they were considered uninformative, clusters with more than 16 members were also removed to avoid multi-mapping. In the next step, sequences were aligned onto the prepared catalog to estimate the number of alleles for each cluster and the catalog was re-filtered to retain clusters with 2 to 16 alleles. Individual sequences were mapped onto the re-filtered catalog once again and loci with more than 40% missing data per locus were removed. Here "locus" is used to describe a unique reference sequence 95 bp long between the EcoRI and AciI cut sites with a cluster of mappable sequence reads.

SNP calling and characterization

The in-house and Rad-loci pipeline only create a catalog of reference loci and therefore mapping of the individual reads onto the catalog and SNP calling relies on a third-party software. For both of these pipelines demultiplexed, quality and contamination filtered individual reads were mapped onto the reference catalog using the Burrows-Wheeler Aligner v 0.7.13 (Li et al. 2010). Reference catalog loci with a minimum coverage of 10x were used

in an effort to minimize the effect of allele dropout (ADO). SNP calling was performed by two independent programs: Genome Analysis Toolkit (GATK) v 3.5.0 (McKenna et al. 2010) and VarScan v 2.4.1 (Koboldt et al. 2009). Only SNPs consistently called by both programs were selected using SelectVariants module of GATK software. SNPs in concordance identified by GATK and VarScan were further filtered using VCFtools v 0.1.14 (Danecek et al. 2011).

For all three pipelines non-biallelic loci were removed with the *--min-alleles* 2 *--max-alleles* 2 options and only one SNP per locus was retained using the *--thin* 95 option. Filtering of rare alleles was performed using the minor allele frequency filtering option *--maf* set to 0.1. The Linkage Disequilibrium (LD) filter was used to minimize the occurrence of linked loci (option *--min-r2*) with $r^2 > 0.8$ threshold in addition to retaining only one SNP per locus. Individuals and loci with more than 20% missing data were also removed from any further analysis. In addition to commonly used metrics for pipeline comparison (number of loci, number of SNPs), a filtered set of SNPs was further divided into two panels according to calculated F_{ST} value compared to neutral distribution (Excoffier et al. 2010) – neural loci and loci putatively under selection (further referred to as outlier loci). One of the most commonly used methods is based on global F_{ST}, which is employed to identify genetic markers with higher genetic distance than expected in a neutral distribution. A number of demographical scenarios have been shown to influence these estimations and result in type I (false positive) and type II (false negative) errors (Excoffier et al. 2009).

Dependent on the level of selection, mutation rate and hierarchical structure of the populations, different implementations of simulation methods may perform better than others (Narum et al. 2011) and in order to obtain the most conservative estimation, markers under selection were identified with several F_{ST} outlier tests. BAYESCAN v2.1 (Perez-Figueroa et al. 2010) was used with 10 000 iterations and a burn-in of 200 000 steps. A Prior model of 10 000 was used in order minimize the number of false positives (Lotterhos et al. 2015). Arlequin v3.5.2.2 (Excoffier et al. 2010) runs were implemented using 10 000 simulations and 1000 demes. 500 000 simulations and a false discovery rate of 0.1 were used when running Lositan (Antao et al. 2008). A Q-value of 0.05 was used in BAYESCAN runs and the p-value was set at 0.05 in Arlequin and Lositan runs. Only SNPs consistently identified by all three programs were used for further analysis with all three pipelines. Furthermore, the outlier panel was divided into loci under balancing and divergent selection, only loci under divergent selection were retained for analysis.

An analysis of technical replicates was performed to control for possible batch effects of the library preparation and sequencing, and to validate the use of the SNPs discovered. Spatial distribution of replicates was assessed using a Principal Component Analysis (PCA) performed in the R environment with the package adegenet v.1.4-1 (Jombart et al. 2011). In addition, hierarchical clustering was performed using the *hclust* function from package stats-package v3.4.1 (Team 2017).

Population genetics analysis

All population genetics analyses were performed using neutral, neutral+outlier and outlier marker panels. The same tests used for the population genetics analysis of adult lobsters were used for the juvenile and adult + juvenile datasets. F-statistics, observed (H_o) and expected heterozygosity (H_E) were calculated using Arlequin v3.5.2.2 (Excoffier et al. 2010) and the R package diveRsity (Keenan et al. 2013). A two-sample Wilcoxon test was used to detect whether mean H_o differed significantly from mean H_E for both SNP panels. Population structure was analysed using a Bayesian model-based clustering algorithm implemented in STRUCTURE (Pritchard et al. 2003). In addition, a non model-based multivariate analysis, Discriminant Analysis of Principal Components (DAPC), was implemented in the R package adegent (Jombart et al. 2011).

Results

Genotyping results

A total of 130 juveniles were sequenced and no samples were removed from the dataset for missing more than 20% of data. A new reference catalogue was generated for an adult + juvenile dataset and levels of missing data were assessed for a new set of SNPs, which consistently occurred in juvenile and adult lobsters. Only 15 individuals were excluded from the combined dataset based on the missing data threshold. Individuals were then sorted into two groups for replicate and population genetics analysis (**Table 4.2**). Only one unique individual per two or more replicates was included in the dataset for population genetic analysis.

Table 4.2. Final counts of individuals in the datasets analysed.

	Juvenile dataset	Adult + Juvenile dataset
Total sequenced	130	864
Excluded	0	15
Replicates	10	85
Population analysis	120	764
Replicate analysis	20	152

The average number of sequence reads per individual across the two libraries was close to three million reads with approximately 5 % being removed as reads resulting from contamination and approximately 3% were removed due to low quality or ambiguous barcodes/RAD-tags. After filtering, the reference catalogue retained more than 9,000 and 11,000 loci for the juvenile and adult + juvenile datasets, respectively. From the reads mapped onto the juvenile lobster catalogue, I identified a total of 3970 filtered SNPs, while 2307 SNPs were discovered for the adult + juvenile dataset. Comprehensive statistics of all data processing and filtering stages are presented in **Table 4.3**.

SNPs were then ranked according to single loci F_{ST} and divided into markers that were putatively neutral (73.4%), and putatively under selection (referred to as outliers), which were further divided into markers under balancing selection (8.4%) and under diversifying selection (18.2%). Interestingly, the juvenile dataset markers had a higher percentage of neutral markers (87.4%) and a lower proportion of diversifying SNPs (8.4%) (**Fig.4.3**) when compared to the previously analysed adult dataset (**Chapter 3**) and the adult + juvenile dataset analysed in this chapter (**Fig.4.4**).

Table 4.3. Stages of sequencing data filtering.

Stage	Statistic	Juvenile	Adult + Juvenile
Sequence filtering	Average number of reads per individual	2831067	2900144
	Average percentage of contamination-filtered reads	5.4	4.6
	Average percentage of low quality and ambiguous barcode reads	3.1	2.8
Reference catalogue filtering	Total loci	19794	22899
	Filtered loci	11254	9604
SNP filtering	Called with VarScan	86615	166162
	VarScan filtered	5228	3282
	Called with GATK	132765	223537
	GATK filtered	9739	7537
	VarScan/GATK concordance	3970	2307
	Neutral	2141	1639
	Under balancing selection	1815	541
	Under divergent selection	14	127

Fig.4.3. Number of SNPs identified as neutral (panel A) and outliers (panel B) for the juvenile dataset using three genome scan methods.
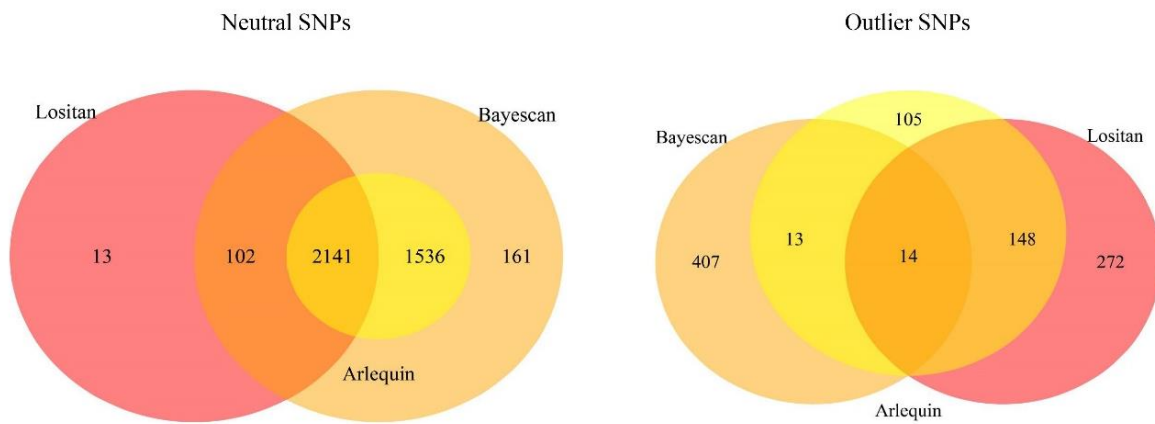
Fig.4.4. Number of SNPs identified as neutral (panel A) and outliers (panel B) for the juvenile and adult datasets combined using three genome scan methods.
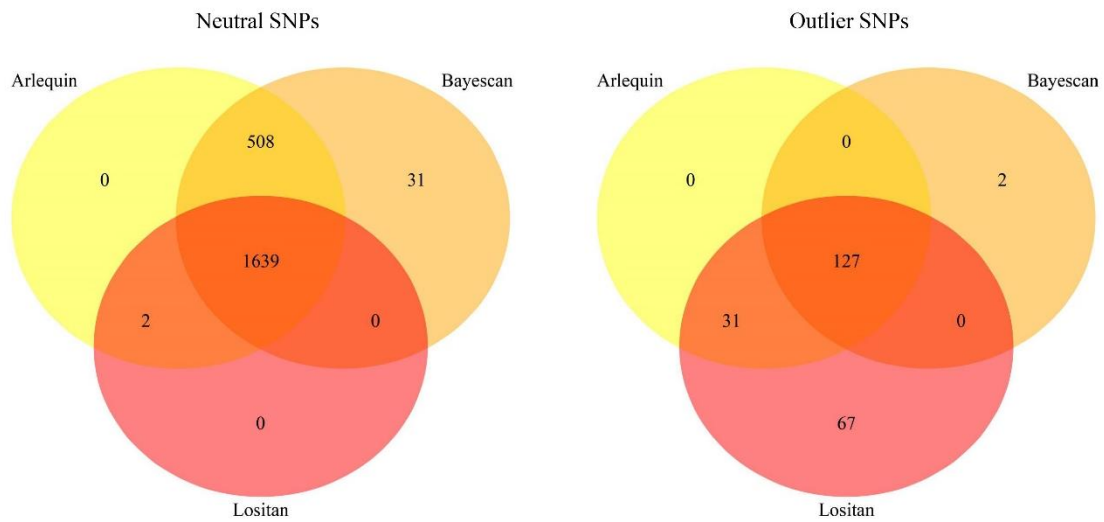
Replicate analysis

Principal Component Analysis of the technical replicates showed that technical replicates of the same individual were consistently placed within the shortest distance of each other, regardless of it being intra- or inter- library replicate for both the juvenile (**Fig.4.5,** panel A) and adult + juvenile datasets (**Fig.4.6,** panel A). Hierarchical clustering with the Unweighted Pair Group Method with Arithmetic Means (UPGMA) resulted in technical replicates being clustered together within juvenile (**Fig.4.5,** panel B) and juvenile + adult datasets (**Fig.4.6,** panel B). PCA and UPGMA clustering analysis indicated reproducibility of wet-lab protocols and consistency of sequencing and SNP identification.

Fig.4.5. PCA of the juvenile dataset replicates shown in Panel A, UPGMA hierarchical clustering is shown in Panel B.
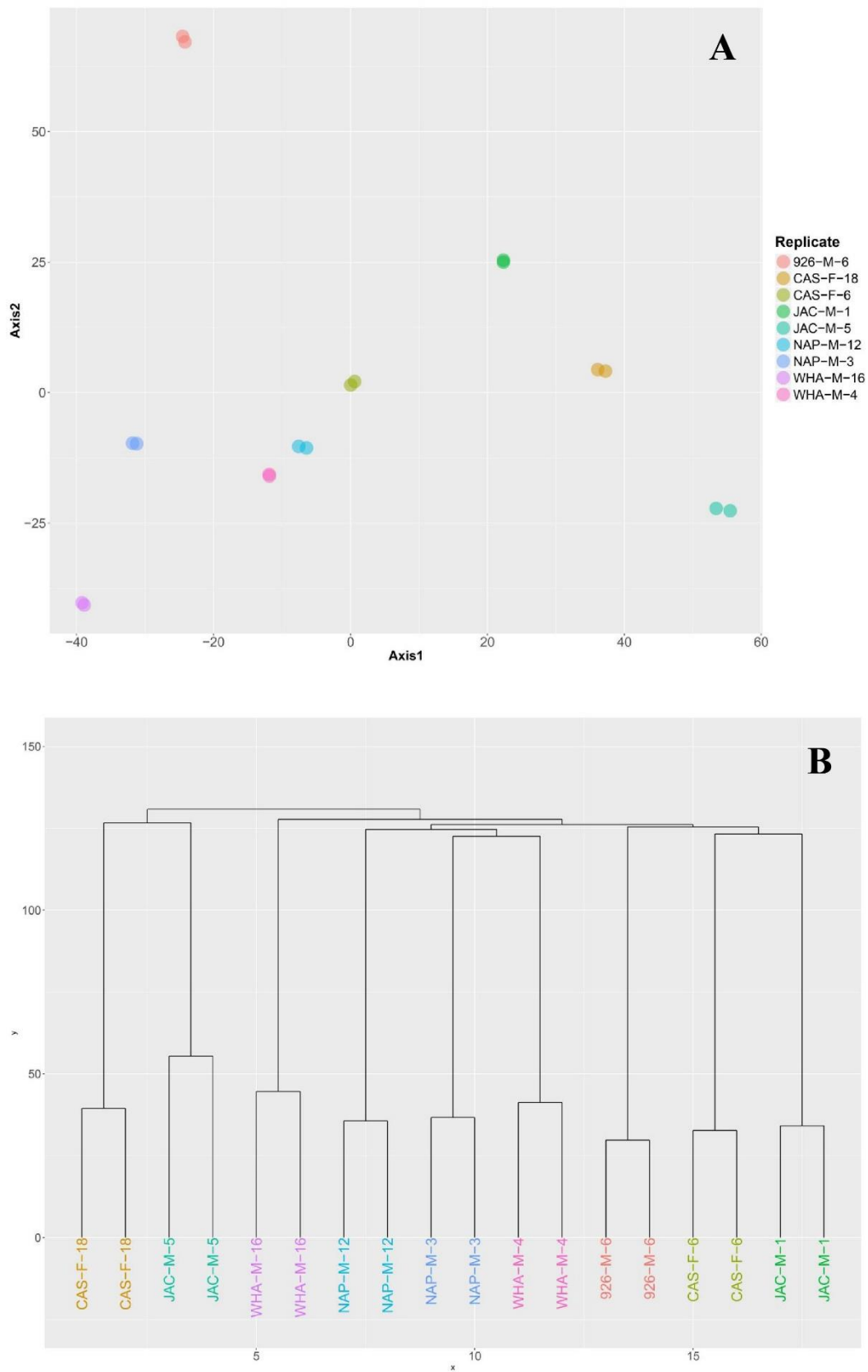
Fig.4.6. PCA of the adult + juvenile dataset replicates is shown in Panel A, UPGMA hierarchical clustering is shown in Panel B.
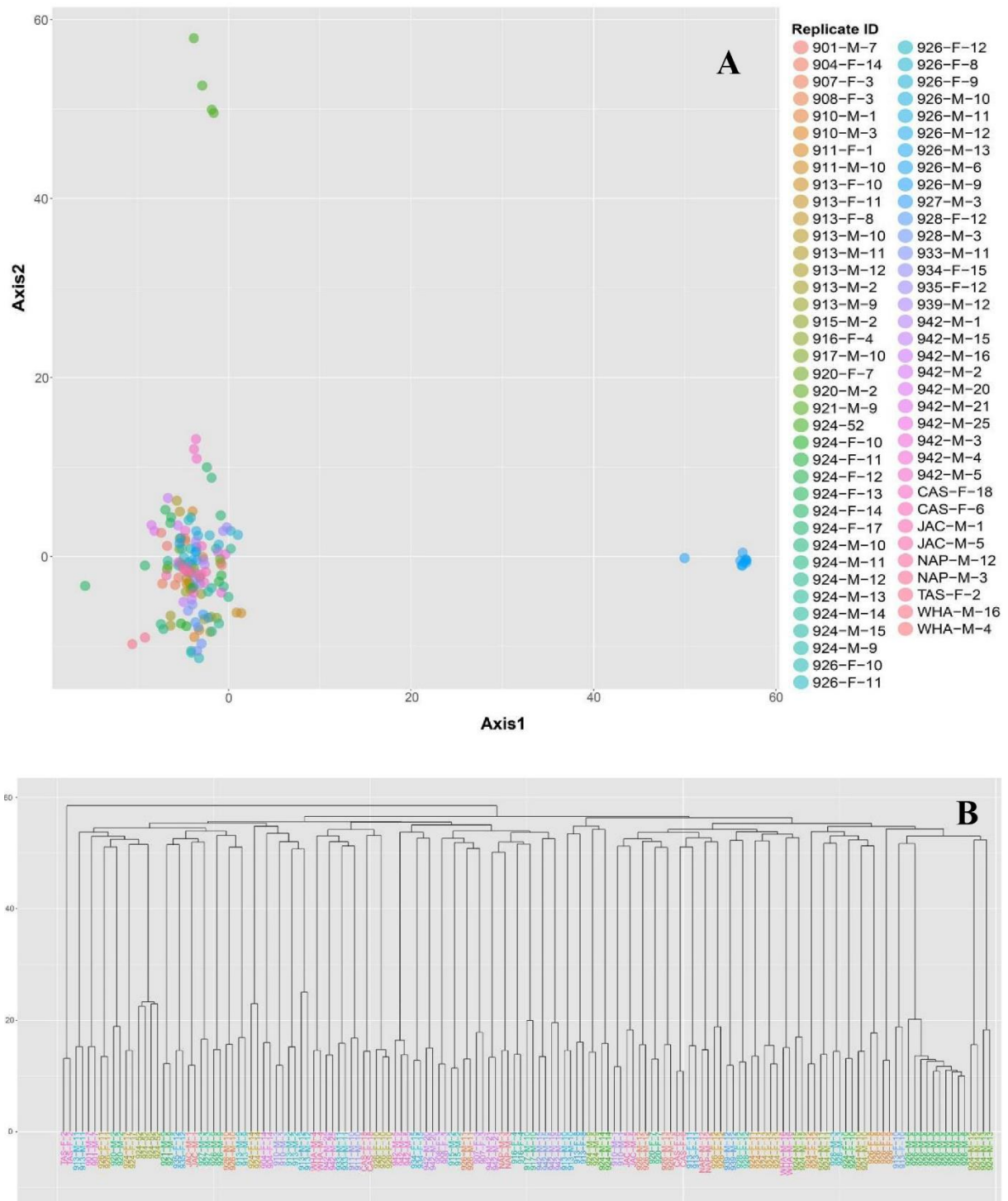
Population genetics analysis

F-statistics, Bayesian clustering analysis and DAPC were performed to assess patterns of population differentiation between the three broad genetically divergent regions described in **Chapter 3**:

- NW region (stat areas 935, 939, 901 and 902);
- NE region (stat areas 904, 907, 908, 910, 911, 913, 915, 942, 916, 917, 933, 934 and juvenile sampling sites WHA, NAP, CAS);
- S region (stat areas 920, 921, 924, 926, 927, 928 and juvenile sampling site JAC).

As shown in **Fig.4.2.** my samples likely included a range of ages from late pueruli/early juvenile to fully formed juveniles. PCA analysis did not reveal any batch effects between late pueruli and juveniles of the three broad age brackets using either neutral and outlier markers (**Fig.4.7.**) thus allowing to analyse all 120 samples as a single dataset.

Fig.4.7. Results of comparison of juvenile age brackets using Principal Coordinates Analysis. Top three panels represent spatial distribution inferred with neutral markers, bottom three panels – with outlier markers.
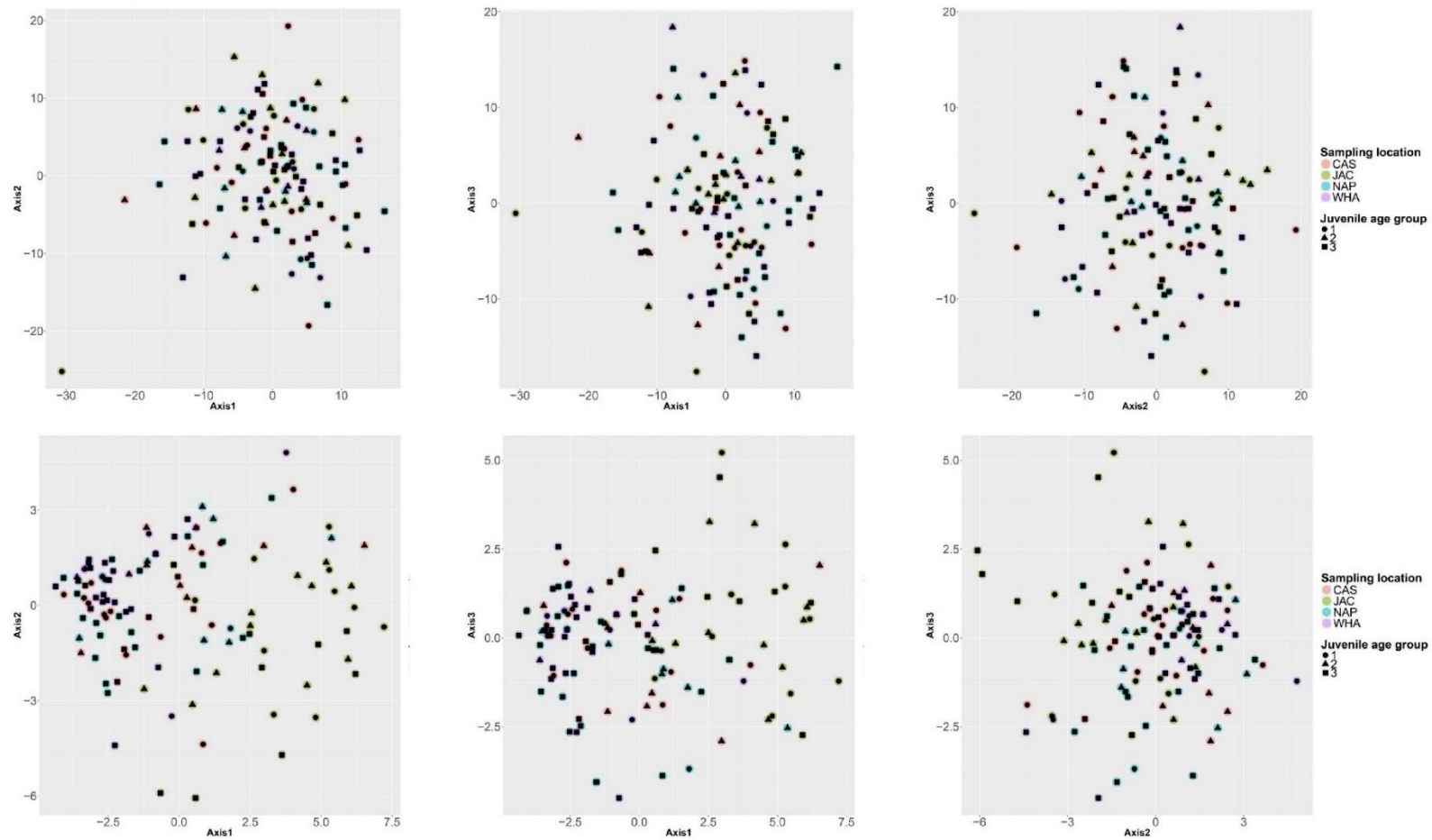
F-statistics

Pairwise F_{ST} values calculated for outlier, neutral and combined set of markers were highly significant at p<0.0001 (**Table 4.4**). However, the F_{ST} value for the neutral panel was very low and unlikely to indicate any ecologically meaningful population subdivision. In contrast, outlier panel F-statistics indicated the division between the NE and S regions. Neutral+outlier set of markers demonstrated intermediate level of differentiation when compared to neutral and outlier markers alone. For the combined dataset, highly significant (p<0.0001) differentiation was detected with outlier SNPs for the NE region, while the NW and S regions were noticeably less divergent from each other, although this difference was still significant. Neutral markers exhibited significant but very low levels of differentiation while neutral+outlier markers showed higher level of differentiation although not as high as outlier markers.

Table 4.4. Pairwise F_{ST} values calculated after pooling of the sampling locations into subpopulations. Results are presented for neutral, outlier and neutral+outlier panels separately. Values in bold are statistically significant at p<0.0001.

