The molecular ecology of