		Juvenile			Adult + Juvenile		
Neutral panel		NE	S		NW	NE	S
	NW			NW	0		
	NE	0		NE	**0.0002**	0	
	S	**0.0006**	0	S	**0.0008**	**0.0005**	0
Outlier panel		NE	S		NW	NE	S
	NW			NW	0		
	NE	0		NE	**0.0373**	0	
	S	**0.049**	0	S	**0.0162**	**0.0403**	0
Neutral + Outlier panel		NE	S		NW	NE	S
	NW			NW	0		
	NE	0		NE	**0.004**	0	
	S	**0.002**	0	S	**0.002**	**0.004**	0

Levels of heterozygosity obtained were very similar for both datasets, however inbreeding coefficients F_{IS} calculated for the juvenile dataset with outlier and neutral+outlier markers

showed positive values, which is very different from the adult and adult + juvenile datasets that showed negative values for all regions (**Table 4.5**).

Table 4.5. Descriptive statistics calculated after pooling of the sampling locations into subpopulations. Results are presented for neutral, outlier and neutral+outlier panels separately.

Dataset		Neutral panel			Outlier panel			Neutral + Outlier panel		
		H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}
Juvenile	NW	0.442 (±0.230)	0.385 (±0.102)	-0.15	0.33 (±0.121)	0.356 (±0.109)	0.07	0.322 (±0.113)	0.327 (±0.089)	0.015
	S	0.44 (±0.239)	0.380 (±0.109)	-0.16	0.349 (±0.145)	0.361 (±0.116)	0.035	0.321 (±0.131)	0.321 (±0.099)	0.00
Adult + Juvenile	NW	0.418 (±0.201)	0.371 (±0.104)	- 0.126	0.394 (±0.130)	0.378 (0.100)	- 0.042	0.416 (±0.196)	0.371 (0.104)	- 0.118
	NE	0.416 (±0.201)	0.372 (±0.103)	- 0.119	0.357 (±0.148)	0.348 (±0.121)	- 0.028	0.411 (±0.198)	0.370 (±0.105)	- 0.111
	S	0.417 (±0.203)	0.317 (±0.105)	- 0.125	0.359 (±0.139)	0.353 (±0.113)	- 0.019	0.412 (±0.199)	0.369 (±0.106)	- 0.116

Discriminant Analysis of Principal Components

Barplot of probabilities of assignment of individuals to the clusters identified suing outlier markers showed the northern-most juvenile sampling location (WHA) had the highest admixture of genotypes, depicted in green, and their proportion declined at sites further South (NAP and CAS) (**Fig.4.8**). The proportion of genotypes in red increased along a North – South gradient with the southern-most location (JAC) having the highest admixture of "red" genotypes among locations sampled.

Fig.4.8. DAPC results obtained with outlier markers viewed via compoplot graphic function of the *adegenet* package showing the probabilities of assignment of individuals to the clusters identified.
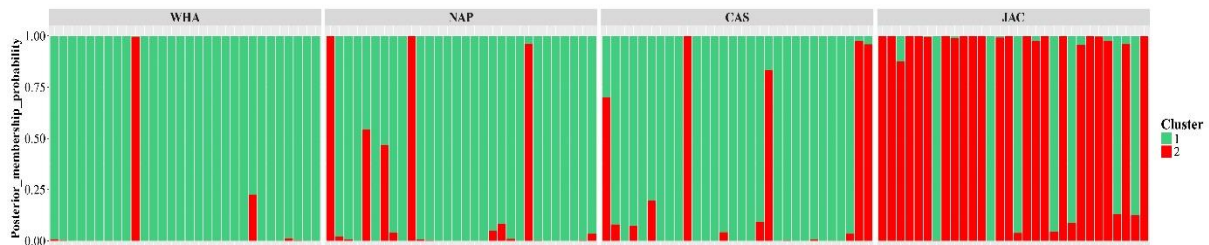
DAPC analysis of the juvenile only dataset performed with the outlier markers identified two clusters almost exclusively comprised of individuals from the NE region (Cluster 1) and S region (Cluster 2), except for admixture of a single individual in each cluster which accounted for 1% admixture of S region in Cluster 1 and 3% admixture of NE region in Cluster 2 (**Fig.4.9**, panel C). The neutral panel similarly identified two clusters within the dataset, however both clusters contained almost equal proportions of NE and S regions (**Fig.4.9**, panel A). Combined set of markers showed slightly improved discriminating ability between the two regions when compared to the neural markers alone (**Fig.4.9**, panel B). DAPC using outlier markers identified three clusters as the most favorable number of genetically differentiated groups in the adult + juvenile dataset. The clustering pattern was very similar to one observed for adults except for higher levels of admixture of individuals from other clusters. The most distinct was Cluster 2, 92% of which was comprised of lobsters originating from the NE region. DAPC detected a much lower level of differentiation between the NW and S genotypes, compared to the adult dataset clustering pattern (**Fig.4.10**, panel C). Pattern of differentiation inferred with neutral markers was very similar to the results obtained for adult lobsters only (**Chapter 3**) – two clusters identified, however both include individuals from NW, NE and S regions in almost equal proportions (**Fig.4.10**, panel A). Similarly to the results of previous analysis and results obtained for adult dataset, combined set of markers identified two clusters: mostly NE region and NW+S region (**Fig.4.10**, panel B).

Fig.4.9. Discriminant Analysis of Principal Components (DAPC) of genetic differentiation among two regions in the juvenile dataset. Panel A represents clustering produced by the neutral SNP panel, panel B – produced by the outlier SNP panel.
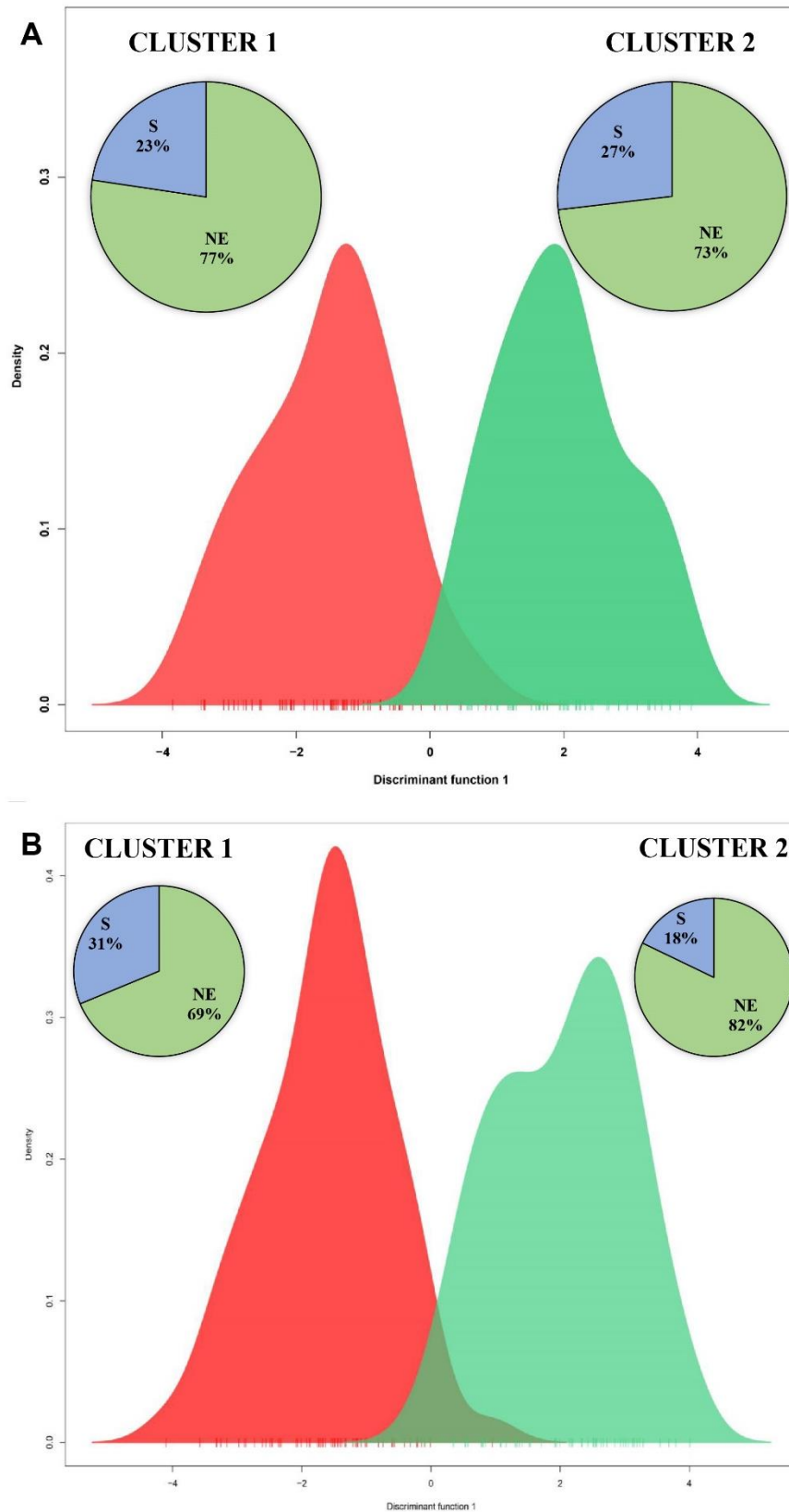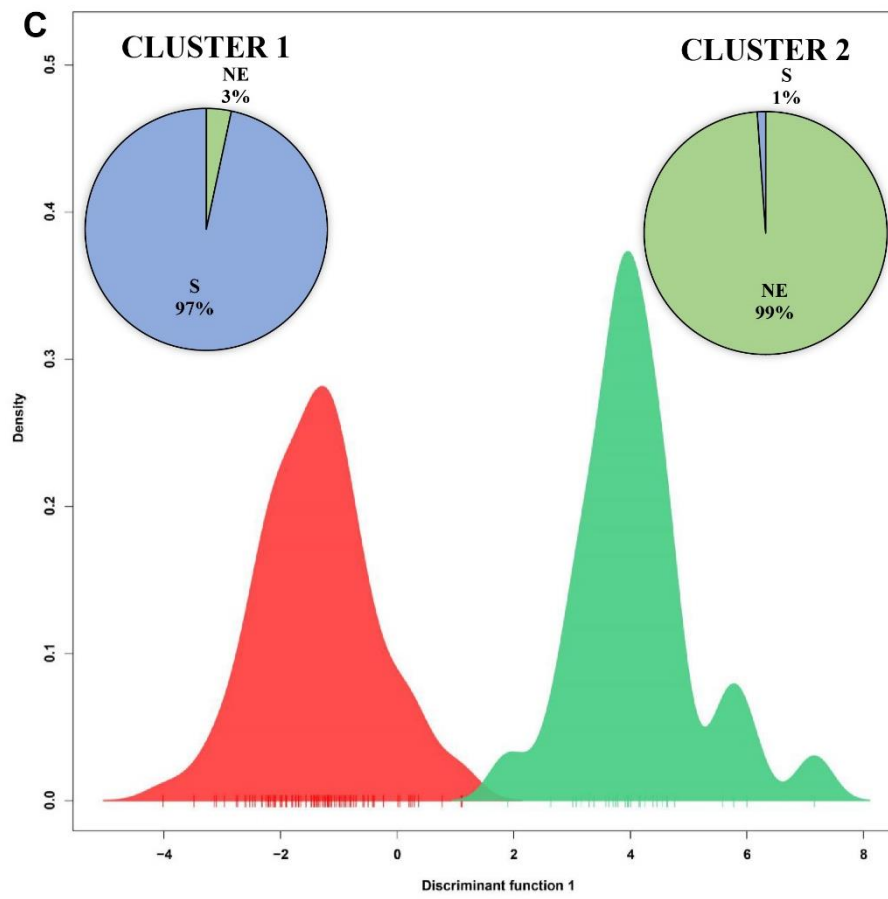
Fig.4.10. Discriminant Analysis of Principal Components (DAPC) of genetic differentiation among two regions in the adult and juvenile datasets combined. Panel A represents clustering produced by the neutral SNP panel, panel B – produced by the outlier SNP panel.
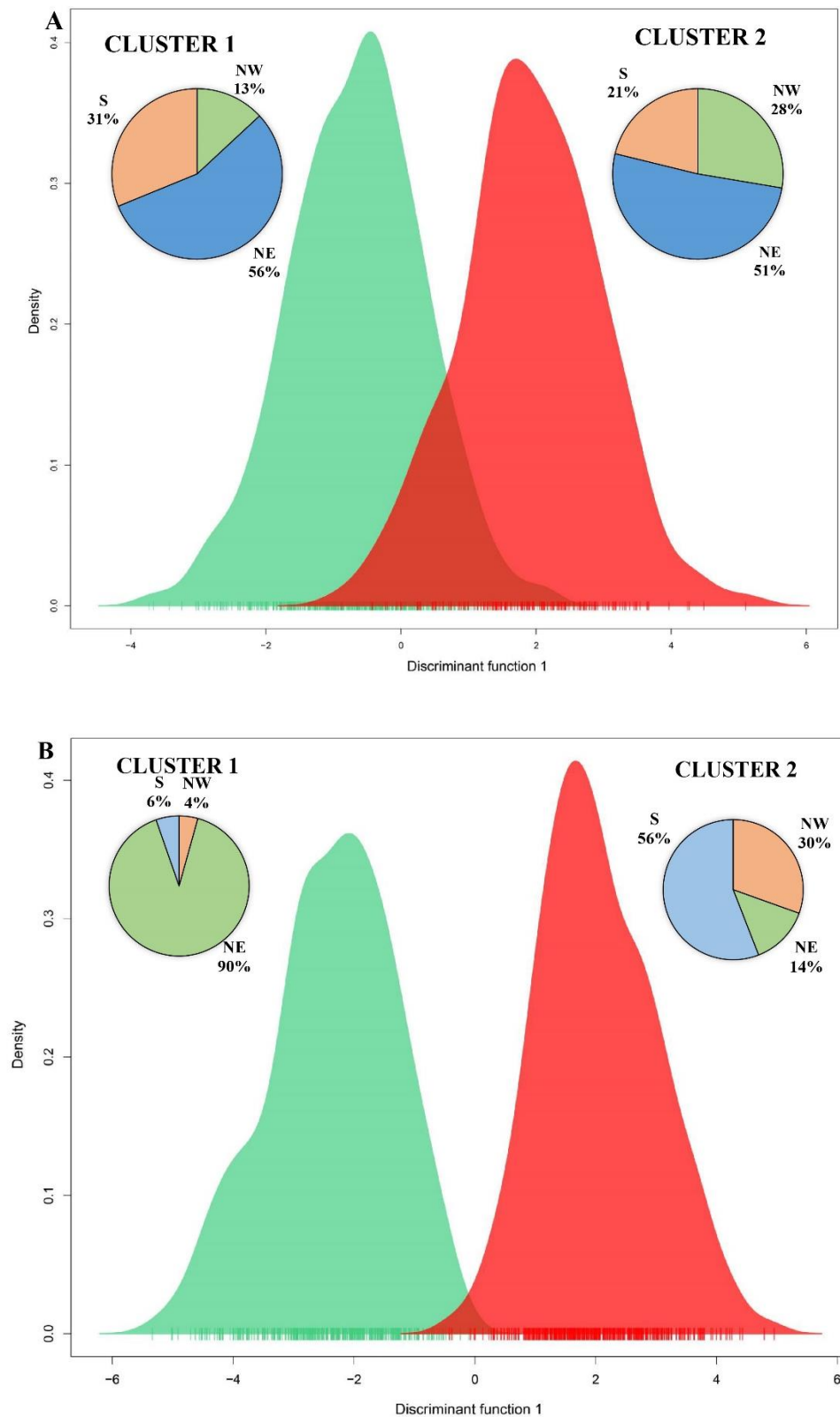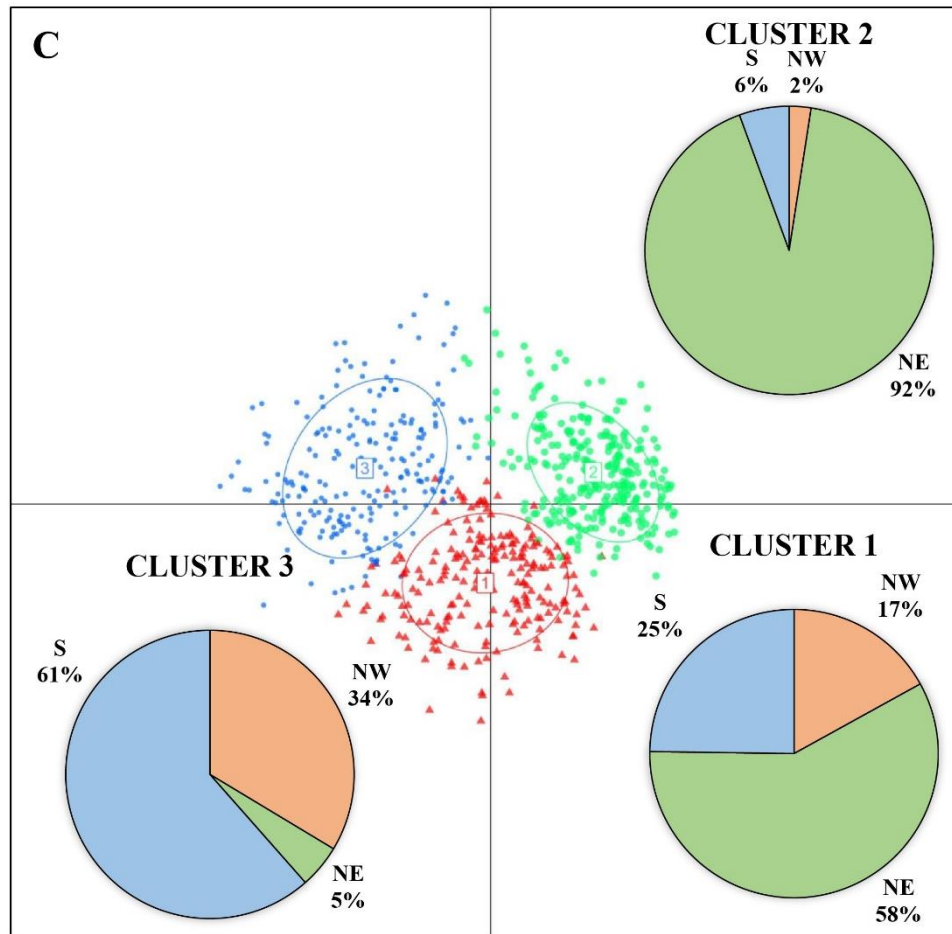
Bayesian clustering analysis with STRUCTURE

Bayesian clustering identified K = 2 as the most favorable number of clusters for all three sets of markers in the juvenile dataset. The markers under selection exhibited strong differentiation between the NW and S regions with most individuals in the NW region being characterised by the high levels of genotypes at multiple loci depicted in green, while the S region had the highest admixture of individuals with genotypes shown in red. A decrease in admixture of green/increase in admixture of red in the locations on the East coast of North Is in a North-South direction wasn't consistent enough to identify any further subpopulation differentiation (**Fig.4.11,** panel C), but was suggestive of a cline of genetic differentiation, which was consistent with the DAPC analysis. When analysed with neutral and neutral+outlier markers, lobsters from all sampling locations appear to have varying levels of admixture of "red" or "green" genotypes at multiple loci, but it is not consistent with the geographical position of the populations analysed (**Fig.4.11,** panel A, B).

For the markers under selection in the adult + juvenile data K = 3 was most supported, however, only individuals from the NE region had a characteristic genetic makeup, while the NW and S regions appear to be comprised of individuals with two of three genotypes at multiple loci in approximately equal proportions, showing little differentiation between these two regions (**Fig.4.12,** panel C). Combined set of markers produced a very similar differentiation pattern to purely outlier markers, although K=2 was identified as most favorable (**Fig.4.12,** panel B). Neutral markers, consistent with previous analyses, showed no noticeable population subdivision (**Fig.4.12,** panel A).

Fig.4.11. Bayesian clustering analysis (STRUCTURE v.2.3.4) of genetic differentiation between two regions sampled for the juvenile dataset produced by the neutral (panel A), neutral+outlier (panel B) and outlier (panel C) SNP panel.
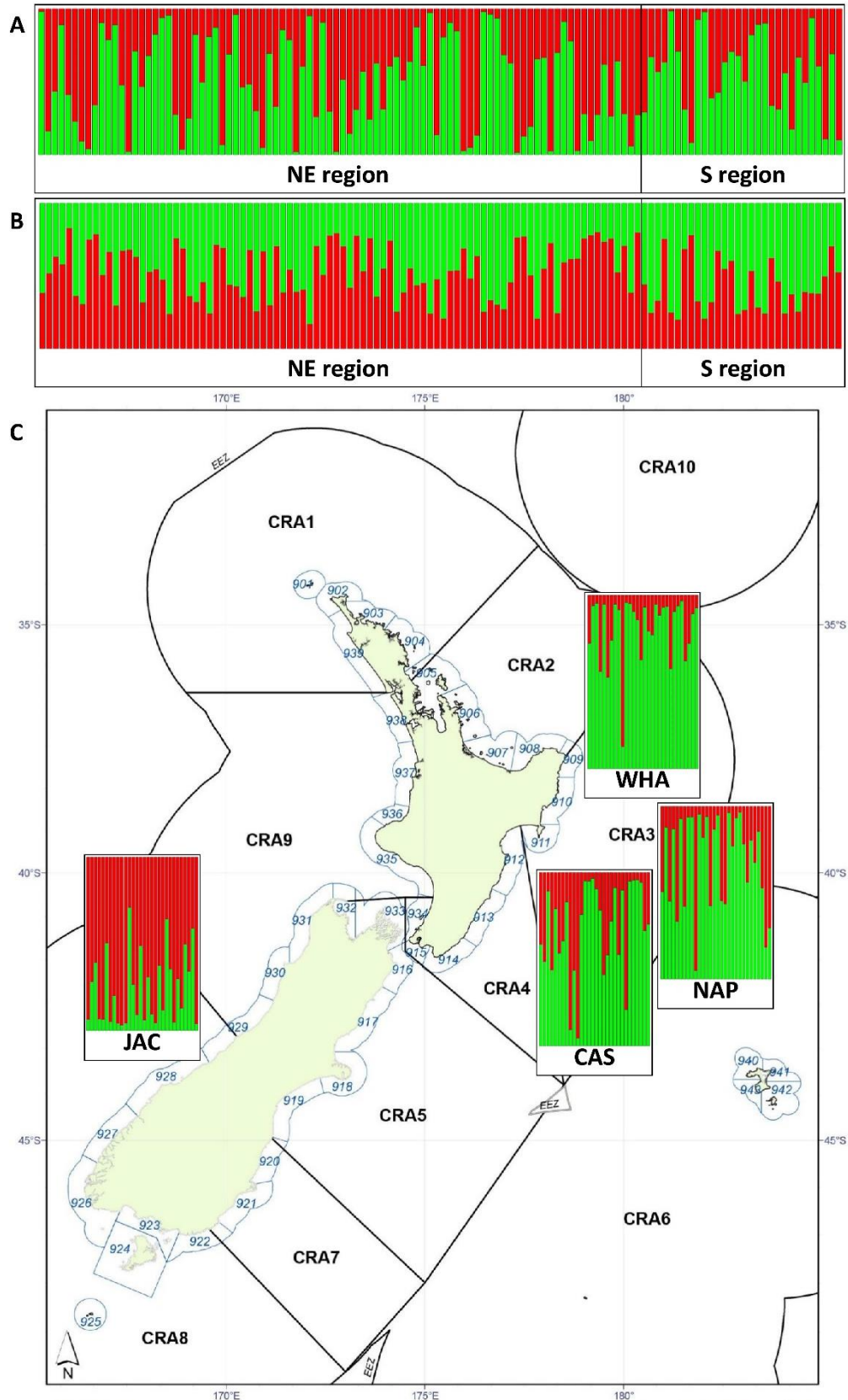
Fig.4.12. Bayesian clustering analysis (STRUCTURE v.2.3.4) of genetic differentiation between two regions sampled for adult + juveniles produced by the neutral (panel A), neutral+outlier (panel B) and outlier (panel C) SNP panel.
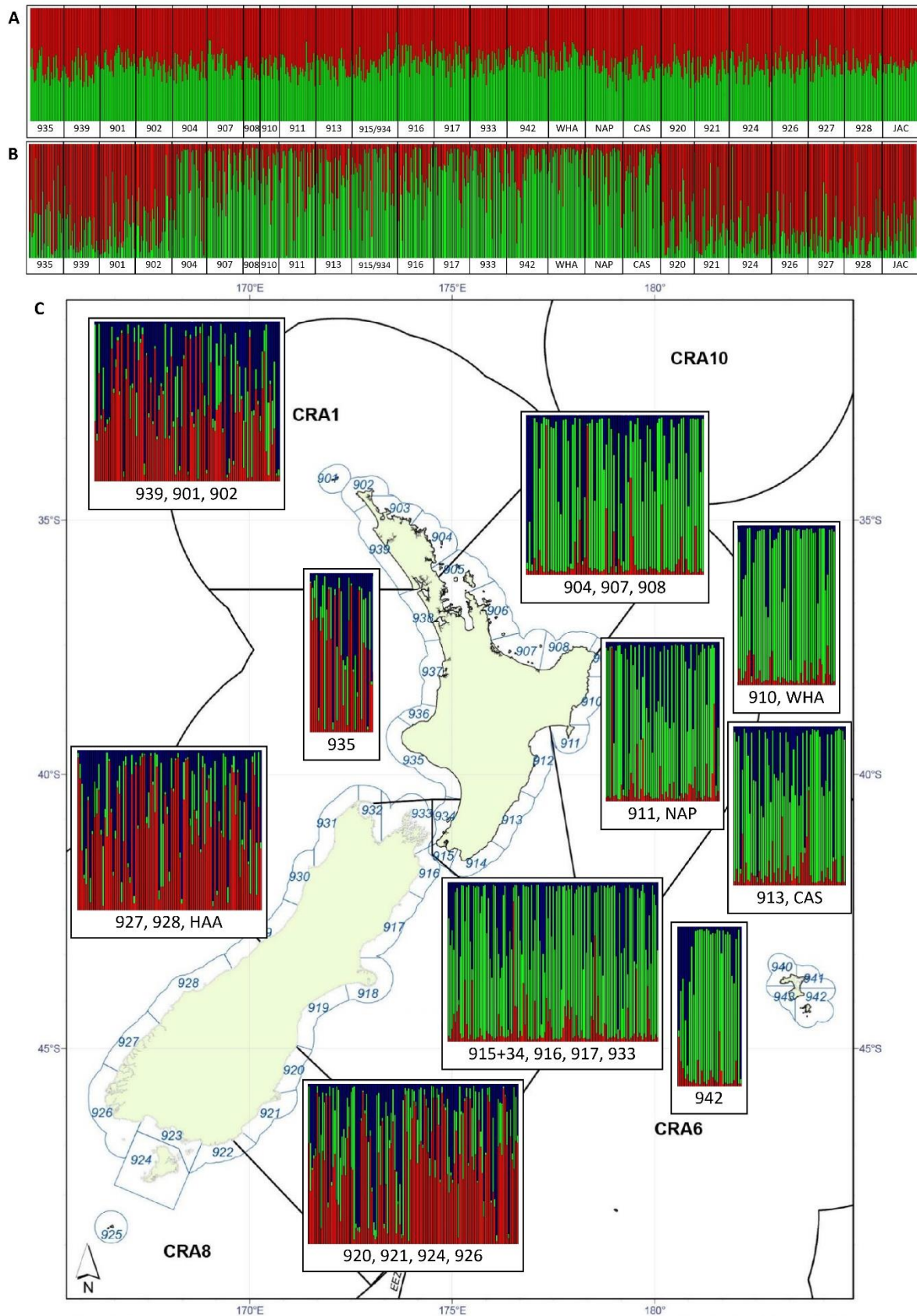
Discussion

The analysis of the genetic differentiation between juveniles originating from different locations can provide invaluable information about the mechanisms promoting divergence and may offer insight into the selective pressures of the environment (Waples et al. 2008; Berumen et al. 2012). Understanding local adaptation and possible mechanisms that promote adaptive divergence are increasingly recognised as critical for species conservation and management in the face of climate change and its effect on the distribution range of marine animals (Last et al. 2011). Thus, the main aim of this chapter was to investigate the levels of connectivity and adaptive divergence of juvenile red rock lobster *Jasus edwardsii*. Panels of putatively neutral and putatively under selection markers were developed and validated for the juvenile and adult + juvenile rock lobsters to permit population differentiation analysis. Allele frequencies of the neutral markers were homogenous, which indicated high levels of gene flow. Pueruli/juvenile population differentiation patterns discovered using markers under selection were very similar to those for adult lobsters: two genetically different regions were identified (North-East and South). Unfortunately, the presence of the previously described North – Western region wasn't tested due to the lack of juvenile samples from the area due to low levels of settlement. Contrasting patterns of neutral and outlier markers indicate a phenotype – environment mismatch, which results in post-settlement mortality of non-adapted genotypes (Marshall et al. 2010). However, my analysis of the juvenile samples demonstrated high similarity between adult and juvenile divergence patterns, which is indicative of local selection acting on earlier stages of *Jasus edwardsii* development than were sampled here.

What can patterns of divergence at neutral vs outlier loci reveal about post-settlement barriers to connectivity?

All the datasets analysed (including the adult only dataset investigated in **Chapter 3**) consistently show a lack of genetic divergence of neutral markers, which is indicative of effective dispersal and high levels of gene flow between all the populations studied. Contrastingly, markers under selection demonstrate genetic divergence between three broad regions (NW, NE and S). This pattern of divergence of loci is indicative of a phenotype – environment mismatch type of post-settlement connectivity barrier (Marshall et al. 2010). However, the selective pressure affecting the post-settlement mortality is not strong enough

to exclude the settlement of non-adapted genotypes completely, which is evident from the low levels of admixture between the regions and is confirmed by the lack of differentiation of the neutral loci. Despite a relatively small number of migrants arriving into an apparently non-optimal environment and surviving into maturity and reproducing, it appears to be sufficient to homogenise the allele frequencies of the next generation of lobsters. The selective post-settlement mortality of juveniles has been reported in other marine species with wide larval dispersal such as the common tropical damselfish *Neopomacentrus filamentasus* (Vigliola et al. 2007), the Caribbean octocoral *Eunicea flexuosa* (Prada et al. 2014), and the Baldchin groper *Choerodon rubescens* (Cure et al. 2017). Allele frequencies of the putative mitochondrial control region have been shown to change between early settlers and 3-month old damselfish recruits that coincides with size-selective mortality among settlers. Genetic variation in this mitochondrial control region has been described to affect the metabolic rate and weight gain and a haplotype-phenotype link has been confirmed by artificial size selection (Vigliola et al. 2007). Two lineages of Caribbean octocoral have arisen as an adaptation to different depth and light conditions with strong evidence of both morphological and genetic clines in the coral and its algal symbiont, which was consistent with translocation studies (Prada et al. 2013). Interestingly, patterns of lineage segregation were stronger in adults when compared to juveniles, which was attributed to a phenotype – environment mismatch and prolonged pre-reproductive selection of non-optimal settlers (Prada et al. 2014). A study on juvenile Baldchin groper using genome-wide SNP markers detected differentiation patterns of outlier loci most likely produced by local temperature adaptation along with low levels of population divergence inferred from neutral markers, which was consistent with a previous microsatellite study of adult populations (Cure et al. 2017).

Comparison of population differentiation patterns between adult and pueruli/juvenile lobsters

Markers putatively under selection identified two distinct populations within NZ: North – East of the North Island and the Western part of the South Island. This is consistent with the absence of barriers to gene flow based on the analysis of adult *Jasus edwardsii* (**Chapter 3**). Only four sampling locations were analysed in this chapter, which limits the conclusions that can be made about broader patterns of differentiation, however, the population structure discovered is very consistent with the genetically distinct regions described in Chapter 3. I

performed a full range of analyses (F_{ST}, DAPC and Structure) using outlier markers on the juvenile and adult datasets combined to further test the apparent consistency between these two datasets. F-statistics, DAPC and Bayesian clustering demonstrated similarity in the genetic makeup of the juveniles from locations WHA, NAP and CAS to the genotypes of the individuals from the NE region. Similarly, juvenile lobsters from the JAC location clustered together with adult lobsters from the S regions. Despite general similarity of population differentiation patterns between Adult, Juvenile and Adult+Juvenile datasets this appears to not be the case for analysis of neutral+outlier marker set. In Adult and Adult+Juvenile datasets combined set of markers tends to demonstrate differentiation patterns more similar to purely outlier markers than to the purely neutral. However, in Juvenile dataset, only inferences from neutral+outlier markers appear to be closer to inferences from purely neutral markers. The likely explanation for this phenomenon is different proportions of outlier/neutral markers in the combined dataset. A much smaller number of outlier SNPs have been identified for Juvenile samples than for Adult (or Adult+Juvenile) samples: 14 vs 127 correspondingly. In addition to that, the Juvenile dataset contains a higher number of neutral markers - 2141 vs 1648 SNPs in the Adult dataset. This explanation is confirmed by the analysis of Adult+Juvenile dataset, where the increase in a number of outlier markers produces clearer differentiation patterns in the combined set of markers which closely resemble population differentiation of purely outlier markers.

Interestingly, a much smaller number of SNPs under divergent selection were identified for juvenile lobsters compared to the adults – only 14 vs 127 SNPs. In this study we used a methylation sensitive endonuclease – AciI in order to detect changes in gene expression. Chromatin structure and DNA methylation have been described as the main mechanisms of gene expression regulation at the DNA level (Grewal et al. 2003). Methylation patterns have been shown to differ between larval and adult stages and control the metamorphosis via programmed death of the larval cell lineage and differentiation of adult cell lineage (Jacobs et al. 2006; Azumi et al. 2007). Thus, if the major changes in the methylation patterns were to occur between juvenile and adult lobsters, restriction with EcoRI – AciI enzyme pair would produce adult and juvenile libraries with different sets of DNA fragments (Jiang et al. 2016). Consequently, a noticeably lower number of markers would be shared between adults and juveniles combined when compared to the number of loci for adults and juveniles separately. However, analysis of juvenile and adult dataset combined yielded 127 outlier and 1639 neutral SNPs, which is comparable to the number of SNPs identified for adults only and is a

much larger number of outlier SNPs than discovered for juvenile dataset only. This discrepancy between the datasets most likely arises from the lower number of juvenile locations sampled, especially the absence of any genotypes from the North – West region of NZ which have been identified as genetically distinct in the **Chapter 3**. Outlier calling performed on a subset of adult individuals from the four sampling locations closest to the sampling locations of juveniles confirmed this hypothesis by consistently identifying only 10 SNPs as being putatively under selection (**Appendix 3**). This suggests that more extensive sampling will produce a highly similar pattern of differentiation in the juvenile dataset compared to the adult dataset. In support of this hypothesis, population differentiation analysis of the combined adult and juvenile data revealed genetic divergence patterns similar to that observed in the adult dataset only with poorer discriminating power for the NW region. This result indicates that selective mortality of less adapted migrant genotypes has occurred prior to the last pre-metamorphic pueruli stage/early post-metamorphic juvenile stage. However, *Jasus edwardsii* has a highly complex life cycle with multiple instances of drastic morphological and functional change between life stages: metamorphosis of egg into naupliosoma, naupliosoma into phyllosoma which includes 17 instars, phyllosoma into pueruli and pueruli into juvenile.

The adaptive differentiation patterns of adult and late pueruli/early juvenile lobsters appear to be very similar, despite a limited number of sampling locations in the juvenile dataset. This similarity indicates that the selective pressure has been exerted on a previous lobster developmental stage, long enough ago to fully form the pattern of genetic divergence observed in both the juvenile/pueruli and adult lobster datasets. Results obtained in this study also allow me to exclude mortality of subadults and adults as a mechanism for the local adaptation pattern observed. Thus, the early pueruli stage of *Jasus edwardsii* is the most likely target for the local selection, however this hypothesis requires more extensive sampling of preceding life stages and from all three genetically distinct regions identified.

Chapter 5

Temperature as a driver of adaptive population divergence of the red rock lobster *Jasus edwardsii* across New Zealand
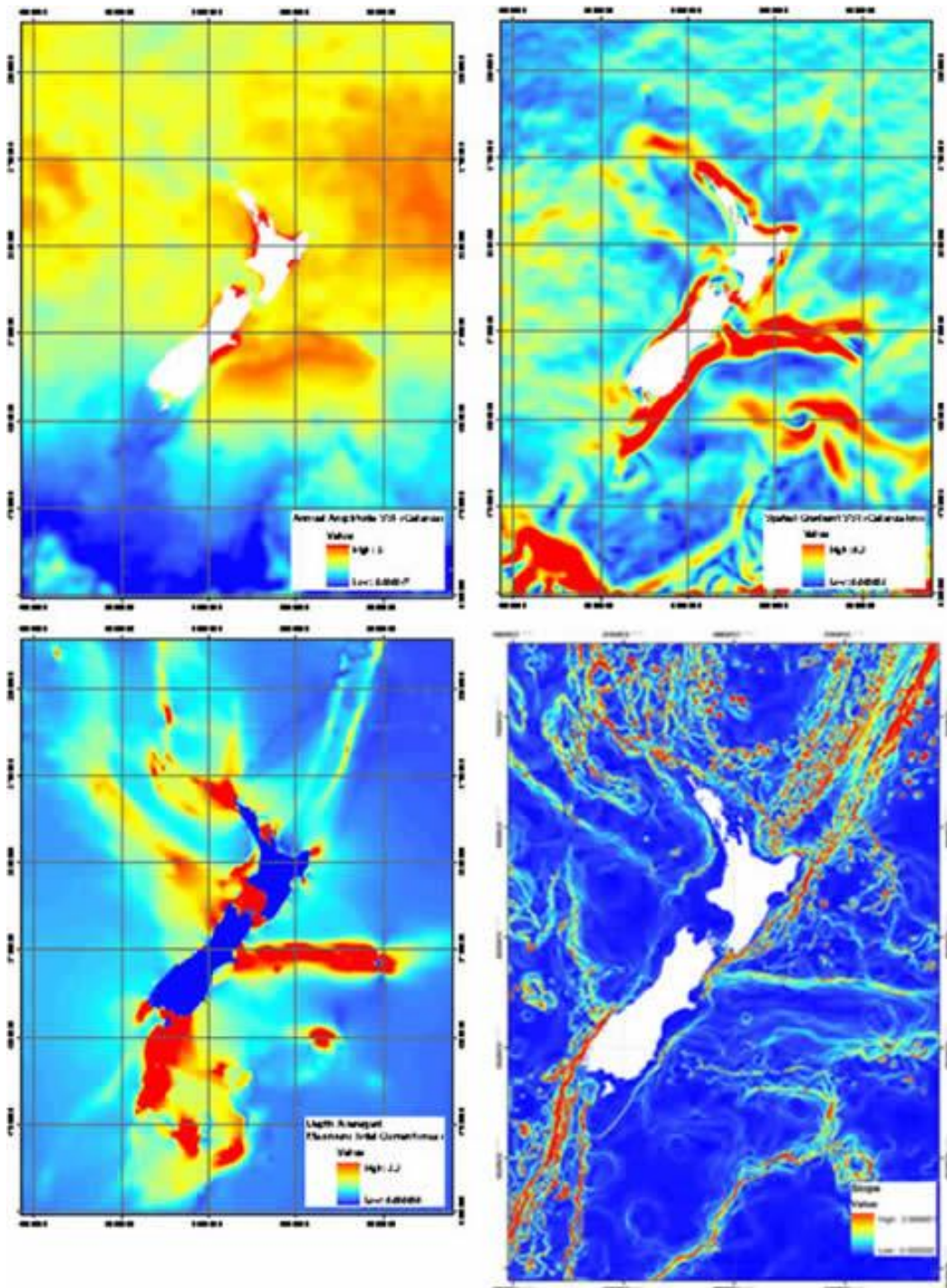
Image credits: The New Zealand Marine Environment Classification (Snelder et al. 2005).

Abstract

Selective pressure resulting from the environment shapes the genotypes at multiple loci of populations and can promote adaptive changes and local divergence. The importance of incorporating information about the factors underlying adaptive divergence of fished species has been recognised for some time as it is important for the conservation and sustainable management of these species. In this chapter I aimed to explore environment–genotype associations of the commercially important red rock lobster, *Jasus edwardsii*. Lobsters from 21 populations throughout New Zealand were genotyped and range of analyses were used to identify 127 SNPs that were putatively under selection. Biological Environment Stepwise (BEST), redundancy analyses (RDA)and generalized linear modelling (GLM) were used to identify a significant correlation between the annual amplitude of sea surface temperature and adaptive population divergence. In addition, the RDA highlighted the contribution of spatial distribution to the formation of adaptive population divergence. An environmental association analysis identified 43 SNPs as candidate loci for annual amplitude of sea surface temperature (SST) adaptation. Furthermore, alignment of translated reference catalog sequences containing the candidate SNPs to databases of annotated protein sequences offered a candidate gene for thermal adaptation - UDP-glycosyltransferase. This is an important enzyme involved in the detoxification of a variety of environmental pollutants and its activity and gene expression patterns have been linked to temperature. My results have identified a possible mechanism for local adaptation in the red rock lobster: adaptive divergence appears to be driven by the latitudinal gradient of SST amplitude. The gradient of environmental selection together with local patterns of connectivity produce three distinct temperature adapted genotypes (North-West, North-East and South) with clines of admixture of locally adapted genotypes between these regions.

Introduction

Marine environments generally lack obvious barriers to dispersal, which is predicted to result in high levels of gene flow and subsequent homogenisation of genotypes across large geographical distances (Levin 2006). However, more recent studies have reported population differentiation and self-recruitment in a number of marine species (Almany et al. 2007; Puebla et al. 2012; Benestan et al. 2015; Pocwierz-Kotus et al. 2015). Furthermore, it is becoming increasingly apparent that environmental gradients can influence local recruitment success and counteract the effects of the gene flow, which results in locally adapted genotypes and has been reported in a number of marine species (Orsini et al. 2013; D'Aloia et al. 2015). Non-random mortality associated with dispersal and settlement to the new habitats can occur as a result of a decreased fitness caused by the mismatch between a migrant's adaptations to its parental environment and the new environment it is attempting to settle in (Marshall et al. 2010). This mechanism was first described by DeWitt et al. (1998), who coined the term phenotype–environment mismatch. Two mechanisms have been described to explain the formation of phenotype-environment mismatches: evolution of phenotypic plasticity and local adaptation (Marshall et al. 2010), although a combination of these two processes has been considered to best describe adaptation to heterogeneous environments (DeWitt et al. 2004). Although phenotypic plasticity offers a higher chance of survival when dispersal is passive and migrants have little choice in their final settlement location, it may come at a cost that limits its fitness advantage (DeWitt et al. 1998; Relyea 2002; Van Buskirk et al. 2009). Phenotypic plasticity is also hypothesised to be constrained by number of factors: inability to produce the more extreme phenotypes at the limits of a species distribution range, time required for trait adjustment, ontogenic changes and environmental factors not connected to the trait under selection preventing the expression of the most optimal phenotype (Auld et al. 2010). On the other hand, any fitness benefit of local adaptation inherently reduces the performance of adapted genotype in alternative environments and is likely to be selected against (Hereford 2009). Both of these strategies (if plasticity is viewed as a trait) can also be constrained by a lack of genetic variation, genetic linkage to another trait or phylogenetic history (Pigliucci 2001; Bridle et al. 2007).

Although the existence of the genes responsible for modulating the level of plasticity (as a trait *per se*) is still under debate, it has been hypothesised that adaptations for plastic responses are likely to be connected to regulatory genes, while local adaptation tends to directly affect the trait under selection (Windig et al. 2004; Auld et al. 2010).

Adaptive divergence produced by any phenotype (genotype) – environment mismatch is thought to have important consequences for conservation and sustainable management (Selkoe et al. 2008; Iacchei et al. 2013; Selkoe et al. 2016a). For example, the population structure of white hake has been attributed to water depth and is potentially connected to differences in water temperatures at different depths (Cimmaruta et al. 2005). The population boundaries were not consistent with established fishery management units with the smallest and most genetically isolated population being the most exploited. This past study demonstrates the importance of understanding the genetic structure of harvested populations in the context of variable environments (Roy et al. 2012). Similarly, knowledge of habitats promoting genetic diversity and adaptive resilience is valuable in marine reserve design (Gebremedhin et al. 2009; Matala et al. 2014; Selkoe et al. 2016b). The protection of habitats promoting adaptability and increasing genetic diversity becomes even more important in the context of climate change and climate-induced range shifts. Knowledge of the diversifying parameter/parameters of the environment is essential to maintain resilience and genetic diversity of the species and can offer long-term conservation strategies (Allendorf et al. 2010).

Advancement of molecular genetics and improvements to statistical and computational tools have supported the combination of population genetics with landscape ecology creating the field of landscape genetics – a study of how geospatial and environmental features can impact microevolutionary processes (environmental selection, gene flow, genetic drift) (Manel et al. 2003). Seascape genetics is a sub-discipline of landscape genetics and is applied to marine environments. There are major challenges for applying these approaches to linking genetic and environmental variation in marine systems, for example, marine environments often exhibit higher temporal variability and connectivity than terrestrial environments (Liggins et al. 2013). There is an increasing number of seascape genetic studies highlighting the importance of understanding the relationship between the environmental and genetic variation. Temperature variation is one of the most commonly studied drivers of adaptive divergence and has been reported in a number of marine animals. Parallel allele frequency clines of 40 temperature adapted SNPs have been identified for Atlantic cod on both sides of the Atlantic Ocean (Bradbury et al. 2010). A seascape genetics study of American lobster revealed that divergence of selectively neutral SNPs originated mainly from ocean current-mediated larval connectivity (21%), while minimum annual sea surface temperature (SST) explained 8.1% of divergence of 28 outlier SNPs (Benestan et al. 2016). Isolation-by-

adaptation has also been proposed as a mechanism of population differentiation and was discovered in North Atlantic shrimp using ten microsatellite markers, where 26% of genetic divergence was explained by bottom temperature (Jorde et al. 2015). Diversifying selection of a single codon in the mitochondrial cytochrome b gene was detected in one of the mitochondrial DNA clades of the European anchovy and temperature explained 59 % of the genetic variance between spatially distant clades (Silva et al. 2014).

Another broad environmental gradient is salinity, which has been reported to produce adaptive fine-scale genetic differentiation between populations of Atlantic herring; a number of outlier SNPs identified in that study belonged to the genes associated with physiological response to salinity and osmoregulation (Lamichhaney et al. 2012). Results from laboratory experiments on the correlation of allele frequency variation in response to different salinity treatments in European sea bass were confirmed by population genetics studies reporting genetic differentiation between sea and lagoon populations at a single gene-associated microsatellite marker (Guinand et al. 2015). Twenty genic (i.e. located within a gene) microsatellite markers demonstrated a significantly higher population divergence than non-genic markers in threespine sticklebacks, additionally, 45% of genic markers were identified as F_{ST} outliers, and their allele frequency variation was correlated with differences in salinity (DeFaveri et al. 2013).

Both sea surface temperature and salinity have been significantly correlated with allele frequencies of outlier SNPs indicating adaptation to the local environment in European hake (Milano et al. 2014). Nine transcriptome-derived outlier SNPs markers have been associated with temperature and salinity differences between North Sea and North Atlantic habitats of Atlantic herring (Limborg et al. 2012) Outlier SNP detection combined with landscape genetic methods revealed local adaptations to differences in salinity, oxygen and temperature between the North Sea and the Baltic Sea in Atlantic cod (Berg et al. 2015).

Many seascape genetics studies have historically employed neutral microsatellite genetic markers for inferring environmental barriers to the gene flow that might cause divergence of local populations via genetic drift (Slatkin 1987). Variation of non-neutral markers might have a direct effect on fitness and survival into the reproductive stage of the individuals under selection. Neutral marker variation does not bring any fitness advantages, therefore it is not informative with regard to the adaptive or evolutionary potential of the population (Holderegger et al. 2006). Despite that, neutral markers remain a useful tool for population

connectivity assessments and in identifying demographical processes like migration/gene flow and changes in population size (Riginos et al. 2013). In the absence of obvious geographical barriers to gene flow in marine environments, seascape analysis of neutral markers can provide insight into biological barriers or environmental selection that consistently reduced or counteracted the gene flow, thus allowing for local adaptation over a number of preceding generations (Wei et al. 2013; Constable 2014; Hannan 2014; Silva et al. 2015; Benestan et al. 2016). However, neutral markers only reflect long-term connectivity patterns of populations, while markers under adaptive selection provide information about more recent processes (Holderegger et al. 2006). The divergence of markers under selection across habitats can be linked to environmental factors and ultimately may lead to the identification of candidate genes on which selection is acting, which can provide insights into the genotype-environment interactions (Schmidt et al. 2008).

The red rock lobster *Jasus edwardsii* (Hutton 1875) (Palinuridae) inhabits rocky reefs, between 1 m and 200 m depth, along the coastlines of southern Australia, Tasmania and New Zealand. In New Zealand, its range extends from Three Kings Islands in the north (34° S, 172° E) to the Auckland Islands in the south (51° S, 166° E), and to the Chatham Islands in the east (44° S, 176° W) (Kensler 1967). Thus, lobster populations inhabit a number of environmentally distinct locations within New Zealand's boundaries that appear to impose selective pressure sufficient to counteract high levels of gene flow (**Chapter 3**). The New Zealand marine environment ranges from subtropical to subantarctic waters and as such conditions vary along multiple environmental gradients on several spatial scales, which provides a range of selective conditions that might explain the adaptive population differentiation patterns of *Jasus edwardsii* described in **Chapter 3**. The largest scale refers to variation across large geographical distances such as latitudes. Environmental parameters most likely to vary on this scale are temperature (-8.1 – 9.9 °C) and salinity (34.2 – 35.6 psu) (Leathwick et al. 2009). On an intermediate scale, for example an estuary or lagoon, differences in sediment type and size, freshwater runoff, tidal currents and organic matter dissolved in the water can create distinct habitats or environmental clines. Finally, some environmental parameters, like tidal height, can affect genetic variation on a fine geographical scale (Schmidt et al. 2008). However, temperature has been reported to affect all of the life stages of rock lobsters: phyllosoma (Smith et al. 2002; Bermudes et al. 2008), postpueruli (Dubber et al. 2004), juvenile (Crear et al. 2000) and adult (Thomas et al. 2000) making it one of the top candidates for promoting adaptive divergence in *Jasus edwardsii*. In

the case of *Jasus edwardsii*, the effective dispersal of the next generation migrants appears to negate the effects of local selection to some extent. Despite the selective pressure exerted by the environment, the levels of an admixture of genotypes of three broad regions within NZ investigated in **Chapter 3** indicate occasional settlement and survival into the reproductive stage occurs in non-optimal conditions. Mating between local and migrant genotypes homogenises allele frequencies and is reflected in the absence of any population differentiation signal from neutral markers, thus investigating neutral genetic variation in relation to the local environmental conditions is not likely to produce meaningful associations.

Therefore, the aim of this chapter is to identify the environmental factors promoting local selection and adaptive genetic divergence between populations of the red rock lobster *Jasus edwardsii* in New Zealand.

Specific aims of this chapter are:

- To elucidate the contribution of spatial and environmental factors to adaptive divergence patterns of *Jasus edwardsii*.
- To conduct environmental association analysis to identify candidate SNPs for adaptive selection exerted by the specific environmental variables identified.
- To identify the possible function of loci under selection and confirm possible association with environmental factors.

Materials and methods

Genetic data

A panel of 127 putatively under selection SNPs was employed to assess potential environmental associations with the patterns of differentiation in *Jasus edwardsii*. Data for 611 individuals collected from 21 statistical areas within NZ were used for further analysis (stat areas 915 and 934 were pooled because they were collected together and only 1 geographical point was provided, see **Chapter 3**, **Fig. 3.1**). For distance-based linear model (DistLM)/BEST analysis allele frequencies were calculated for each locus using Genepop 4.5.1 (Rousset 2008). Pairwise Wright's fixation index (F_{ST}) was calculated for the 21 statistical areas using Arlequin ver 3.5.2.2 (Excoffier et al. 2010) and averaged across sampling locations to obtain a single descriptive value for each population in order to perform univariate GLM analysis. For redundancy analysis (RDA) a SNP marker minor allele frequency (MAF) was calculated for each sampling location using Plink v.1.9 and detrended in order to improve the association with spatial vectors using the function *decostand* from the vegan package in R (Oksanen et al. 2013). MAF variation among genotypes was summarized using Principal Component Analysis (PCA) and Principal Components (PCs) with eigenvalues >1 were retained for association with geographical and environmental data.

Spatial data

Following the analysis performed by Wei et al. (2013), three spatial variables were used: latitude (Lat), longitude (Lon) and an index of geographic distance (Geo_dist). Latitude and longitude were obtained from the GPS coordinates of sampling locations and expressed in decimal degrees. An index of geographic distance was calculated as the sum of the shortest possible coastal distances between each pair of populations in R using the geo*sphere* package.

Following Benestan et al (2016), GPS coordinates of each sampling location were converted into Cartesian coordinates with the *geoXY* function available in the SoDa package in R v. 3.1.3 (Team R core 2014). A Euclidian distance matrix was calculated using the function *dist* and distance-based Moran's eigenvectors, which are suitable for constrained ordination (Borcard et al. 2012), and obtained with the PCNM function (PCNM package of R software),

(Dray et al. 2006). The vectors obtained are independent vectors that summarize the spatial relationship between the sampling locations (Perez et al. 2010).

Environmental data

Site-specific environmental variables were obtained from New Zealand Marine Environment Classification (MEC) (Snelder et al. 2007) and Benthic-optimized Marine Environment Classification (BOMEC) (Leathwick et al. 2009) datasets. The MEC dataset is comprised of gridded data layers of environmental variables for the NZ Exclusive Economic Zone (EEZ), while BOMEC defines a number of ecosystem types comprehensively describing the variation of benthic environments throughout NZ. Despite differences in the methods used to compile these classifications, both datasets have a resolution of 1 km, which allows MEC and BOMEC environmental variables to be combined. Total of 18 environmental variables (11 from MEC and 7 from BOMEC) were extracted from GIS raster and vector shape files mapped against sampling sites in ArcMap v 10.4.1 (**Table 5.1**).

Table 5.1. List of environmental variables obtained from the New Zealand Marine Environment Classification (MEC) and Benthic-optimized Marine Environment Classification (BOMEC).

MEC			
Environmental variable	Abbreviation	Description	Units
Depth	Depth	Bathymetry grid (1 km resolution)	m
Annual mean solar radiation	Rad_mean	Mean extra atmospheric solar radiation modified by mean annual cloud cover	Wm^{-2}
Winter solar radiation	Rad_wint	Extra atmospheric solar radiation in June, modified by mean annual cloud cover	Wm^{-2}
Wintertime sea surface temperature	SSTwint	Mean of daily data from early September when SST is typically lowest	°C
Annual amplitude of sea surface temperature	SSTanamp	Smoothed annual amplitude of SST	°C
Spatial gradient annual mean sea surface temperature	SSTgrad	Smoothed magnitude of the spatial gradient of annual mean SST	°C km^{-1}
Summertime sea surface temperature anomaly	SSTanom	Spatial anomalies with scales between 20 and 450 km in late February when SST is typically highest	°C
Mean orbital velocity	Orb_v_mean	Orbital velocity at the bed for the mean significant wave height calculated from a 20-year wave hindcast	m/s

Tidal current	Tidal	Depth-averaged maximum tidal current	m/s
Seabed rate of change of slope	Bed_slope	The rate of change of slope for each cell	$0.01\ m^{-1}$
Freshwater fraction	FW	The proportion of fresh water based on river inputs	proportion
BOMEC			
Environmental variable	Abbreviation	Description	Units
Temperature	Temperature	Temperature obtained from World Ocean Atlas normalized to a depth	°C
Salinity	Salinity	Salinity obtained from World Ocean Atlas	psu
Productivity	Productivity	Modelled	$mgCm^{-2}d^{-1}$
Sediment type	Sed_size	Sediment size (categorical variable)	mm
Sediment resuspension	Sed_resusp	Wave model and sediment data	Arbitrary units
Suspended particle matter	Susp_pm	Remote sensed	Arbitrary units
Dissolved organic matter	Diss_org_mat	Remote sensed	Arbitrary units

Variable selection

All environmental variables were tested for independence via a Pearson's correlation test and Principal Component Analysis (PCA) using R packages *Hmisc* and *stats.* Variables correlated at the 0.05 level were removed from further analysis. Vectors of spatial variation were treated as independent.

Genotype-environment association analysis

BEST analysis was carried out in the PRIMER v6 software environment with PERMANOVA (Clarke et al. 2006; Anderson et al. 2008) following Wei et al. (2013)(Wei et al. 2013). A Euclidean resemblance matrix was calculated for previously normalized environmental data and a Bray-Curtis resemblance matrix was calculated for the genetic data. To test for any correlation between the two matrices, the BIOENV procedure of the BEST routine was implemented. The BIOENV procedure (Clarke et al. 1993) aims to identify a subset of the explanatory variables whose Euclidean distance matrix correlates maximally with the Bray-Curtis dissimilarity matrix of genetic data. Further, the procedure implements the Spearman correlation coefficient method to rank all combinations of variables. Models were considered significant at $p<0.05$ after 999 permutations. SPSS v23 (Green et al. 2010) was used for GLMs, analyses were performed following Wei et al. (2013). The Akaike information criterion (AIC) was used to rank the models at the significance level of $p<0.05$. A test of all effects was performed using the Wald Chi-squared statistic.

Redundancy analysis (RDA) is an extension of multiple linear regression, which allows the assessment of multiple response variables and multiple explanatory variables. This method employs constrained ordination of linear combinations of explanatory variables along RDA axes that are designed to maximize their relationship with these variables. A model with the highest explanatory potential is selected according to the highest adjusted coefficient of determination. Redundancy analyses (RDA) were performed with *ordistep* and *rda* functions from the vegan package in R following Benestan et al. (2016). Analysis of variance (ANOVA) of the RDA framework was performed using 1000 permutations in order to estimate the global significance of the model. Furthermore, partial RDA (pRDA) was performed to examine the effect of the environment in the RDA framework of explanatory variables while controlling for the spatial variables and vice versa. To achieve this all other explanatory variables were placed in a matrix of control variables, which allows partitioning of their effects. pRDA produces a single RDA axis and eigenvalue representing how much variation the investigated variable accounts for. A marginal ANOVA with 1000 permutations was performed to assess the significance of each contribution of the selected environmental factors and spatial variables into adaptive genetic variation separately.

Candidate SNPs for local adaptation

The 127 SNPs described in the **Chapter 3** as being putatively under adaptive selection were tested for the association with smoothed annual amplitude of SST using LFMM v.1.4 (Frichot et al. 2013). Only the contribution of SST was studied further in order to identify candidate genes because it was the only environmental variable to have significant influence on adaptive divergence when analysed separately. Latent Factor Mixed Models (LFMMs) are linear regression models in which the response variable is a genotype, and the explanatory variable is an environmental (or spatial) variable. LFMM uses an MCMC algorithm for regression analysis based on the residuals of PCA. Additional factors are designed to partition genetic variation unexplained by the environmental variables, which allows LFMM to account for demographic history or population genetic structure. The number of genetic clusters used in LFMM analyses was determined using the *snmf* function, which calculates ancestry coefficients using the sparse non-negative matrix factorization. Several runs of *snmf* function with the most favourable numbers of ancestral populations were assessed as advised by the package authors. Candidate SNPs for local SST amplitude identified with LFMM have been tested further using logistic regression approach. GLM with a binomial error was employed to model the probability of allele "A" to be one of the two alleles. Candidate genotypes were assessed for significant correlations with the SSTanamp parameter by Poisson regression.

For candidate SNPs identified as correlated with the amplitude of SST corresponding DNA sequences from the reference catalog was used for a Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) search against the non-redundant protein database RefSeq and the more well annotated SWISS-PROT database (Bairoch et al. 2000). The BLASTx tool was employed in order to search protein databases using a translated nucleotide query.

Results

Correlation with environmental variables

Environmental variables were ranked using PCA, (**Fig.5.1**) and variables with the highest proportion of explained variance that were not correlated with other variables (at 0.05) (**Fig.5.2**), were retained for further analysis: Bed_slope, Orb_v_mean, SSTanamp, SSTwint, and Salinity.

Fig.5.1. Principal Components Analysis of all environmental variables. Black dots represent sampling locations; blue arrows represent environmental variables (abbreviation per **Table 5.1**). The length of the arrows is proportionate to the variance explained by that variable.
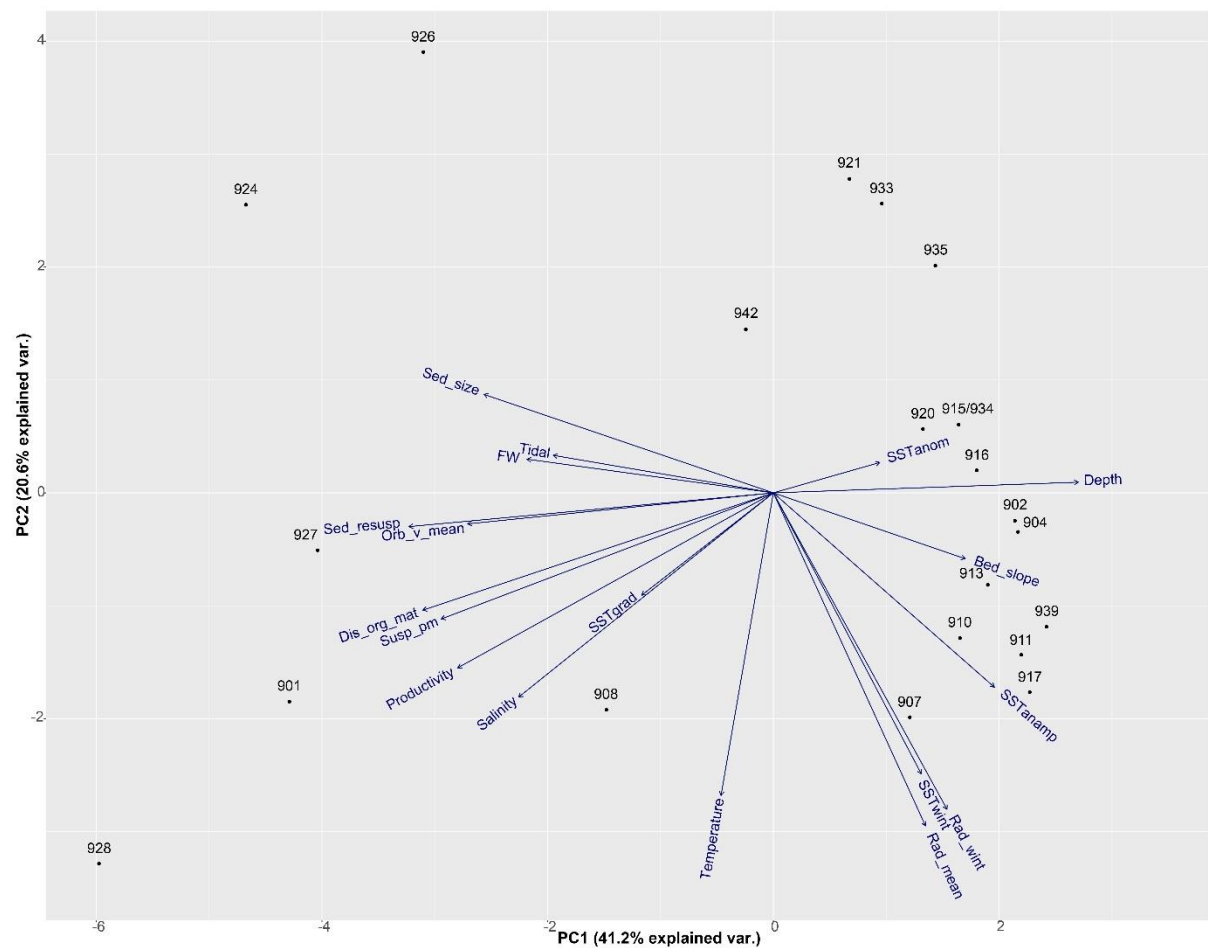
Fig.5.2. The plot of Pearson's correlation of environmental variables. Red color indicates negative correlations, blue color – positive correlations. Colour intensity and the size of the circle are proportional to the correlation coefficients.
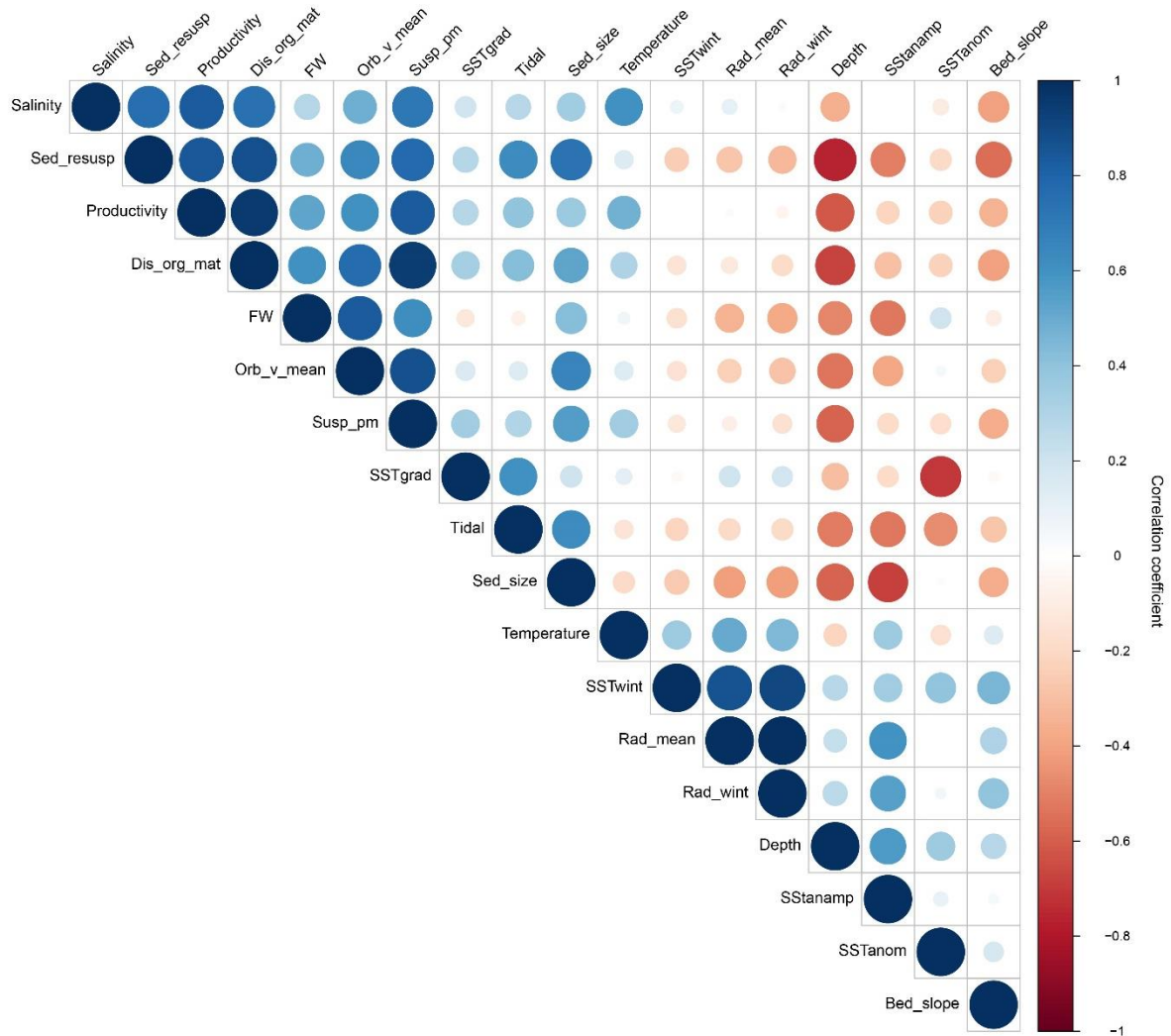
Generalized Linear Models

The pairwise F_{ST} values between sampling locations ranged from 0.002 to 0.053 with calculated mean F_{ST} values ranging from 0.02 for statistical area 935 to 0.03 for area 928 (**Appendix 4**).

Generalized linear model analysis based on data from the 21 statistical areas identified smoothed annual amplitude of SST (SSTanamp) as the single most significant factor (p = 0.036) with a Wald Chi-Square of 4.41 and standard error of 0.0025. A generalized linear model consisting of only SSTanamp produced an adjusted coefficient of determination of 0.13, which was globally significant (p = 0.00531). None of the three spatial variables used in GLM analysis (following the approach used by Wei et al. 2013) were significant at a p-value threshold of 0.05 and were not included in the models.

Biological Environmental STepwise

BEST analyses showed that the top best fitting model was significant at p < 0.05 and included the variables SSTanamp and SSTwint (r_S = 0.337, p-value = 0.026). The two remaining best fitting models had Spearman's r values of 0.331 and 0.33 and were significant (p < 0.05). The sstanamp variable was included in all the best fitting models (**Table 5.2.**).

Table 5.2. Results of BEST analysis presenting Spearman's r values for the top best fitting models.

# of Variables	r_S	Variables
2	0.337	SSTanamp, SSTwint
3	0.331	SSTanamp, SSTwint, Bed_slope
1	0.33	SSTanamp

Redundancy analysis

MAF variation between sampling locations have been decomposed into 21 primary components and three PCs were retained for RDA analysis based on the Kaiser–Guttman criterion ($(\lambda) > 1$) (Jackson 1993), which together accounted for 62.6% of the adaptive genetic variation. The constrained ordination method selected smoothed annual amplitude of SST (SSTanamp) and two spatial vectors (PCNM1, PCNM2) for the RDA framework. The first two RDA axes accounted for 64.6% and 21.1% of the genetic variation, respectively (**Fig.5.3.**). The RDA of environmental and spatial predictors combined was globally significant ($p = 0.001$) and had an adjusted coefficient of determination of 0.609.

Fig. 5.3. The plot of RDA framework showing the relationship between the distribution of 21 sampling locations and explanatory variables. Black arrows represent spatial vectors (PCNM1 and PCNM2) and temperature (SSTanamp). Markers are filled according to the three previously identified (**Chapter 3**) genetically distinct regions: NW (green), NE (blue) and S (red).
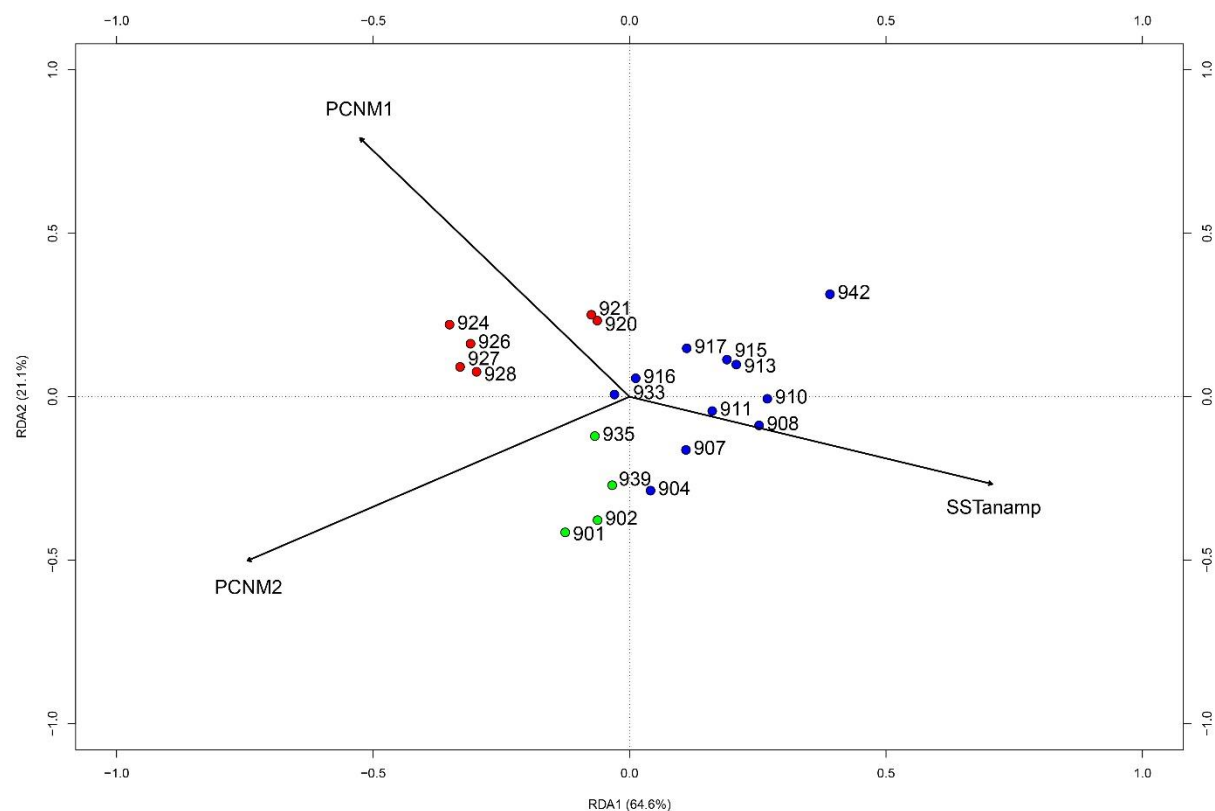
pRDA and ANOVA performed on the single environmental variable included in the model (SSTanamp) showed that annual SST amplitude *per se* is a significant predictor of the adaptive genetic divergence ($R_{adj}^2 = 0.305$, p= 0.001) (**Fig. 5.4**). On the other hand, pRDA performed on the spatial variables (PCMN1 and PCNM2) demonstrated relatively small and non-significant contribution of spatial variation to the pattern of adaptive population differentiation of red rock lobster ($R_{adj}^2 = 0.004$, p = 0.325) (**Fig. 5.5**).

Fig. 5.4. pRDA showing the distribution of 21 sampling locations in relation to SSTanamp along a single RDA axis. Sampling locations are coded based on the three previously identified (**Chapter 3**) genetically distinct regions: NW (green), NE (blue) and S (red).
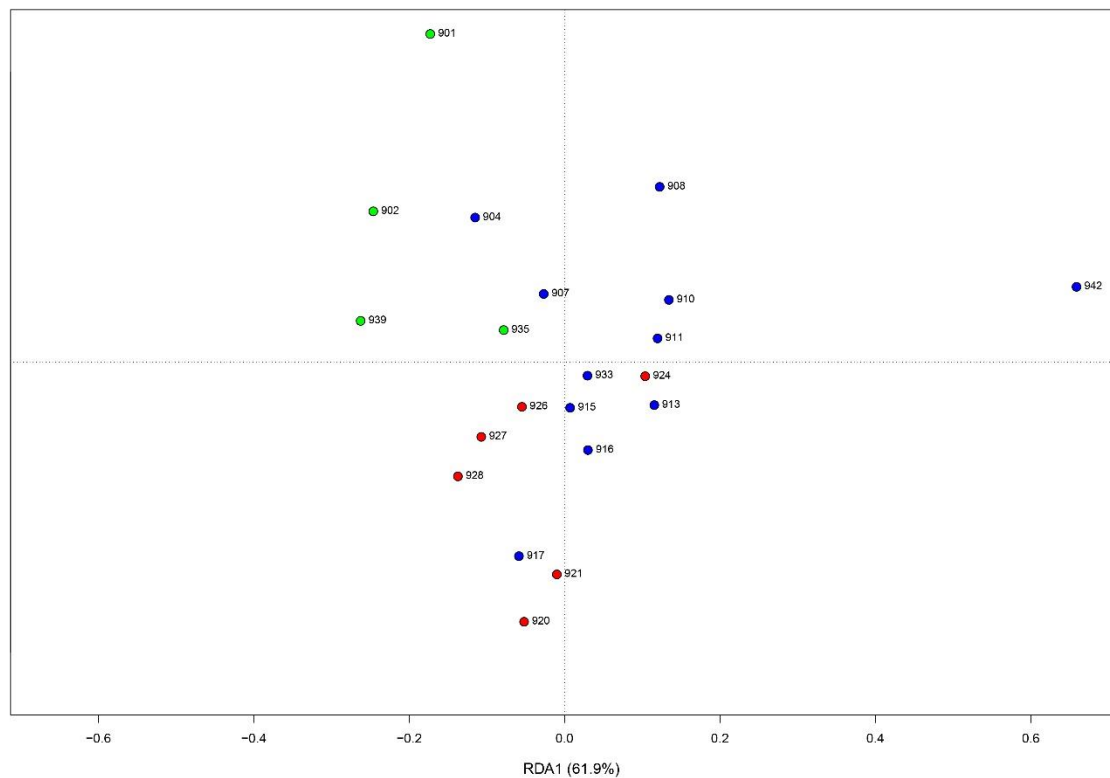
Fig. 5.5. pRDA showing the distribution of 21 sampling locations in relation to spatial explanatory variables along a single RDA axis. Sampling locations are coded based on the three previously identified (**Chapter 3**) genetically distinct regions: NW (green), NE (blue) and S (red).
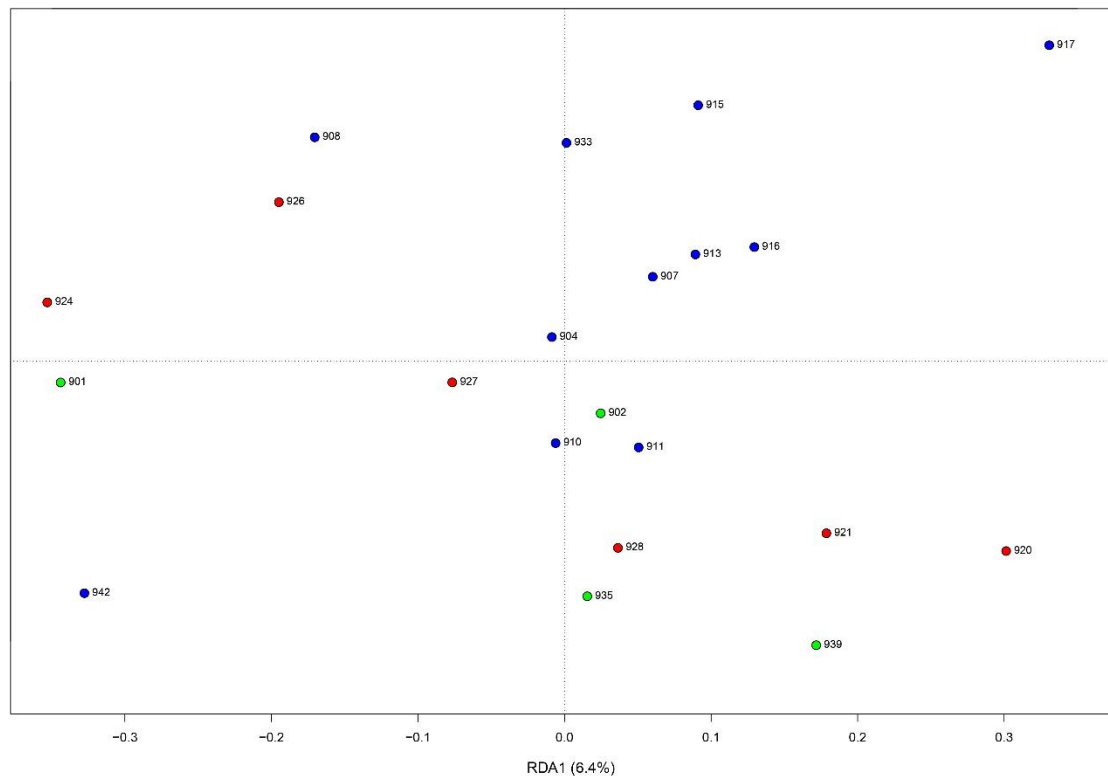
Candidate SNPs

The LFMM method identified 43 candidate SNPs as being significantly associated with smoothed annual SST amplitude. The significance of the environmental association for all of 43 candidate markers was confirmed by the analysis of logistic regression. BLASTx searches against the RefSeq database identified five matches which are shown in **Table 5.3**. For sequence 92714_782 only one top hit has been shown due to the rest of the matches being to the protein from the same protein superfamily but belonging to different organisms and with considerably lower E-values. A UniProtKB/Swiss-Prot database search for sequence 92714_782 yielded multiple matches to the same protein - UDP-glycosyltransferase with the top hit having an E-value of 0.002, 60% of identity and covering 63% of the query (UniProtKB/Swiss-Prot ID: O97951.1).

Table 5.3. Results of BLASTx searches against of DNA sequences associated with the annual amplitude of SST.

Reference catalogue ID	Accession ID	% identity	E-value	Query cover, %	Description
92714_782	AHX56958.1	59	0.00049	91	UDP-glycosyltransferase 214A1 [Strigamia maritima]
57301_11437	XP_015624625.1	72	9.4	56	PREDICTED: RING-H2 finger protein ATL80 [Oryza sativa Japonica Group]
3829_6668	XP_018959922.1	56	7.9	56	PREDICTED: potassium voltage-gated channel subfamily KQT member 5-like, partial [Cyprinus carpio]
61392_2223	PIN26099.1	50	9.4	69	Ferric reductase, NADH/NADPH oxidase [Handroanthus impetiginosus]
38735_1633	OCT55208.1	45	6.6	91	hypothetical protein XELAEV_18003867mg [Xenopus laevis]

Discussion

Investigating the factors underlying adaptive divergence of fished species is important for their conservation and sustainable management (Russello et al. 2012; Fitzpatrick et al. 2016). In order to investigate any environment–genotype association of adult *Jasus edwardsii* I performed a range of analyses on 127 SNPs identified as putatively under selection and discovered a significant contribution of the annual amplitude of sea surface temperature to adaptive population divergence. An environmental association analysis linked annual amplitude of SST to 43 candidate SNPs. The search for alignment with annotated protein sequences produced a candidate gene coding UDP-glycosyltransferase, which is an important enzyme involved in the detoxification response to a variety of environmental pollutants and the level of its activity and gene expression have been linked to the temperature (Stroomberg et al. 2004; Viñuela et al. 2011; Hu et al. 2015).

Genotype-environment associations

Different genotype-environment association tests, including GLM, BEST and RDA, identified SSTanamp (smoothed annual amplitude of SST) as the most significant predictor of the adaptive genetic divergence in *Jasus edwardsii*. SSTanamp represents the annual variation in temperature created by the seasonal stratification/mixing processes and is calculated via spatial smoothing of the annual harmonic of sea surface temperature across the region (Snelder et al. 2005). A potential limitation to this result is that only sea surface temperature estimates were available for the locations sampled, and adult lobsters are more likely affected by the bottom temperatures compared to the dispersive larval stages. However, air, sea surface and bottom temperatures tend to be correlated and they have been employed in a number of studies of the relationship between environmental and genetic divergence of marine organisms. In the absence of water temperature data, air temperatures have been successfully used as a proxy for surface temperatures and significant correlations with adaptive divergence have been reported in steelhead trout (Matala et al. 2014) and chinook salmon (Matala et al. 2011). Sea surface temperatures have been successfully associated with adaptive population differentiation patterns of seafloor dwelling organisms, such as the greenshell mussel and American lobsters (Wei et al. 2013; Benestan et al. 2016). When both surface and bottom temperature measurements were available for genotype-environment associations, Jorde et al. (2015) found that both parameters were significantly correlated with

genetic divergence, although bottom temperature accounted for a higher proportion of genetic variation. Results of these studies provide confidence in the genotype – sea surface temperature association found in this study.

The BEST analysis identified a combination of SSTanamp and SSTwint as the best fitting model, however, inclusion of winter SST into the model only marginally improved the Spearman's r values when compared to the model that consisted of only SST annual amplitude (0.337 for SSTanamp+SSTwint against 0.33 for SSTanamp alone). However, a different group of explanatory variables that wasn't limited to the SSTanamp variable was identified from the redundancy analysis and included two spatial vectors. In contrast to the BEST model, the RDA best fitting model included the PCNM1 and PCNM2 variables in addition to the explanatory temperature variable, which produced an almost twofold increase in the proportion of adaptive genetic variability explained (for SSTanamp the only $R_{adj}^2 = 0.305$, while for SSTanamp+PCNM1+PCNM2 $R_{adj}^2 = 0.609$). Eigenvectors generated by the PCNM function represent spectral decomposition of spatial connectivity. Essentially, they are independent variables that describe a variety of spatial patterns, like autocorrelation or repeated patterns (Borcard et al. 2004). This result highlights the importance of geographical location as a contributor to adaptive population differentiation, however, PCNM vectors do not easily translate into more relatable measures of distance like latitude or longitude. Results from the RDA analysis have been visualized using type II scaling, which emphasizes the relationship between the PC vectors (represented by the black arrows **Fig. 5.3.**). The correlation between the explanatory variables is equal to the cosine of the angle between the vectors in these plots, which can provide some insight into the relationship between those variables. Angles less than 90° suggest a positive correlation between vectors and angles of more than 90° are suggestive of a negative correlation (Legendre et al. 2012). The angle between the spatial vectors PCNM1 and PCNM2 is approaching 90°, which might be indicative of a relatively small correlation between these two variables. Other pairs of explanatory variables (SSTanamp and PCNM1, SSTanamp and PCNM2) meet at very similar angles that are noticeably larger than 90°, indicating a relatively strong negative correlation.

Environmental association analysis

Environmental association analysis with LFMM identified 43 markers as candidate SNPs for association with an annual amplitude of SST, which is 33.9% of all SNPs described as putatively under selection in **Chapter 3**. Generally, shorter query sequences mean much higher chances of finding a match in the database of particular size purely by chance (Karlin et al. 1990), which is reflected in the high E-values for most of the database matches found in the **Table 5.3.** Only one query sequence passed the E-value threshold of 1 x 10^{-6}, reference sequence 92714_782, which has been aligned to a conservative domain of UDP-glycosyltransferase in both RefSeq and Swiss-Prot databases. This enzyme belongs to a large superfamily that consists of many isoforms that can be classified according to the amino acid similarity into four gene families (Burchell et al. 1997). The highest quality match for the reference sequence 92714_782 is isoform 214A1, which belongs to a UGT2 family that is followed by a number of matches to the isoforms from both UGT1 and UGT2 families. Comparison of the UGT1 and UGT2 families indicated a high level of sequence conservation in the C-terminal domain that allows the enzymes to interact with a common UDP-GlcUA-binding site and most variable amino acids are found in the N-terminal domain responsible for substrate-binding (Iyanagi 2007). Many of the isoforms are produced via alternative splicing (Iyanagi 1991), which makes alignment of the translated DNA sequence to the protein sequence more challenging. Therefore, even though the query sequence may not be sufficient to identify a particular isoform, alignment to the conservative domain of UDP-glucuronosyltransferase provides confidence in connecting reference catalogue fragment to this protein and its function (**Appendix 5**).

Candidate loci for adaptive divergence

UDP-glycosyltransferase (UGT) is an enzyme that catalyses the conjugation of a variety of small lipophilic groups to sugars and is a major phase II enzyme involved in detoxification processes (Bock 2016). This catalytic reaction has been characterised in humans to produce lipophilic glucuronides from a variety of substrates: steroids, bile acids, bilirubin, hormones, dietary constituents, drugs and environmental pollutants (Tukey et al. 2000). Metabolism of natural and artificial compounds (xenobiotics) starts with a decrease in the biological activity of the substrate with cytochrome P450 monooxygenases (Phase I). In phase II toxic by-products generated in the previous phase are modified to increase their solubility by UDP-

glucuronosyltransferases (UGTs), glutathione S-transferases or carboxylesterases. During phase III transmembrane transporters actively export the conjugated toxins out of the cell (Misra et al. 2011). Stages of the detoxification process have been studied and most well described in humans and model animals, however, UDP-glycosyltransferase has been confirmed to occur in fish (Clarke et al. 1992; Leaver et al. 2007) and marine invertebrates and appear to perform a similar function (Livingstone 1998).

Even though UGT activity can be influenced by endogenous compounds as well as diet, non-model animal studies have focused on investigating increased UGT activity in response to a range of environmental pollutants (van der Oost et al. 2003). In addition to the investigation of the activity levels of UGT, an increase in mRNA expression of the UGT gene after exposure to an environmental pollutant, tributyltin, has been described in salmon (Mortensen et al. 2007). Interestingly, marginal differential expression of UDP-glycosyltransferase has also been reported in coral larvae in response to elevated temperatures (Polato et al. 2013). A combination of these two factors – toxicants and temperature – have been studied in depth in *Caenorhabditis elegans.* Animals have been exposed to combinations of chlorpyrifos, diazinon and two temperature regimes - 16°C and 24°C. Vinuela et al. 2011 found that in addition to the exposure to pollutants, the increased temperature treatment also upregulated expression of detoxification genes, especially genes with UDP domains. Variation in the MAF of the UGT gene detected in this study might be due to the variation of the proportion of CpG methylated alleles of the UGT gene across red rock lobster populations, which was identified by using methylation-sensitive enzyme AciI for ddRADseq library preparation.

Genotypic variations and differential expression of UDP-glucuronosyltransferase have been hypothesised to affect the survivorship of plaice *Pleuronectes platessa* due to the difference in ability to detoxify carcinogenic polycyclic aromatic hydrocarbons (PAHs) such as benzo(a)pyrene (George et al. 1998). Six alternative variants resulting from SNP variation in the UGT1B1 have been identified, however, they have been described as synonymous and most likely functionally neutral. Additionally, a variation in the protein originating from dinucleotide (AT) deletion in UGT1B1 gene produced a disadvantageous genotype with non-functional gene product. George et al. (2002) speculated that pollution–sensitive phenotypes observed in the wild are likely to have inherited the poorer-performing variant of UGT1B1 and survivorship of this phenotype in polluted environments is likely to be affected (George et al. 2002). A similar selective mechanism might be at work at 92714_782 loci with temperature interacting with bioaccumulation of environmental pollutant/pollutants thus

triggering differential expression of UGT genes and an overall increase in their activity. Interestingly, no selective variation has been observed at this locus in the NE and S regions, and it appears to underlie adaptive divergence of NW region that is characterised by the highest SST of the location surveyed (Snelder et al. 2005; Leathwick et al. 2009) and includes the most populated metropolitan area of New Zealand. Waitemata and Manukau Harbours and the Tamaki Estuary in Auckland region have been reported to contain above-background levels of PAH - one of the most common environmental pollutants (Depree et al. 2008; Park 2014). However, there are no data available on environmental pollutants from the specific locations sampled in this study.

Genotypic variations and differential expression of UDP-glucuronosyltransferases have been shown to produce disadvantageous phenotypes in humans (hyperbilirubinemia and autoimmune disorders) in the absence of environmental pollutants (Tukey et al. 2000). Unfortunately, no literature exists to date on the occurrence of similar conditions in non-model species although this data highlights the importance of detoxification processes for normal functioning of the organism and indicates that malfunctioning of the UDP-glucuronosyltransferase gene product most likely will significantly impact the survivorship even in relatively non-polluted conditions.

Even though variation at the locus associated with UDP-glycosyltransferase offers a compelling mechanism for selective mortality of non-adapted migrant genotypes, it is still only one of 43 loci showing adaptive divergence that was attributed to the amplitude of SST. Unfortunately, it wasn't possible to link the other 42 loci to specific genes with confidence, thus limiting insight into the molecular mechanisms underlying the local adaptations. Despite that, it is highly likely that gradient of annual temperature amplitude provides a range of selective environments and is likely to promote settlement of higher-temperature adapted genotypes in the more Northern parts of New Zealand and lower-temperature adapted genotypes in the Southern parts. However, due to several factors, we do not observe as clear-cut population structure as would be expected from selective pressure of temperature gradient. First of all, phenotypic plasticity of a small proportion of the migrants appears to allow them to survive into maturity in less suitable habitats (Ghalambor et al. 2007) and is not detectable by the methods employed in this study. Secondly, larval transport via ocean currents differs significantly between the East and West coasts of New Zealand; with highly effective mixing of the larvae via a system of eddies on the East coast of NZ and comparatively reduced dispersal along the West coast of New Zealand with only one major

current – the Westland current (Chiswell et al. 2010; Chiswell et al. 2011). Unequal production of locally adapted genotypes is contributing to the discontinuity of the temperature-adapted genotypic cline with lowest producers located along most of the west coast (CRA9) and lower part of the eastern South Island (CRA7) (Chiswell et al. 2008). Thus, effects of distance and difference in the connectivity patterns applied to the latitudinal gradient of temperature produce three distinct temperature adapted genotypes – NW, NE and S with clines of an admixture of locally adapted genotypes between these three regions.

Chapter 6

General discussion

Summary of key findings

This study aimed to assess patterns of connectivity of the red rock lobster *Jasus edwardsii* and investigate the mechanisms forming population differentiation patterns. To achieve this goal, SNP markers have been chosen as they permit analysis of both markers which are selectively neutral and those under putative divergent selection. For a non-model species where there is limited knowledge of the genome, the preparation of a reduced representation library and SNP discovery process requires careful optimisation. Therefore, in Chapter 1 of my thesis I optimized a ddRADseq protocol for specific features of the *Jasus edwardsii* genome and searched for appropriate bioinformatic tools that would provide sufficient marker discovery rates without sacrificing the quality or impact the downstream population genetics analysis. The molecular (modified ddRADseq) and bioinformatic protocols (Rad-loci pipeline) chosen have been further assessed using technical replicates which confirmed the adequacy of the data for population genetics inferences. Analysis of neutral and outlier markers revealed how two contrasting microevolutionaty processes shape the population structure of the red rock lobster: local selective pressure promotes genetic divergence of NZ lobsters into three broad regions (North-East NZ, North-West NZ and South NZ), but this is counteracted by effective larval dispersal ensuring high levels of gene flow and connectivity between populations. The strength of local selection appears to be insufficient to prevent admixture of alternatively adapted genotypes entirely and allows for mating with locally-adapted genotypes, which causes homogenization of the allele frequencies. Post-settlement mortality has been proposed as the central mechanism in the formation of the three genetically distinct regions and warranted investigation of population differentiation patterns of earlier life stages of the red rock lobster. Pueruli/juveniles exhibited patterns of population differentiation that were very similar to adult lobsters: neutral markers showed a high level of homogenisation, while outlier markers indicated a divergence between North-East NZ and South NZ (so samples from North-West NZ were available for analysis). Matching patterns of divergence of these two life stages provide evidence that selective mortality is occurring on earlier stages of red rock lobster development. In addition, population genetics analysis revealed low levels of admixture of non-local genotypes in both adult and pueruli/juvenile samples that appeared to form a latitudinal gradient of allele frequencies on both coasts of NZ. Analysis of environment–genotype associations, identified the annual amplitude of sea surface temperature (SST) as a single significant selective force explaining 30.7% of genetic variation according to a redundancy analysis. Although vectors of spatial variation *per se* did

164

not hold significant explanatory power, in combination with SST amplitude two vectors of spatial variation created a framework that accounted for 60.9% of the genetic variation. These results suggest that the latitudinal gradient of SST amplitude across NZ is a primary selective force creating a gradient of temperature – adapted genotypes, while local spatial structures also affect connectivity and contributing to the formation of three genetically distinct regions. Environmental association analysis identified 43 SNPs that correlated with the annual amplitude of SST, however, only one of them has been annotated as belonging to the gene coding UDP-glycosyltransferase (UGT) – a detoxification enzyme which activity and gene expression has been shown to respond to levels of environmental pollution and temperature.

Results of my thesis indicate that a prolonged pelagic larval duration in combination with the efficient mechanism of larval dispersal via currents ensure delivery of the homogenous mix of divergently adapted genotypes throughout NZ. At settlement phenotype – environment mismatches of non-optimal genotypes promote post-settlement mortality and the establishment of three genetically diverged regions. Variation in the peak settlement times throughout the year in combination with phenotypic plasticity allows for a low level of admixture of "mismatched" genotypes that ensure homogenisation of allele frequencies of the next generation of red rock lobsters. This system appears to have an effective mechanism of adjustment to the effects of changing seawater temperature. The levels of genetic diversity and adaptability are likely to facilitate the long-term sustainability of harvesting of the red rock lobster *Jasus edwardsii*.

Limitations and considerations of the outlier marker analysis

Majority of tests for detection of markers putatively under selection relies on estimates of global F_{ST} relatively to the total heterozygosity in order to identify markers exhibiting larger genetic distance than expected under the assumption of neutral distribution (Narum et al. 2011). Some of the most extensively used F_{ST} outlier tests based on making strict assumptions about the demographic history of the population and are devised to control for evolutionary nonindependence of samples (Lotterhos et al. 2015). However, this approach is prone to potential problems arising from the fact that true population history is rarely known with confidence and may be quite complex. Outlier detection tests outcome has been shown to vary due to sampling design (Mita et al. 2013) and demographic history of the population, especially in case of recent range expansion (Lotterhos et al. 2014). Thus, often a

combination of outlier detection software used in order to account for most common or most likely demographic scenarios when no prior knowledge is available (Ahrens et al. 2018).

In this study, in order to minimize the impact of sampling design extensive sampling have been performed throughout the NZ range of *Jasus edwardsii*. Additionally, some prior information about demographic history is available for the red rock lobster. No recent range contractions/extensions have been detected for the species via the annual catch and settlement monitoring performed by the red rock lobster fishery (NZ 2016; Forman et al. 2017). The weak pattern of isolation by distance has been detected for the NZ population of red rock lobster using microsatellite markers (Thomas et al. 2013). However, re-analysis of eight microsatellites used in this study revealed one of the markers to be linked to the genomic region under selection (**Appendix 1**) which is likely to contribute to some extent to the IBD pattern observed.

One of the most widely used approaches to outlier detection is based on coalescent simulations of the F_{ST} distribution and expected heterozygosity under the island model assumption (Beaumont et al. 1996) - FDIST2 method implemented in Lositan (Antao et al. 2008). The FDIST2 method has been characterised as having relatively low type II error rate (less false negatives) and type I error rate (less false positives) for divergent selection and similar power when compared to other available software for outlier detection (Narum et al. 2011). However, neutral parametrisation applied to this method under the IBD model has been shown to cause a higher incidence of false positives in the right tail of the F_{ST} distribution (Lotterhos et al. 2014). Therefore, neutral parametrisation of this test has not been employed in this study. The assumption of an island model may lead to a narrow range of F_{ST} distribution and hence produce more false positives if the demographic history of the population may be more appropriately described by the hierarchical model (Excoffier et al. 2009). Another commonly used piece of software – Arlequin is designed to account for hierarchical genetic structure and bottlenecks (Excoffier et al. 2010) and has been employed for this study alongside Lositan. An alternative to strict demographic history assumptions for outlier detection is a Bayesian method based on the multinomial-Dirichlet distribution (Beaumont 2005). This approach is capable of simulating gene frequencies under a wide range of demographic scenarios and accounts for unequal effective sizes and immigration rate between subpopulations (Foll et al. 2008). One of the main parameters of outlier detection in this approach is prior odds for a number of neutral:selective loci. Increase in the value of prior odds from the default value of 10 has been shown to substantially decrease the

occurrence of false positives without sacrifice of the discriminating power, therefore prior odds of 10 000 have been used in this study. Thus, three outlier detection methods based on different demographic models and mathematical approaches were utilized in order to provide robust statistical support for outlier marker characterisation. Nonetheless, the occurrence of false positives cannot be completely ruled out due to a potential violation of the neutrality assumption in the case of *Jasus edwardsii*. However, set of neutral markers obtained using Arlequin, Lositan and BayeScan in this study was able to detect population differentiation between Australia and New Zealand consistently with previously published research which indicates that marker sequestration is likely biologically meaningful (Morgan et al. 2013; Thomas et al. 2013; Villacorta-Rath et al. 2016).

Another important limitation of results of this study is the utilization of Bayesian inference with STRUCTURE software applied to outlier and neutral+outlier sets of markers as it may violate the model assumption of neutrality of STRUCTURE algorithm (Pritchard et al. 2003). On the other hand, analysis of markers under selective environmental pressure may provide an invaluable insight into the evolutionary or adaptively significant populations. A number of studies successfully employed STRUCTURE for analysis of non-neutral markers based on careful selection of outlier markers based on LD trends and deviations from HWE (Limborg et al. 2012; Bernardi et al. 2016; Gleason et al. 2016). In this study, all markers have been screened for deviations from HWE and very stringent LD filtering have been performed. However, it is important to acknowledge that relatively high LD cut off of 0.8 and retention of single SNP per loci are the only available methods of LD screening in species without the reference genome and may not guarantee complete independence of the markers. Therefore, assumptions about population differentiation from Bayesian clustering analysis have been made only if supported by other methods of inference (DAPC, F_{ST} tests).

Adaptive potential of red rock lobsters in the changing climate

Ecosystems of New Zealand's marine area have been estimated to contribute approximately US$357 billion worth of services per year with coastal marine ecosystems contributing a substantial share (MacDiarmid et al. 2013). However, anthropogenically induced climate change has a potential to reduce the ecological and socio-economic value of many coastal marine ecosystems (Smithson 2002). The red rock lobster (*Jasus edwardsii*) fishery represents one of the most valuable NZ fisheries as it generates NZ$230 million to the

economy annually. Red rock lobster also has high recreational and customary value, which makes it an iconic NZ species (NZ 2016). Lobsters also play essential role as part of rocky reef ecosystems as the keystone predators and the scavengers (Fielder 1965).

Rising sea temperatures is one of the central concerns in understanding the influence of a changing climate on distribution of species' and therefore has become the focus of many studies (Perry et al. 2005; Sorte et al. 2010; Chen et al. 2011). The distribution range of a species is to a large extent determined by its 'climate envelope' – physiological tolerance of the species to the ranges of environmental conditions (Walther et al. 2002). Increasing temperatures are predicted to cause species distribution range shifts that can include the expansion, contraction, or both, of a species' range (Sorte et al. 2010). A meta-analysis of the demographic responses to climate change suggests that the most likely direction of distribution range shifts for marine species' is poleward (Perry et al. 2005). For terrestrial species climate change-induced altitudinal range shifts are quite common, however very few marine species respond to rising sea temperatures via a change in depth (Burrows et al. 2011). Management and conservation of a species in the face of climate change relies heavily on both the prediction of distribution range shifts and the species' ability to adjust to the shift their 'climate envelopes'. Acclimation to the changing conditions is limited by the physiological tolerance threshold and once it is exceeded mobile marine organisms have the advantage of an immediate response to a 'climate envelope' shift as they can potentially move, while for sedentary and sessile species biological or evolutionary response times can lag behind fast local environmental changes (Brierley et al. 2009). Lobsters are considered to be primarily sedentary organisms, but immature red rock lobsters have been shown to migrate substantial distances: from eastern Foveaux Straight and Stewart Island along the southeast coast of Stewart Island to Fiordland. (McKoy 1983). Red rock lobsters appear to occupy all the physiologically suitable habitat available for the species within NZ waters however despite the ability to move their latitudinal range expansion seems to be constrained by the vast distances to the next suitable habitat. Despite the ability of red rock lobster larvae to drift with currents for up to two years, the distance and patterns of currents do not allow settlement and survival of pueruli beyond Auckland Islands in the south and Chatham Islands in the east (Annala 1983). In addition to distribution/abundance shifts, species responses to climate change include: a change in phenology and physiology, species interactions and adaptation (Parmesan et al. 2003). Phenological and physiological changes are beyond the scope of this work which is focused on adaptability of the red rock lobster in a changing climate. Results

of this study suggest that the red rock lobster appears to possess an efficient mechanism of larval homogenisation and distribution that provides sufficient genetic diversity for local selection. The system of currents around New Zealand in conjunction with a prolonged pelagic larval duration produces a mix of divergently adapted genotypes, which is evident from the low levels of differentiation found for neutral markers (Chapter 3). Once phyllosoma reach the continental shelf and molt into pueruli, the local environment reduces the chances of settlement of non-optimally adapted individuals and thus produces three types of SST amplitude-adapted genotypes, which each survive at a higher rate in their corresponding environments with an optimal annual amplitude of sea temperature (Chapter 3 and 5).

Levels of admixture of non-adapted genotypes, however low, in the northern-most region of NZ indicate that the current selective pressure of SST amplitude is not sufficient to eliminate the survival of less adapted genotypes entirely, which suggests that there is potential for the populations closest to the equator to withstand further increases in SST amplitude without contraction of the northernmost limit of the distribution range. Overall, the existence of a range of temperature adapted genotypes may indicate genomic resilience and adaptability of the red rock lobster to sea temperatures changes. However, the observed genetic variation of the outlier markers requires identification of the associated genes and their function to determine their adaptive significance. SST around New Zealand is projected to increase by ~0.8°C under greenhouse gas concentration trajectory Representative Concentration Pathways (RCP) 4.5 and ~1°C under RCP8.5, and ~1.1°C under RCP4.5 and ~2.5°C under RCP8.5 by the end of the century (Law et al. 2017). Further investigation of the adaptability of the species requires testing of the physiological tolerance limits of associated phenotypic traits and fitness advantages of allelic variants in regard to environmental selection.

Increasing sea temperatures is only one of the facets of anthropogenically induced climate change, changes in ocean acidity, as a result of increased CO_2 concentrations have had significant impacts on marine life (Harley et al. 2006). Crustaceans appear to be less sensitive to the effects of ocean acidification when compared to other marine taxa such as corals, molluscs and fish (Wittmann et al. 2013). The CO_2-driven ocean acidification appears to have a variable ontogenetic impact on marine life - a direct comparison of tolerance to higher water acidity of larval and adult stages revealed the higher sensitivity of larval stages of echinoderms, bivalves, corals and crustacean species (Kurihara 2008). Ocean acidification has been shown to affect the growth of two species of crabs whereby acidified water did not

change the morphology or the calcium content of juvenile crabs, but it stunted their growth suggesting that crabs are able to maintain the calcification rates at increased energetic cost (Long et al. 2013). Higher metabolic rates driven by an increase in water acidity may have a profound effect on red rock lobster pueruli. In contrast with phyllosoma, red rock lobster pueruli are non-feeding, and their energy resources permit them to actively swim for approximately 200 km (for both horizontal and vertical movements), which is essential in finding a suitable habitat for settlement (Jeffs et al. 2001). Actively feeding life stages of *Jasus edwardsii* are also likely to be affected by a changing climate – laboratory simulations of the development of stage I phyllosoma at higher water temperatures demonstrated the production of smaller individuals with reduced levels of energy reserves (Smith et al. 2002). Embryonic development of the red rock lobster appears to have a higher efficiency at lower water temperatures similar to other marine crustaceans (Laughlin et al. 1989; Walther et al. 2010) and so a higher seawater temperature is likely to lead to a higher pre-settlement mortality.

Laboratory simulations of different climate change scenarios have demonstrated that a decrease in pH has the potential to exacerbate the impact of an increase in water temperature and vice versa (Wittmann et al. 2013). Combined effects of increased temperature and ocean acidification has been shown to prolong larval development and reduce larval growth and fitness of spider crab *Hyas araneus* (Walther et al. 2010). Combined effects of higher temperature and acidity on the early life stages of red rock lobster may profoundly reduce growth rates that may ultimately lead to a decrease in survival, and alter their distribution and abundance (Whiteley 2011). In this study, I was not able to perform environmental association analysis of genetic divergence with pCO_2 levels due to the lack of information about this environmental variable. However, only 43 out of 127 markers under selection were correlated with SST amplitude, and further research into genomic regions under selection may reveal adaptive potential to ocean acidification of the red rock lobster.

Oceanic currents and eddies play a vital role in the connectivity of red rock lobster populations, as well as in the distribution of the mix of temperature-adapted genotypes. The global air and sea warming patterns have been associated with a global change of wind patterns and poleward shift of subtropical western boundary currents (Wu et al. 2012). This change to the wind stress patterns may cause the intensification the East Australia Current (EAC) in a southward direction (Cai et al. 2005). Strengthening of the EAC is likely to affect the warming of the Tasman Sea and impact the marine ecosystem through the southern mid-

latitude ocean (Cai 2006). Changes in the patterns of oceanic circulation may reduce the high levels of connectivity between red rock lobster populations in NZ and impede the transport of adapted genotypes to their optimal habitats.

Temporal variation of population differentiation patterns of the red rock lobster

Here I have investigated patterns of population differentiation of adult and pueruli/juvenile red rock lobster in order to estimate when in the life cycle selective pressure is exerted, and adaptive population divergence is formed. These two life stages have been sampled within a one year (2016). Determination of the age of adult lobsters sampled was beyond the scope of this study, and thus only leg samples were taken. Therefore, once the sampling was completed, it was impossible to age the samples as reliable age estimation in crustaceans requires gastric mill samples (Kilada et al. 2012). It is highly likely that adult lobster dataset is comprised of a range of different ages. However, all the adult lobster sampled were above the minimal legal size (MLS) - tail width of at least 54 mm for males and 60 mm for females. MLS restriction has been established in order to ensure that lobsters reach full maturity and can reproduce at least once before being harvested (MacDiarmid 1989b). Depending on the location, lobster can reach maturity in three to twelve years, so the time gap between the adult and pueruli/juvenile samples obtained for this study is at least three years. Therefore, comparison of these two life stages can also provide insight into the temporal stability of population differentiation patterns.

Analysis of molecular variance (AMOVA) allow to partition and estimate components of molecular variance (Excoffier et al. 1992). An AMOVA between years has been performed using SNPs discovered for a combined adult and pueruli/juvenile dataset. Four locations from the pueruli/juvenile dataset were compared to the corresponding sampling locations in the adult dataset. An AMOVA was performed using the R package *poppr* version 2.1.1 utilizing the ade4 implementation of AMOVA (Kamvar et al. 2017). Results are presented in **Table 6.1.** The AMOVA showed a low level of differentiation between the stages sampled (0.05%). A high degree of consistency of population differentiation patterns is further supported by the similarity of pairwise F_{ST} values between the stages (**Table 6.2**).

Table 6.1. Analysis of molecular variance (AMOVA) among corresponding sampling locations of adult and pueruli/juvenile red rock lobsters.

Source of variation	Variance	% of variation	p
Between stages	0.17669	0.05	0.0001
Within stage	343.7188	99.95	0.0001

Table 6.2. Pairwise F_{ST} between corresponding locations of pueruli/juvenile and adult datasets. Results are presented for neutral and outlier panels, separately. Values in bold are statistically significant at p<0.0001.

		Pueruli/juvenile					Adult			
		WHA	CAS	NAP	JAC		910	911	913	928
neutral	WHA					910				
	CAS	0.0008				911	0.0029			
	NAP	0.0017	0.0023			913	0.0023	0.0013		
	JAC	0.0003	0.0000	0.0005		928	0.0022	0.0004	0.0015	
		WHA	CAS	NAP	JAC		910	911	913	928
outlier	WHA					910				
	CAS	**0.008**				911	**0.004**			
	NAP	**0.047**	0.003			913	**0.014**	**0.02**		
	JAC	**0.097**	**0.094**	**0.109**		928	**0.066**	**0.065**	**0.079**	

Interestingly, for Australian populations of *Jasus edwardsii* patterns of chaotic genetic patchiness have been reported (Villacorta-Rath et al. 2018) that is in contrast to the results of the present study. Villacorta-Rath et al. (2018) compared two populations from Tasmania and South Australia over the period of four years (2009 - 2013) using puerulus (stage 1 - 2) and

postpuerulus samples. Here, due to the preservation of the puerulus in ethanol, I was not able to precisely identify the stage of development of the pueruli samples, although similar uncertainty has been reported by Villacorta-Rath et al. (2018). Therefore, it is impossible to compare the stages of development of these samples with a high degree of certainty. Variable patterns of population differentiation inferred via pairwise F_{ST} and DAPC analysis of neutral markers is further supported by AMOVA and analysis of allele frequencies between SA and Tasmania (Villacorta-Rath et al. 2018). Pairwise F_{ST} indices calculated for four NZ sampling sites using neutral markers were not significant and varied within similar ranges: 0.0000 – 0.0023 for pueruli/juveniles and 0.0004 – 0.0029 for adults.

A single SNP putatively under divergent selection was identified in common for both SA and Tasmania sampling sites (Villacorta-Rath et al. 2018). In contrast, 127 SNPs were characterised as being under selection in both pueruli/juvenile and adult datasets in my study. The disparity in the number of outlier SNPs may originate from the difference in strength and/or consistency of environmental selection between sites in Australia and NZ. However, different sets of outlier calling software were used for SNP characterisation: OUTflank and Lositan were employed by Villacorta-Rath et al. (2018), while I used Arlequin, BayeScan and Lositan for outlier detection. The difference in the approaches employed by the outlier detection methods might at least in part, explain the difference in the number of SNPs characterised.

Selective pressure acting on the locations sampled in Australia appears to be relatively weak and inconsistent when compared to environmental selection on NZ populations of red rock lobsters. Thus, neutral micro-evolutionary processes and pre-settlement stochastic effects on cohorts of migrants play a more substantial role in the formation of patterns of population differentiation in Australia than in NZ. However, more extensive analysis of temporal variation of population differentiation patterns in NZ is needed to confirm this hypothesis.

Implication for red rock lobster fishery management

The sustainable management of fished resources is difficult without knowledge of the population structure and connectivity that provide important information about dispersal, recruitment and stock identification for managerial decisions. Many fisheries rely on assessments of factors underlying short-term demographic changes ('quantitative' change), however, the use of molecular genetic methods to investigate the processes forming

population structure ('qualitative' change) influence both short-term alterations in phenotypic traits and longer-term responses to natural and anthropogenic perturbations (Hauser et al. 2008). Decoupling of the population structure from management units has led to reduced productivity and a decrease in the size of local populations for many fished species (Kough et al. 2013; Knutsen et al. 2014). Stock identification is one of central questions in fisheries management as it allows an estimate of the sustainable harvesting levels that will not compromise stock productivity (Botsford 1997). The term stock has been defined as a "reproductively isolated unit that is genetically distinct" (Kochzius 2009). The analysis of neutral micro-evolutionary processes (gene flow, genetic drift) performed in this study demonstrates the absence of genetically distinct and reproductively isolated populations within the NZ population of red rock lobster, thus supporting the current management regime that has been largely based on the assumption of a panmixia. However, sustainable management of the fishery requires the incorporation of all biologically relevant processes acting on the fished populations. Here, I identified an adaptive genetic response to the selective force (annual amplitude of SST) that produced three distinct temperature-adapted genotypes: higher SST amplitude adapted genotypes (mostly distributed in the NW region), lower SST amplitude adapted genotypes (mostly distributed in the S region), and the NE region that is mostly populated by the genotypes adapted to the intermediate conditions. The adaptively divergent populations identified in this thesis appear to coincide to some extent with the existing management units – CRA areas. CRA1 and CRA9 are mostly comprised of individuals with NW genotype although samples obtained from statistical area 904 (belonging to CRA1) are genetically more similar to the NE genotypes. Environmental selection in CRA areas 2, 3, 4, 5, and 6 favours settlement of NE genotypes. S region includes CRA7 and CRA8, although the apparent exact match is most likely produced by the lack samples from statistical areas immediately adjacent to CRA7 and CRA8. Temporal stability of patterns of adaptive divergence is crucial in the context of management (Bonin et al. 2007) and based on the comparison of the population genetic structure of adult and pueruli/juvenile lobsters, which represent separate time points, the observed adaptively diverged regions appear to be relatively stable over time. However, the mechanism of establishment of adaptive population structure suggests that it is highly responsive to variable environmental selection. Selective mortality of the non-optimally adapted genotypes at the time of settlement can ensure development of the divergence patterns within single generation (Hauser et al. 2008). In *Jasus edwardsii*, settlement occurs year-round with the most common settlement peaks in Austral summer and winter thus creating potential for

continuous genotypic adjustment of the population structure. Therefore, the minor mismatches between adaptively divergent regions and management areas are not likely to have any negative effects on the sustainability of the harvesting. Additionally, levels of heterozygosity and inbreeding calculated for each sampling location and for broader regions indicate high genetic diversity of the harvested populations of red rock lobster. Overall, the contemporary level of genetic diversity supports the established management units and harvesting regime.

However, concerns of long-term sustainability arise in the context of climate change, as the impact of changing environment combined with the effects of fishing pressure may lead to depletion of stocks and collapse of the fishery (Harley et al. 2006). Furthermore, overexploited fish populations can undergo a loss of genetic diversity (Hauser et al. 2002), and with it the ability for the species to adapt and evolve to new environmental circumstances (Perry et al. 2010). Preservation of the genetic diversity of the fished stocks is recognised to be one of the central concerns for long-term sustainability (Conover et al. 2002; Bonin et al. 2007; Reiss et al. 2009). Therefore, monitoring of the patterns of adaptive variation described in this study would be highly desirable in order to maintain the resilience of NZ stock of red rock lobster to environmental change.

Future directions

Based on the results obtained in this study, the monitoring and maintenance of adaptive genetic diversity are crucial for conservation and sustainable harvesting of the red rock lobster long-term. Three distinct temperature-adapted genotypes described here have corresponding "climate envelopes", and genotype range shifts are a likely mode of response to sea temperature rise. NW genotypes are likely to be more negatively impacted by the increase in SST than NE and S genotypes as they are closer to limits of adaptive thermal tolerance (Thomas et al. 2000). Latitudinal range expansion of the NW genotypes, which are the most vulnerable to climate change as closest to the equator, may indicate a potential contraction of the northernmost boundary of the *Jasus edwardsii* distribution range and increase the severity of the effects of climate change on lobster physiology. Early detection of possible contraction of the distribution range is especially important because the geographical distribution of red rock lobsters does not allow for latitudinal expansion of the species range. This information might be invaluable for the timely adjustment of the rates of harvesting of

the adapted genotype affected by the distribution range contraction in order to preserve the levels of adaptive variation in the *Jasus edwardsii* population.

Further investigation into the candidate genes for SST amplitude and associated phenotypic traits could provide invaluable information about the local and NZ-wide selective pressures and adaptive mechanisms for the red rock lobster. Incorporation of data concerning the levels of ocean acidification into environmental analyses might provide a more informative insight into the impacts of climate change on the genetic divergence of *Jasus edwardsii* the species' susceptibility to a change to the environmental conditions.

Early life stages of the red rock lobster appear to be more vulnerable to the effects of climate change than adult lobsters and so they might become a survival bottleneck in the life cycle. Laboratory experiments testing the tolerance of larval stages of *Jasus edwardsii* to changes in sea temperature and acidity under different climate change scenarios may provide an indication of the physiological tolerance thresholds. Furthermore, these results could inform conservation and management bodies as to which environmental conditions would result in larval connectivity between red rock lobster populations being reduced to a critical level that threatens the demographic and genetic connectivity between populations.

Additionally, monitoring of the allele frequencies of the candidate genes over time will allow for an estimate of the levels of DNA change that may reflect the adaptability of the species to the specific types of environmental selection. The decay of a particular candidate gene associated with a physiological trait will indicate evolutionary constraints of that trait.

Finally, in the present study conclusions on the relative consistency of a selective pressure on the NZ population of red rock lobsters and their adaptive response are drawn from the similarity of neutral and adaptive population structure of two life stages that represent two separate time points: adult lobster above minimal legal size and late pueruli/early juvenile lobsters. Monitoring of the population differentiation patterns over successive years will produce a more confident assessment of the levels of temporal variability of the selection and adaptive response.

Another limitation to the conclusions drawn in this study is that population genetics analysis of pueruli/juveniles were conducted only on individuals settling in wintertime. Low winter sea temperature is likely to have negatively affected the survival of NE genotypes in the S regions which can be seen from higher outlier pairwise F_{ST} values in pueruli/juvenile dataset when compared to levels of differentiation between corresponding locations in the adult

dataset (**Table 6.2**). Red rock lobster settlement has two main settlement peaks (Austral winter and summer) with lower levels of settlement during the rest of the year (Booth et al. 1986). Settling in summer sea water temperatures may allow for higher survival rate of NE genotypes in S region during the initial post-settlement period. However, differentiation levels of adult populations suggest that NE genotypes might experience higher mortality rates during the first winter after settlement. Phenotypic plasticity of some of the summer settlers may allow them to survive into maturity and reproduce, however, investigation of population differentiation patterns of pueruli/juveniles settling in summer is needed to prove this hypothesis.

Concluding remarks

In conclusion, here I described the current population genetics structure of the red rock lobster *Jasus edwardsii* using selectively neutral and putatively under divergent selection SNP markers. Screening for correlations between a range of environmental variables and the genetic diversity identified annual amplitude of SST to be the selective agent most likely driving adaptive divergence of three broad regions in NZ: NW, NE and S. The selective pressure of SST amplitude is not sufficient to entirely eliminate settlement and survival into maturity of genotypes adapted to alternative environments. Therefore, breeding can occur between non-optimally and optimally adapted genotypes within an environment causing homogenisation of allele frequencies. High levels of connectivity and gene flow between the populations effected by the system of currents and eddies, in combination with an approximately two year-long PLD, provides highly efficient mixing of larval genotypes and results in a delivery of genetically diverged settlers throughout NZ. A mismatch between phenotypic fitness of divergently adapted genotypes and environmental conditions promotes selective mortality and increases recruitment of optimally adapted genotypes. The combination of the selective pressure and high levels of gene flow creates a system capable of fast adjustment to the changing sea temperature.

References

Aguirre JD, et al. 2012. Genetic diversity increases population productivity in a sessile marine invertebrate. Ecology 93:1134-1142.

Ahrens CW, et al. 2018. The search for loci under selection: trends, biases and progress. Mol Ecol 27:1342-1356.

Allendorf FW, et al. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697-709.

Almany GR, et al. 2007. Local replenishment of coral reef fish populations in a marine reserve. Science 316:742-744.

Altschul SF, et al. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Amaral AR, et al. 2012. Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (Genus Delphinus). PLoS One 7:e31482.

Andersen Ø, et al. 2015. Evolutionary history and adaptive significance of the polymorphic Pan I in migratory and stationary populations of Atlantic cod (Gadus morhua). Marine Genomics 22:45-54.

Anderson EC, et al. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567-2582.

Anderson M, et al. 2008. Permanova+ for Primer: Guide to Software and Statisticl Methods: Primer-E Limited.

Anderson SC, et al. 2011. Rapid global expansion of invertebrate fisheries: trends, drivers, and ecosystem effects. PLoS One 6:e14735.

Andolfatto P. 2001. Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev 11:635-641.

Andrews KR, et al. 2014. Recent novel approaches for population genomics data analysis. Mol Ecol 23:1661-1667.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Annala. 1983. New Zealand Rock Lobsters: Biology and Fishery.

Annala JH. 1996. New Zealand's ITQ system: have the first eight years been a success or a failure? Reviews in Fish Biology and Fisheries 6:43-62.

Annala JH, et al. 1987. Fecundity of the New Zealand red rock lobster,Jasus edwardsii. New Zealand Journal of Marine and Freshwater Research 21:591-597.

Annala JH, et al. 1988. Growth of rock lobsters (Jasus edwardsii)in Fiordland, New Zealand. New Zealand Journal of Marine and Freshwater Research 22:29-41.

Annala JH, et al. 1980. Size at the onset of sexual maturity in female Jasus edwardsii (Decapoda: Palinuridae) in New Zealand. New Zealand Journal of Marine and Freshwater Research 14:217-227.

Antao T, et al. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323.

Araneda C, et al. 2016. Adaptive genetic variation distinguishes Chilean blue mussels (Mytilus chilensis) from different marine environments. Ecology and Evolution 6:3632-3644.

Auld JR, et al. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proceedings of the Royal Society B: Biological Sciences 277:503-511.

Azumi K, et al. 2007. Gene expression profile during the life cycle of the urochordate Ciona intestinalis. Developmental Biology 308:572-582.

Baird NA, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376.

Bairoch A, et al. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research 28:45-48.

Barnes RSK, et al. 1999. An introduction to marine ecology: John Wiley & Sons.

Barth JM, et al. 2017. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. Mol Ecol.

Barton N. 2005. Fitness landscapes and the origin of species. In: BioOne.

Barton NH. 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society of London B: Biological Sciences 355:1553-1562.

Batta-Lona PG, et al. 2011. Population genetic variation of the Southern Ocean krill, Euphausia superba, in the Western Antarctic Peninsula region based on mitochondrial single nucleotide polymorphisms (SNPs). Deep Sea Research Part II: Topical Studies in Oceanography 58:1652-1661.

Baum JK, et al. 2003. Collapse and conservation of shark populations in the Northwest Atlantic. Science 299:389-392.

Baxter SW, et al. 2011. Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism. PLoS One 6:e19315.

Beaumont MA. 2005. Adaptation and speciation: what can F_{ST} tell us? Trends Ecol Evol 20:435-440.

Beaumont MA, et al. 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. Proceedings of the Royal Society of London. Series B: Biological Sciences 263:1619-1626.

Becker BJ, et al. 2007. Complex larval connectivity patterns among marine invertebrate populations. Proceedings of the National Academy of Sciences 104:3267-3272.

Benestan L, et al. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). Mol Ecol 24:3299-3315.

Benestan L, et al. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). Mol Ecol 25:5073-5092.

Bentley DR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53.

Berg PR, et al. 2015. Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (Gadus morhua L.). Genome Biology and Evolution 7:1644-1663.

Bermudes M, et al. 2008. Response of early stage spiny lobster Jasus edwardsii phyllosoma larvae to changes in temperature and photoperiod. Aquaculture 281:63-69.

Bernardi G, et al. 2016. Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. Mol Ecol 25:3384-3396.

Berumen ML, et al. 2012. Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. Ecology and Evolution 2:444-452.

Bierne N. 2010. The distinctive footprint of local hitchhiking ina varied environment and global hitchhiking in a subdivided population. Evolution 64:3254-3272.

Bock KW. 2016. The UDP-glycosyltransferase (UGT) superfamily expressed in humans, insects and plants: Animal vs. plant arms-race and co-evolution. Biochemical Pharmacology 99:11-17.

Bolger AM, et al. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bonin A, et al. 2004. How to track and assess genotyping errors in population genetics studies. Mol Ecol 13:3261-3273.

Bonin A, et al. 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. Conservation Biology 21:697-708.

Booth J, et al. 2000. New Zealand's rock lobster fisheries. Spiny lobsters: fisheries and culture:78-89.

Booth J, et al. 1992. Distribution of phyllosoma larvae of the red rock lobster Jasus edwardsii off the east coast of New Zealand in relation to the oceanography. Larval biology. Proceedings 15:138-148.

Booth JD. 1979. Settlement of the rock lobster,Jasus edwardsii (Decapoda: Palinuridae), at Castlepoint, New Zealand. New Zealand Journal of Marine and Freshwater Research 13:395-406.

Booth JD. 1997. Long-distance movements in Jasus spp. and their role in larval recruitment. Bulletin of Marine Science 61:111-128.

Booth JD, et al. 1994. Early life history of spiny lobster. Crustaceana:271-294.

Booth JD, et al. 1990. Systematic status of the rock lobsters Jasus edwardsii from New Zealand and J. novaehollandiae from Australia. New Zealand Journal of Marine and Freshwater Research 24:239-249.

Booth JD, et al. 1986. Settlement of the red rock lobster,Jasus edwardsii, near Gisborne, New Zealand. New Zealand Journal of Marine and Freshwater Research 20:291-297.

Borcard D, et al. 2012. Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. Ecology 93:1473-1481.

Borcard D, et al. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826-1832.

Bossart J, et al. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. Trends Ecol Evol 13:202-206.

Botsford LW. 1997. The Management of Fisheries and Marine Ecosystems. Science 277:509-515.

Bradbury IR, et al. 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proc Biol Sci 277:3725-3734.

Bridle JR, et al. 2007. Limits to evolution at range margins: when and why does adaptation fail? Trends Ecol Evol 22:140-147.

Brierley AS, et al. 2009. Impacts of Climate Change on Marine Organisms and Ecosystems. Current Biology 19:R602-R614.

Bruford MW, et al. 1993. Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939-943.

Buerkle C, et al. 2013. Population genomics based on low coverage sequencing: how low should we go? Mol Ecol 22:3028-3035.

Burchell B, et al. 1997. The Structure and Function of the UDP-Glucuronosyltransferase Gene Family. In: Goldstein DS, Eisenhofer G, McCarty R, editors. Advances in Pharmacology: Academic Press. p. 335-338.

Burgess SC, et al. 2014. Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. Ecological Applications 24:257-270.

Burrows MT, et al. 2011. The Pace of Shifting Climate in Marine and Terrestrial Ecosystems. Science 334:652-655.

Butler M. 1999. Sperm economy and limitations in spiny lobsters.

Butler M, et al. 2000. Puerulus and juvenile ecology. Spiny Lobsters: Fisheries and Culture, Second Edition:276-301.

Cai W. 2006. Antarctic ozone depletion causes an intensification of the Southern Ocean super-gyre circulation. Geophysical Research Letters 33.

Cai W, et al. 2005. The response of the Southern Annular Mode, the East Australian Current, and the southern mid-latitude ocean circulation to global warming. Geophysical Research Letters 32.

Callicrate T, et al. (Callicrate2014 co-authors). 2014. Genomic resources for the endangered Hawaiian honeycreepers. BMC Genomics 15:1098.

Candy JR, et al. 2015. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (Thaleichthys pacificus, Osmeridae), an anadromous Pacific smelt. Mol Ecol Resour 15:1421-1434.

Castoe TA, et al. 2010. Rapid identification of thousands of copperhead snake (Agkistrodon contortrix) microsatellite loci from modest amounts of 454 shotgun genome sequence. Mol Ecol Resour 10:341-347.

Catchen J, et al. 2013. Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124-3140.

Catchen JM, et al. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda) 1:171-182.

Charlesworth B, et al. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetics Research 70:155-174.

Charlesworth D. 2006. Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. PLoS Genet 2:e64.

Chen I-C, et al. 2011. Rapid Range Shifts of Species Associated with High Levels of Climate Warming. Science 333:1024-1026.

Chistiakov DA, et al. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255:1-29.

Chiswell SM, et al. 2005. Distribution of mid- and late-stageJasus edwardsiiphyllosomas: Implications for larval recruitment processes. New Zealand Journal of Marine and Freshwater Research 39:1157-1170.

Chiswell SM, et al. 2008. Sources and sinks of larval settlement in Jasus edwardsii around New Zealand: Where do larvae come from and where do they go? Marine Ecology Progress Series 354:201-217.

Chiswell SM, et al. 2011. Larval connectivity of harbours via ocean currents: A New Zealand study. Continental Shelf Research 31:1057-1074.

Chiswell SM, et al. 2010. The East Cape Current and two eddies: A mechanism for larval retention? New Zealand Journal of Marine and Freshwater Research 32:385-397.

Chong Z, et al. 2012. Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics 28:2732-2737.

Church DM, et al. 2009. Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. PLoS Biol 7:e1000112.

Cimmaruta R, et al. 2005. Genetic structure and environmental heterogeneity in the European hake (Merluccius merluccius). Mol Ecol 14:2577-2591.

Clarke DJ, et al. 1992. Multiplicity of UDP-glucuronosyltransferases in fish. Purification and characterization of a phenol UDP-glucuronosyltransferase from the liver of a marine teleost, Pleuronectes platessa. Biochemical Journal 284:417.

Clarke K, et al. 1993. A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series:205-219.

Clarke K, et al. 2006. Primer. Primer-E, Plymouth.

Cobb JS. 1997. Oceanic processes affecting lobster larvae: report from a workshop. Marine and Freshwater Research 48:771-775.

Conover DO, et al. 2006. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. Journal of Fish Biology 69:21-47.

Conover DO, et al. 2002. Sustaining Fisheries Yields Over Evolutionary Time Scales. Science 297:94-96.

Constable HB. 2014. Population structure, temporal stability and seascape genetics of two endemic New Zealand Pleuronectiformes, Rhombosolea plebeia (sand flounder) and R. leporina (yellowbelly flounder).

Conway T, et al. 2012. Gossamer--a resource-efficient de novo assembler. Bioinformatics 28:1937-1938.

Coscia I, et al. 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.

Crear B, et al. 2000. Growth of juvenile southern rock lobsters,< i> Jasus edwardsii</i>, is influenced by diet and temperature, whilst survival is influenced by diet and tank environment. Aquaculture 190:169-182.

Cruickshank TE, et al. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133-3157.

Cure K, et al. 2017. Genomic signatures of local adaptation reveal source-sink dynamics in a high gene flow fish species. Scientific Reports 7:8618.

D'Aloia CC, et al. 2015. Patterns, causes, and consequences of marine larval dispersal. Proceedings of the National Academy of Sciences 112:13940-13945.

DaCosta JM, et al. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One 9:e106713.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Davey JW, et al. 2011a. RADSeq: next-generation population genetics. Briefings in Functional Genomics 9:416-423.

Davey JW, et al. 2013. Special features of RAD Sequencing data: implications for genotyping. Mol Ecol 22:3151-3164.

Davey JW, et al. 2011b. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499-510.

DeAngelis MM, et al. 1995. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Research 23:4742.

DeFaveri J, et al. 2013. Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. Evolution 67:2530-2546.

Deiana A, et al. 1999. Genome size and AT-DNA content in thirteen species of Decapoda. Crustaceans and the biodiversity crisis. Edited by FR Schram and JC von Vaupel Klein. Koninklijke Brill NV, Leiden, the Netherlands:981-985.

Depree C, et al. 2008. Polycyclic aromatic hydrocarbons in Auckland's aquatic environment: sources, concentrations and potential environmental risks: Auckland Regional Council.

DeWitt T, et al. 2004. Integrated solutions to environmental heterogeneity: theory of multimoment reaction norms. Phenotypic Plasticity: Functional and Conceptual Approaches:98-111.

DeWitt TJ, et al. 1998. Costs and limits of phenotypic plasticity. Trends Ecol Evol 13:77-81.

Dieringer D, et al. 2003. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res 13:2242-2251.

Dou J, et al. (Dou2012 co-authors). 2012. Reference-free SNP calling: improved accuracy by preventing incorrect calls from repetitive genomic regions. Biology Direct 7:17.

Dray S, et al. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). ecological modelling 196:483-493.

Dubber GG, et al. 2004. The effects of temperature and diet on the survival, growth and food uptake of aquarium-held postpueruli of the rock lobster Jasus lalandii. Aquaculture 240:249-266.

e-Costa BH, et al. 2014. Tropicalization of fish assemblages in temperate biogeographic transition zones. Marine Ecology Progress Series 504:241-252.

Eaton DA. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844-1849.

Eckert CG, et al. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol Ecol 17:1170-1188.

Ekblom R, et al. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity (Edinb) 107:1-15.

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

Elshire RJ, et al. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379.

Etter PD, et al. 2012. SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. 772:157-178.

Evanno G, et al. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620.

Everett MV, et al. 2011. Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Mol Ecol Resour 11 Suppl 1:93-108.

Excoffier L, et al. 2009. Detecting loci under selection in a hierarchically structured population. Heredity (Edinb) 103:285.

Excoffier L, et al. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567.

Excoffier L, et al. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Fevolden SE, et al. 2012. Settling-depth vs. genotype and size vs. genotype correlations at the Pan I locus in 0-group Atlantic cod Gadus morhuaÃ‚Â Marine Ecology Progress Series 468:267-278.

Fielder D. 1965. A dominance order for shelter in the spiny lobster Jasus lalandei (H. Milne-Edwards). Behaviour 24:236-245.

Fitzpatrick SW, et al. 2016. Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. Evolutionary Applications 9:879-891.

Fogarty MJ, et al. 2007. Population connectivity and spatial management of marine fisheries. Oceanography 20:112-123.

Foll M, et al. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977-993.

Forman J, et al. 2017. Settlement indices for 2015/16 fishing year for the red rock lobster (Jasus edwardsii). Wellington, N.Z. :: Ministry of Primary Industries.

Freamo H, et al. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Mol Ecol Resour 11 Suppl 1:254-267.

Frichot E, et al. 2013. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Mol Biol Evol 30:1687-1699.

Gagnaire P-A, et al. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. Evolutionary Applications 8:769-786.

Galindo HM, et al. 2006. Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. Curr Biol 16:1622-1626.

Galindo HM, et al. 2010. Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. Mol Ecol 19:3692-3707.

Galindo J, et al. 2009. Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail Littorina saxatilis. Mol Ecol 18:919-930.

Gautier M, et al. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol 22:3165-3178.

Gavrilets S. 2004. Fitness landscapes and the origin of species (MPB-41): Princeton, NJ: Princeton university press.

Gebremedhin B, et al. 2009. Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation. Animal Conservation 12:107-109.

George S, et al. 2002. Allelic variations in the plaice UGT1B1 gene. Marine Environmental Research 54:259-262.

George SG, et al. 1998. Structural Studies of a UDP-glucuronosyltransferase gene from the plaice (Pleuronectes platessa). Marine Environmental Research 46:33-35.

Ghalambor CK, et al. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394-407.

Gleason LU, et al. 2016. Genomic evidence for ecological divergence against a background of population homogeneity in the marine snail Chlorostoma funebralis. Mol Ecol 25:3557-3573.

Glover KA, et al. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genet 11:2.

Green SB, et al. 2010. Using SPSS for Windows and Macintosh: Analyzing and understanding data: Prentice Hall Press.

Grewal SIS, et al. 2003. Heterochromatin and Epigenetic Control of Gene Expression. Science 301:798-802.

Groeneveld JC, et al. 2012. High connectivity and lack of mtDNA differentiation among two previously recognized spiny lobster species in the southern Atlantic and Indian Oceans. Marine Biology Research 8:764-770.

Groth DM, et al. 2009. Characterization of polymorphic microsatellite loci in the western rock lobster (Panulirus cygnus). Conservation Genetics Resources 1:163-166.

Guinand B, et al. 2015. From the laboratory to the wild: salinity-based genetic differentiation of the European sea bass (Dicentrarchus labrax) using gene-associated and gene-independent microsatellite markers. Marine Biology 162:515-538.

Hannan DA. 2014. Population genetics and connectivity in Paphies subtriangulata and Paphies australis (Bivalvia: Mesodesmatidae).

Hare MP, et al. 2005. Nonrandom larval dispersal can steepen marine clines. Evolution 59:2509-2517.

Harley CD, et al. 2006. The impacts of climate change in coastal marine systems. Ecol Lett 9:228-241.

Harrison N, et al. 2011. Next-generation sequencing and systematics: What can a billion base pairs of DNA sequence data do for you? Taxon 60:1552-1566.

Hauser L, et al. 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). Proceedings of the National Academy of Sciences 99:11742-11747.

Hauser L, et al. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9:333-362.

Hedrick PW. 2006. Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. Annual Review of Ecology, Evolution, and Systematics 37:67-93.

Hellberg ME, et al. 2002. Genetic assessment of connectivity among marine populations. Bulletin of Marine Science 70:273-290.

Helyar SJ, et al. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol Ecol Resour 11 Suppl 1:123-136.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. The American Naturalist 173:579-588.

Hess JE, et al. 2013. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Mol Ecol 22:2898-2916.

Hess JE, et al. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Mol Ecol Resour 11 Suppl 1:137-149.

Hilborn R, et al. 2003. Biocomplexity and fisheries sustainability. Proceedings of the National Academy of Sciences 100:6564-6568.

Hohenlohe PA, et al. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour 11 Suppl 1:117-122.

Hohenlohe PA, et al. 2012. Population Genomic Analysis of Model and Nonmodel Organisms Using Sequenced RAD Tags. In: Pompanon F, Bonin A, editors. Data Production and Analysis in Population Genomics: Methods and Protocols. Totowa, NJ: Humana Press. p. 235-260.

Holderegger R, et al. 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecology 21:797-807.

Hu F, et al. 2015. Bioaccumulation and detoxification responses in the scallop Chlamys farreri exposed to tetrabromobisphenol A (TBBPA). Environmental Toxicology and Pharmacology 39:997-1007.

Hudson ME. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol Ecol Resour 8:3-17.

Hutchinson WF. 2008. The dangers of ignoring stock complexity in fishery management: the case of the North Sea cod. Biology Letters 4:693-695.

Iacchei M, et al. 2013. Combined analyses of kinship and FST suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. Mol Ecol 22:3476-3494.

Iyanagi T. 1991. Molecular basis of multiple UDP-glucuronosyltransferase isoenzyme deficiencies in the hyperbilirubinemic rat (Gunn rat). Journal of Biological Chemistry 266:24048-24052.

Iyanagi T. 2007. Molecular Mechanism of Phase I and Phase II Drug-Metabolizing Enzymes: Implications for Detoxification. In. International Review of Cytology: Academic Press. p. 35-112.

Jackson DA. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology 74:2204-2214.

Jacobs MW, et al. 2006. The effect of larval age on morphology and gene expression during ascidian metamorphosis. Integrative and Comparative Biology 46:760-776.

Jeffs AG, et al. 2001. Distribution and condition of pueruli of the spiny lobster Jasus edwardsii offshore from north-east New Zealand. Marine and Freshwater Research 52:1211-1216.

Jeffs AG, et al. 2005. How do spiny lobster post-larvae find the coast? New Zealand Journal of Marine and Freshwater Research 39:605-617.

Jeffs AG, et al. 2004. Identifying potential prey of the pelagic larvae of the spiny lobster Jasus edwardsii using signature lipids. Comp Biochem Physiol B Biochem Mol Biol 137:487-507.

Jiang Z, et al. 2016. Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. International journal of biological sciences 12:100.

Johannesson K, et al. 2006. Invited review: Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol Ecol 15:2013-2029.

Jombart T, et al. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070-3071.

Jombart T, et al. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13:484.

Jorde PE, et al. 2015. Genetically distinct populations of northern shrimp, Pandalus borealis, in the North Atlantic: adaptation to different temperatures as an isolation factor. Mol Ecol 24:1742-1757.

Jump AS, et al. 2009. Environmental change and the option value of genetic diversity. Trends in Plant Science 14:51-58.

Jump AS, et al. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. Ecol Lett 8:1010-1020.

Kamvar ZN, et al. 2017. Developing educational resources for population genetics in R: an open and collaborative approach. Mol Ecol Resour 17:120-128.

Kao D, et al. 2016. The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. eLife 5:e20062.

Karlin S, et al. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proceedings of the National Academy of Sciences 87:2264-2268.

Keenan K, et al. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4:782-788.

Kennington WJ, et al. 2012. Evaluation of population genetic structure in the western rock lobster: University of Western Australia.

Kensler CB. 1967. The distribution of spiny lobsters in New Zealand waters (Crustacea: Decapoda: Palinuridae). New Zealand Journal of Marine and Freshwater Research 1:412-420.

Kilada R, et al. 2012. Direct determination of age in shrimps, crabs, and lobsters. Canadian Journal of Fisheries and Aquatic Sciences 69:1728-1733.

Kim Y, et al. 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. Genetics 160:765-777.

Kirk H, et al. 2011. Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. International Journal of Molecular Sciences 12:3966-3988.

Kirkpatrick M, et al. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419-434.

Kittaka J, et al. 2005. Development of the red rock lobster, Jasus edwardsii, from egg to juvenile. New Zealand Journal of Marine and Freshwater Research 39:263-277.

Knutsen H, et al. 2014. Does population genetic structure support present management regulations of the northern shrimp (Pandalus borealis) in Skagerrak and the North Sea? ICES Journal of Marine Science 72:863-871.

Koboldt DC, et al. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25:2283-2285.

Kochzius M. 2009. Trends in Fishery Genetics. In: Beamish RJ, Rothschild BJ, editors. The Future of Fisheries Science in North America. Dordrecht: Springer Netherlands. p. 453-493.

Kough AS, et al. 2013. Larval Connectivity and the International Management of Fisheries. PLoS One 8:e64970.

Kritzer JP, et al. 2010. Marine metapopulations: Academic Press.

Krück NC, et al. 2013. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (Mugil cephalus). Mol Ecol Resour 13:715-725.

Kumar S, et al. 2012. SNP Discovery through Next-Generation Sequencing and Its Applications. Int J Plant Genomics 2012:831460.

Kurihara H. 2008. Effects of CO_2-driven ocean acidification on the early developmental stages of invertebrates. Marine Ecology Progress Series 373:275-284.

Lamichhaney S, et al. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. Proc Natl Acad Sci U S A 109:19345-19350.

Last PR, et al. 2011. Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. Global Ecology and Biogeography 20:58-72.

Laughlin R, et al. 1989. Interactions between temperature and salinity during brooding on subsequent zoeal development of the mud crab Rhithropanopeus harrisii. Marine Biology 102:377-386.

Law CS, et al. 2017. Climate change projections for the surface ocean around New Zealand. New Zealand Journal of Marine and Freshwater Research:1-27.

Leathwick J, et al. 2009. Benthic-optimised marine environment classification for New Zealand waters. New Zealand Ministry of Fisheries Final Research Report, BEN2006–01, Wellington.

Leaver MJ, et al. 2007. Piscine UDP-glucuronosyltransferase 1B. Aquatic Toxicology 84:356-365.

Legendre P, et al. 2012. Chapter 11 - Canonical analysis. In: Legendre P, Legendre L, editors. Developments in Environmental Modelling: Elsevier. p. 625-710.

Lepais O, et al. 2014. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour 14:1314-1321.

Levin LA. 2006. Recent progress in understanding larval dispersal: new directions and digressions. Integrative and Comparative Biology 46:282-297.

Li H, et al. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589-595.

Liggins L, et al. 2013. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. Geography Compass 7:173-196.

Limborg MT, et al. 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus). Mol Ecol 21:3686-3703.

Lipcius R, et al. 2000. Introduction: ecology and fishery biology of spiny lobsters. Spiny Lobsters: Fisheries and Culture, Second Edition:1-41.

Lipcius RN. 1985. Size-dependent reproduction and molting in spiny lobsters and other long-lived decapods. Crustacean issues 3:129-148.

Livingstone DR. 1998. The fate of organic xenobiotics in aquatic ecosystems: quantitative and qualitative differences in biotransformation by invertebrates and fish. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 120:43-49.

Lock K, et al. 2007. New Zealand's quota management system: a history of the first 20 years.

Loman NJ, et al. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nature biotechnology 30:434.

Long WC, et al. 2013. Effects of Ocean Acidification on Juvenile Red King Crab (Paralithodes camtschaticus) and Tanner Crab (Chionoecetes bairdi) Growth, Condition, Calcification, and Survival. PLoS One 8:e60959.

Lotterhos KE, et al. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol 23:2178-2192.

Lotterhos KE, et al. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol 24:1031-1046.

Lowe WH, et al. 2010. What can genetics tell us about population connectivity? Mol Ecol 19:3038-3051.

Lowe WH, et al. 2017. Population Genetics and Demography Unite Ecology and Evolution. Trends Ecol Evol 32:141-152.

Luikart G, et al. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981-994.

Lynch M. 2009. Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics 182:295-301.

MacDiarmid A. 1991. Seasonal changes in depth distribution, sex ratio and size frequency of spiny lobster Jasus edwardsii on a coastal reef in northern New Zealand. Marine ecology progress series. Oldendorf 70:129-141.

MacDiarmid AB. 1989a. Moulting and reproduction of the spiny lobster Jasus edwardsii (Decapoda: Palinuridae) in northern New Zealand.

MacDiarmid AB. 1989b. Size at onset of maturity and size-dependent reproductive output of female and male spiny lobsters Jams edwardsii (Hutton) (Decapoda, P~inu~dae) in northern New Zealand.

MacDiarmid AB. 1994. Cohabitation in the spiny lobster Jasus edwardsii (Hutton, 1875). Crustaceana:341-355.

MacDiarmid AB, et al. 2013. New Zealand marine ecosystem services. Ecosystem Services in New Zealand—Conditions and Trends. Manaaki Whenua Press: Lincoln, New Zealand.

Mandeville EG, et al. 2015. Highly variable reproductive isolation among pairs of Catostomus species. Mol Ecol 24:1856-1872.

Manel S, et al. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189-197.

Marshall DJ, et al. 2010. Phenotype–environment mismatches reduce connectivity in the sea. Ecol Lett 13:128-140.

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

Mastretta-Yanes A, et al. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour 15:28-41.

Matala AP, et al. 2014. Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. Evolutionary Applications 7:682-701.

Matala AP, et al. 2011. Resolving Adaptive and Demographic Divergence among Chinook Salmon Populations in the Columbia River Basin. Transactions of the American Fisheries Society 140:783-807.

McGaugh SE, et al. 2012. Recombination Modulates How Selection Affects Linked Sites in Drosophila. PLoS Biol 10:e1001422.

McGlauflin MT, et al. 2011. Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River Lakes, Alaska. Transactions of the American Fisheries Society 140:763-782.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303.

McKoy J. 1983. Movements of rock lobsters, Jasus edwardsii (Decapoda: Palinuridae), tagged near Stewart Island, New Zealand. New Zealand Journal of Marine and Freshwater Research 17:357-366.

McKoy JL, et al. 1981. Growth of rock lobsters (Jasus edwardsii) in the Gisborne region, New Zealand. New Zealand Journal of Marine and Freshwater Research 15:121-136.

Mesak F, et al. 2014. Hundreds of SNPs vs. dozens of SSRs: which dataset better characterizes natural clonal lineages in a self-fertilizing fish? Frontiers in Ecology and Evolution 2.

Mesnick SL, et al. 2011. Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Mol Ecol Resour 11 Suppl 1:278-298.

Metzker ML. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11:31-46.

Milano I, et al. 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). Mol Ecol 23:118-135.

Misra JR, et al. 2011. Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes & development 25:1796-1806.

Mita S, et al. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol 22:1383-1399.

Moen T, et al. 2008. Identification and characterisation of novel SNP markers in Atlantic cod: Evidence for directional selection. BMC Genet 9:18.

Mora C, et al. 2009. Management effectiveness of the world's marine fisheries. PLoS Biol 7:e1000131.

Morgan EM, et al. 2013. Investigation of genetic structure between deep and shallow populations of the southern rock lobster, Jasus edwardsii in Tasmania, Australia. PLoS One 8:e77978.

Morin PA, et al. 2004. SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208-216.

Mortensen AS, et al. 2007. Modulation of xenobiotic biotransformation system and hormonal responses in Atlantic salmon (Salmo salar) after exposure to tributyltin (TBT). Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 145:431-441.

Muir JF. 2012. Fish, feeds, and food security. Animal Frontiers 3:28-34.

Narum SR, et al. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol Ecol 17:3464-3477.

Narum SR, et al. 2011. Comparison of F(ST) outlier tests for SNP loci under selection. Mol Ecol Resour 11 Suppl 1:184-194.

Nayfa MG, et al. 2016. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (Pinctada maxima). Marine Genomics 28:99-106.

Nielsen EE, et al. 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (Gadus morhua). BMC Evol Biol 9:276.

Nordeide JT, et al. 2011. Population connectivity among migratory and stationary cod Gadus morhua in the Northeast Atlantic—A review of 80 years of study. Marine Ecology Progress Series 435:269-283.

Nosil P, et al. 2008. Heterogeneous Genomic Differentiation between Walking-Stick Ecotypes: "Isolation by Adaptation" and Multiple Roles for Divergent Selection. Evolution 62:316-336.

Nosil P, et al. 2009. Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375-402.

Nosil P, et al. 2005. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705-719.

NZ S. 2016. NZ safood exports.

Oksanen J, et al. 2013. Package 'vegan'. Community ecology package, version 2.

ÓLafsdottir GÁ, et al. 2007. Morphological and genetic divergence of intralacustrine stickleback morphs in Iceland: a case for selective differentiation? Journal of Evolutionary Biology 20:603-616.

Olsen MT, et al. 2011. A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (Pusa hispida hispida). Mol Ecol Resour 11 Suppl 1:9-19.

Orsini L, et al. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22:5983-5999.

Ovenden J, et al. 1992. Mitochondrial DNA analyses of the red rock lobsterJasus edwardsii supports an apparent absence of population subdivision throughout Australasia. Marine Biology 112:319-326.

Ozerov M, et al. 2013. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. PLoS One 8:e82434.

Palsbøll PJ, et al. 2007. Identification of management units using population genetic data. Trends Ecol Evol 22:11-16.

Pante E, et al. 2014. Use of RAD sequencing for delimiting species. Heredity (Edinb) 114:450.

Paris JR, et al. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8:1360-1373.

Park SG. 2014. Bay of Plenty Marine Sediment Contaminants Survey 2012: Bay of Plenty Regional Council.

Parmesan C, et al. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37.

Pechenik JA. 2006. Larval experience and latent effects—metamorphosis is not a new beginning. Integrative and Comparative Biology 46:323-333.

Pegadaraju V, et al. 2013. De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. BMC Genomics 14:556.

Perez-Figueroa A, et al. 2010. Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23:2267-2276.

Perez S, et al. 2010. Spatial regression techniques for inter-population data: studying the relationships between morphological and environmental variation. Journal of Evolutionary Biology 23:237-248.

Perry AL, et al. 2005. Climate Change and Distribution Shifts in Marine Fishes. Science 308:1912-1915.

Perry RI, et al. 2010. Sensitivity of marine systems to climate and fishing: Concepts, issues and management responses. Journal of Marine Systems 79:427-435.

Peterson BK, et al. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135.

Pigliucci M. 2001. Phenotypic plasticity: beyond nature and nurture: JHU Press.

Pilot M, et al. 2014. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity (Edinb) 112:428-442.

Pineda J, et al. 2009. Complexity and simplification in understanding recruitment in benthic populations. Population Ecology 51:17-32.

Pinho C, et al. 2008. Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (Podarcis spp.) are an assemblage of incipient species. BMC Evol Biol 8:63.

Pocwierz-Kotus A, et al. 2015. Genetic differentiation of brackish water populations of cod Gadus morhua in the southern Baltic, inferred from genotyping using SNP-arrays. Mar Genomics 19:17-22.

Pogson GH, et al. 2003. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (PanI) locus. Mol Ecol 12:63-74.

Poland JA, et al. 2012. Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5:92-102.

Polato NR, et al. 2013. Variation in the transcriptional response of threatened coral larvae to elevated temperatures. Mol Ecol 22:1366-1382.

Pool JE, et al. 2010. Population genetic inference from genomic sequence variation. Genome Res 20:291-300.

Prada C, et al. 2013. Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. Proceedings of the National Academy of Sciences 110:3961-3966.

Prada C, et al. 2014. Strong Natural Selection on Juveniles Maintains a Narrow Adult Hybrid Zone in a Broadcast Spawner. The American Naturalist 184:702-713.

Pritchard JK, et al. 2003. Documentation for STRUCTURE software: version 2.

Puebla O, et al. 2012. On the spatial scale of dispersal in coral reef fishes. Mol Ecol 21:5675-5688.

Puritz JB, et al. 2014a. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431.

Puritz JB, et al. 2014b. Demystifying the RAD fad. Mol Ecol 23:5937-5942.

Reiskind MOB, et al. 2016. Development of a universal double-digest RAD sequencing approach for a group of nonmodel, ecologically and economically important insect and fish taxa. Mol Ecol Resour 16:1303-1314.

Reiss H, et al. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries 10:361-395.

Relyea RA. 2002. Costs of phenotypic plasticity. The American Naturalist 159:272-282.

Reusch TB, et al. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proc Natl Acad Sci U S A 102:2826-2831.

Riginos C, et al. 2013. Seascape Genetics: Populations, Individuals, and Genes Marooned and Adrift. Geography Compass 7:197-216.

Riginos C, et al. 2001. Larval spatial distributions and other early life–history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. Proceedings of the Royal Society of London. Series B: Biological Sciences 268:1931-1936.

Roesti M, et al. 2014. The genomic signature of parallel adaptation from shared genetic variation. Mol Ecol 23:3944-3956.

Rognes T, et al. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

Roullier C, et al. 2013. On the origin of sweet potato (Ipomoea batatas (L.) Lam.) genetic diversity in New Guinea, a secondary centre of diversity. Heredity (Edinb) 110:594-604.

Rousset F. 2008. Genepop: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103-106.

Rowe H, et al. 2011. RAD in the realm of next-generation sequencing technologies. Mol Ecol 20:3499-3502.

Roy D, et al. 2012. Biocomplexity in a demersal exploited fish, white hake (Urophycis tenuis): depth-related structure and inadequacy of current management approaches. Canadian Journal of Fisheries and Aquatic Sciences 69:415-429.

Russello MA, et al. 2012. Detection of outlier loci and their utility for fisheries management. Evolutionary Applications 5:39-52.

Sanford E, et al. 2011. Local Adaptation in Marine Invertebrates. Annual Review of Marine Science 3:509-535.

Schield DR, et al. 2016. EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. Methods in Ecology and Evolution 7:60-69.

Schierup MH, et al. 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genetical Research 76:51-62.

Schmidt PS, et al. 2008. Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. Ecology 89.

Seeb JE, et al. 1998. Allozyme, mtDNA, and microsatellite variants describe structure of populations of pink and sockeye salmon in Alaska. Bull. NPAFC 1:300-319.

Seeb LW, et al. 2011. Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. Mol Ecol Resour 11 Suppl 1:195-217.

Selkoe KA, et al. 2016a. A decade of seascape genetics: contributions to basic and applied marine connectivity. Marine Ecology Progress Series 554:1-19.

Selkoe KA, et al. 2016b. The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages. Proceedings of the Royal Society B: Biological Sciences 283.

Selkoe KA, et al. 2008. Seascape genetics and the spatial ecology of marine populations. Fish and Fisheries 9:363-377.

Selkoe KA, et al. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615-629.

Selkoe KA, et al. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Marine Ecology Progress Series 436:291-305.

Selkoe KA, et al. 2010. Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Mol Ecol 19:3708-3726.

Shafer A, et al. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution.

Shanks AL, et al. 2003. Propagule dispersal distance and the size and spacing of marine reserves. Ecological Applications:S159-S169.

Shendure J, et al. 2008. Next-generation DNA sequencing. Nature biotechnology 26:1135-1145.

Shima JS, et al. 2010. The legacy of dispersal: larval experience shapes persistence later in the life of a reef fish. Journal of Animal Ecology 79:1308-1314.

Siccha-Ramirez ZR, et al. 2018. SNP identification and validation on genomic DNA for studying genetic diversity in Thunnus albacares and Scomberomorus brasiliensis by combining RADseq and long read high throughput sequencing. Fisheries Research 198:189-194.

Silva CN, et al. 2015. Identifying environmental factors associated with the genetic structure of the New Zealand scallop: linking seascape genetics and ecophysiological tolerance. ICES Journal of Marine Science 73:1925-1934.

Silva G, et al. 2014. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. Proceedings of the Royal Society B: Biological Sciences 281.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science 236:787-793.

Smith CT, et al. 2007. Impacts of marker class bias relative to locus-specific variability on population inferences in Chinook salmon: A comparison of single-nucleotide polymorphisms with short tandem repeats and allozymes. Transactions of the American Fisheries Society 136:1674-1687.

Smith GG, et al. 2002. The effect of embryo incubation temperature on indicators of larval viability in Stage I phyllosoma of the spiny lobster, Jasus edwardsii. Aquaculture 209:157-167.

Smith P. 1980. Genetic variation in the rock lobsters Jasus edwardsii and Jasus novaehollandiae. New Zealand Journal of Marine and Freshwater Research 14:37-41.

Smith PJ, et al. 1980. Genetic variation in the rock lobstersJasus edwardsiiandJasus novaehollandiae. New Zealand Journal of Marine and Freshwater Research 14:55-63.

Smithson PA. 2002. IPCC, 2001: climate change 2001: the scientific basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by JT Houghton, Y. Ding, DJ Griggs, M. Noguer, PJ van der Linden, X. Dai, K. Maskell and CA Johnson (eds). Cambridge University Press, Cambridge, UK, and New York, USA, 2001. ISBN 0-521-80767-0 (hardback). International Journal of Climatology 22:1144-1144.

Snelder T, et al. 2005. The New Zealand marine environment classification. Wellington, Ministry for the Environment.

Snelder TH, et al. 2007. Development of an ecologic marine classification in the New Zealand region. Environmental Management 39:12-29.

Sorte CJB, et al. 2010. Marine range shifts and species introductions: comparative spread rates and community impacts. Global Ecology and Biogeography 19:303-316.

Sovic MG, et al. 2015. AftrRAD: a pipeline for accurate and efficient de novo assembly of RADseq data. Mol Ecol Resour.

Starr B. 2001. An examination of stock definitions for the New Zealand rock lobster fishery.

Storfer A, et al. 2010. Landscape genetics: where are we now? Mol Ecol 19:3496-3514.

Strathmann RR, et al. 2002. Evolution of local recruitment and its consequences for marine populations. Bulletin of Marine Science 70:377-396.

Streiff R, et al. 2001. Isolation and characterization of microsatellite loci in the Norway lobster (Nephrops norvegicus). Molecular Ecology Notes 1:71-72.

Stroomberg GJ, et al. 2004. PAH biotransformation in terrestrial invertebrates—a new phase II metabolite in isopods and springtails. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 138:129-137.

Team RC. 2017. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2014. In.

The International Aphid Genomics C. 2010. Genome Sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biol 8:e1000313.

Therkildsen NO, et al. 2013. Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. Mol Ecol 22:2424-2440.

Thomas C, et al. 2000. The effect of temperature on survival, growth, feeding and metabolic activity of the southern rock lobster, Jasus edwardsii. Aquaculture 185:73-84.

Thomas L. 2012. Population Genetics of the Red Rock Lobster, Jasus edwardsii.

Thomas L, et al. 2011. Characterization of polymorphic microsatellite markers for the red rock lobster, Jasus edwardsii (Hutton 1875). Conservation Genetics Resources 4:319-321.

Thomas L, et al. 2013. Testing the consistency of connectivity patterns for a widely dispersing marine species. Heredity (Edinb) 111:345-354.

Tukey RH, et al. 2000. Human UDP-Glucuronosyltransferases: Metabolism, Expression, and Disease. Annual Review of Pharmacology and Toxicology 40:581-616.

Turner TL, et al. 2005. Genomic Islands of Speciation in Anopheles gambiae. PLoS Biol 3:e285.

Väli Ü, et al. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol 17:3808-3817.

Van Buskirk J, et al. 2009. The fitness costs of developmental canalization and plasticity. Journal of Evolutionary Biology 22:852-860.

van der Oost R, et al. 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environmental Toxicology and Pharmacology 13:57-149.

Van Tassell CP, et al. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247-252.

Venter JC, et al. 2001. The Sequence of the Human Genome. Science 291:1304-1351.

Ventura T, et al. 2015. Redefining metamorphosis in spiny lobsters: molecular analysis of the phyllosoma to puerulus transition in Sagmariasus verreauxi. Scientific Reports 5:13537.

Verdu CF, et al. 2016. Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in Robinia pseudoacacia L. Ecology and Evolution 6:7323-7333.

Vigliola L, et al. 2007. Genetic identity determines risk of post-settlement mortality of a marine fish. Ecology 88:1263-1277.

Vignal A, et al. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34:275-306.

Villacorta-Rath C, et al. 2016. Outlier SNPs enable food traceability of the southern rock lobster, Jasus edwardsii. Marine Biology 163:223.

Villacorta-Rath C, et al. 2018. Temporal genetic patterns of diversity and structure evidence chaotic genetic patchiness in a spiny lobster. Mol Ecol 27:54-65.

Viñuela A, et al. 2011. Gene Expression Modifications by Temperature-Toxicants Interactions in Caenorhabditis elegans. PLoS One 6:e24676.

Walther G-R, et al. 2002. Ecological responses to recent climate change. Nature 416:389.

Walther K, et al. 2010. Effects of ocean acidification and warming on the larval development of the spider crab (Hyas araneus) from different latitudes. Marine Ecology Progress Series 417:159-170.

Wang L, et al. 2017. Genome-Wide Association Study Identifies Loci Associated with Resistance to Viral Nervous Necrosis Disease in Asian Seabass. Marine Biotechnology:1-11.

Wang Z, et al. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57.

Waples RS, et al. 2008. Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9:423-449.

Weersing K, et al. 2009. Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series 393:12.

Wei K, et al. 2013. Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. Marine Ecology Progress Series 477:107-121.

Westgaard JI, et al. 2007. Atlantic cod (Gadus morhua L) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. Fisheries Research 85.

White C, et al. 2010. Ocean currents help explain population genetic structure. Proceedings of the Royal Society of London B: Biological Sciences:rspb20092214.

Whiteley NM. 2011. Physiological and ecological responses of crustaceans to ocean acidification. Marine Ecology Progress Series 430:257-272.

Williams LM, et al. 2010. SNP identification, verification, and utility for population genetics in a non-model genus. BMC Genet 11:32.

Willing E-M, et al. 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics 27:2187-2193.

Windig JJ, et al. 2004. Genetics and mechanics of plasticity. Phenotypic plasticity:31-49.

Wittmann AC, et al. 2013. Sensitivies of extant animal taxa to ocean acidification. Nature Climate Change 3:995.

Wood DE, et al. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 15:R46-R46.

Worm B, et al. 2006. Impacts of biodiversity loss on ocean ecosystem services. Science 314:787-790.

Wu L, et al. 2012. Enhanced warming over the global subtropical western boundary currents. Nature Climate Change 2:161.

Yang G-Q, et al. 2016. Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant Methods 12:39.

Zhang J, et al. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614-620.

Zilberman D, et al. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39:61.

Appendices

Appendix 1. Microsatellite outlier detection by Lositan using 500 000 simulations and a false discovery rate of 0.1.
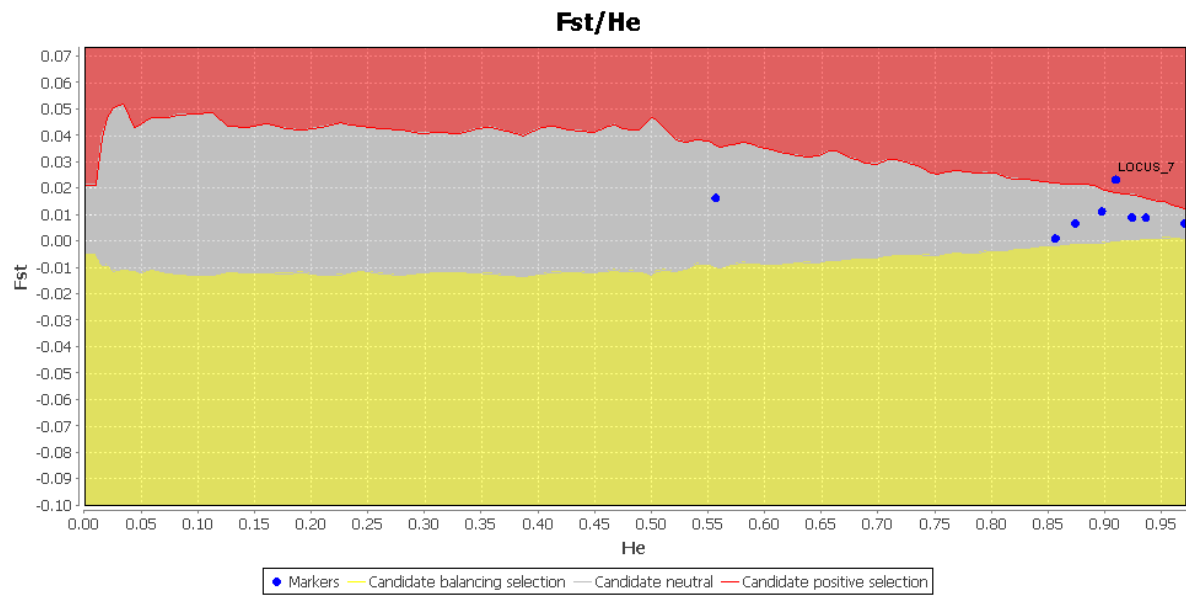
Appendix 2. Descriptive statistics for each sampling location.

Stat Area	Neutral panel			Outlier panel		
	H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}
935	0.413 (±0.208)	0.366 (±0.112)	-0.131	0.628 (±0.321)	0.416 (±0.109)	-0.51
939	0.403 (±0.202)	0.361 (±0.113)	-0.115	0.620 (±0.325)	0.413 (±0.112)	-0.501
901	0.394 (±0.203)	0.360 (±0.114)	-0.098	0.614 (±0.328)	0.414 (±0.113)	-0.487
902	0.410 (±0.200)	0.363 (±0.112)	-0.128	0.629 (±0.319)	0.418 (±0.111)	-0.507
904	0.409 (±0.205)	0.364 (±0.113)	-0.126	0.617 (±0.332)	0.410 (±0.118)	-0.507
907	0.405 (±0.209)	0.361 (±0.115)	-0.121	0.614 (±0.339)	0.406 (±0.126)	-0.516
908	0.408 (±0.224)	0.358 (±0.126)	-0.142	0.613 (±0.345)	0.398 (±0.138)	-0.542
910	0.403 (±0.221)	0.354 (±0.124)	-0.136	0.600 (±0.352)	0.397 (±0.135)	-0.511
911	0.415 (±0.203)	0.368 (±0.110)	-0.131	0.616 (±0.332)	0.406 (±0.122)	-0.518
913	0.407 (±0.202)	0.363 (±0.111)	-0.121	0.614 (±0.334)	0.405 (±0.125)	-0.519
915	0.392 (±0.216)	0.356 (±0.122)	-0.103	0.610 (±0.338)	0.405 (±0.128)	-0.509
916	0.396 (±0.208)	0.361 (±0.113)	-0.098	0.612 (±0.340)	0.403 (±0.127)	-0.519

Stat Area	Neutral panel			Outlier panel		
	H_O (±SD)	H_E (±SD)	F_{IS}	H_O (±SD)	H_E (±SD)	F_{IS}
917	0.412 (±0.199)	0.364 (±0.112)	-0.132	0.620 (±0.332)	0.406 (±0.122)	-0.528
933	0.408 (±0.208)	0.364 (±0.113)	-0.121	0.613 (±0.338)	0.405 (±0.124)	-0.517
942	0.396 (±0.199)	0.361 (±0.112)	-0.094	0.601 (±0.335)	0.404 (±0.126)	-0.491
920	0.399 (±0.208)	0.360 (±0.117)	-0.109	0.614 (±0.336)	0.404 (±0.124)	-0.522
921	0.399 (±0.206)	0.360 (±0.114)	-0.109	0.611 (±0.336)	0.405 (±0.123)	-0.509
924	0.400 (±0.202)	0.361 (±0.115)	-0.105	0.611 (±0.332)	0.407 (±0.120)	-0.504
926	0.419 (±0.200)	0.368 (±0.111)	-0.141	0.621 (±0.329)	0.408 (±0.119)	-0.523
927	0.404 (±0.207)	0.361 (±0.113)	-0.118	0.611 (±0.336)	0.405 (±0.124)	-0.51
928	0.402 (±0.204)	0.364 (±0.113)	-0.108	0.605 (±0.342)	0.401 (±0.125)	-0.509
TAS	0.370 (±0.208)	0.345 (±0.133)	-0.075	0.584 (±0.345)	0.392 (±0.139)	-0.493

Appendix 3. Number of SNPs identified as putatively under selection using three genome scan methods within a subset of four adult lobster sampling locations.
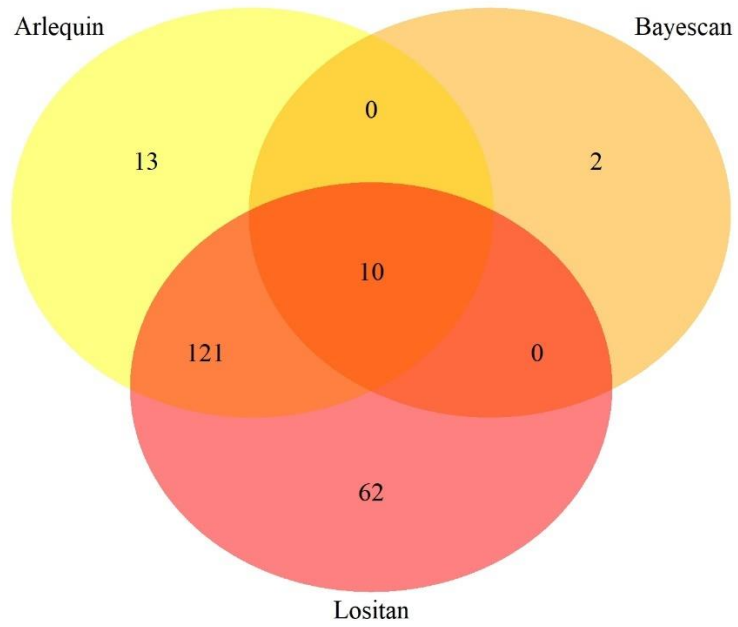
Appendix 4. The matrix of pairwise F_{ST} values between sampling locations calculated using outlier markers.

	935	939	901	902	904	907	908	910	911	913	915	916	917	933	934	942	920	921	924	926	927	928
935	-																					
939	0.006	-																				
901	0.017	0.011	-																			
902	0.020	0.023	0.015	-																		
904	0.031	0.050	0.059	0.052	-																	
907	0.036	0.062	0.063	0.060	0.006	-																
908	0.047	0.080	0.073	0.067	0.022	0.014	-															
910	0.031	0.059	0.065	0.058	0.005	0.003	0.010	-														
911	0.033	0.061	0.064	0.060	0.000	0.004	0.017	0.004	-													
913	0.040	0.069	0.059	0.044	0.023	0.017	0.023	0.014	0.020	-												
915	0.036	0.062	0.051	0.043	0.013	0.008	0.010	0.002	0.007	0.005	-											
916	0.036	0.060	0.068	0.062	0.009	0.000	0.004	-0.003	0.004	0.017	0.009	-										
917	0.054	0.082	0.070	0.053	0.029	0.020	0.025	0.016	0.026	0.008	0.003	0.017	-									
933	0.037	0.067	0.067	0.058	0.007	0.003	0.006	0.001	0.004	0.016	0.001	0.001	0.016	-								
934	0.050	0.074	0.061	0.061	0.017	0.018	0.025	0.016	0.019	0.010	0.008	0.017	0.017	0.012	-							
942	0.048	0.080	0.073	0.063	0.016	0.008	0.012	0.005	0.009	0.008	0.000	0.004	0.009	0.007	0.008	-						
920	0.016	0.037	0.052	0.053	0.035	0.046	0.072	0.032	0.035	0.049	0.045	0.042	0.057	0.050	0.053	0.048	-					
921	0.011	0.028	0.038	0.048	0.037	0.044	0.065	0.033	0.039	0.044	0.042	0.036	0.056	0.043	0.052	0.048	0.009	-				
924	0.021	0.036	0.039	0.051	0.062	0.062	0.076	0.055	0.061	0.054	0.055	0.062	0.068	0.065	0.061	0.064	0.020	0.013	-			
926	0.021	0.039	0.037	0.034	0.060	0.059	0.078	0.054	0.056	0.043	0.052	0.062	0.057	0.067	0.063	0.060	0.020	0.020	0.003	-		
927	0.013	0.029	0.041	0.047	0.050	0.056	0.083	0.048	0.052	0.057	0.057	0.058	0.074	0.059	0.065	0.067	0.010	0.009	0.009	0.013	-	
928	0.016	0.030	0.050	0.057	0.063	0.075	0.100	0.066	0.065	0.079	0.080	0.071	0.093	0.083	0.092	0.083	0.011	0.010	0.012	0.017	0.004	-

Appendix 5. Results of BLAST Conserved domains analysis of the reference sequence 92714_782.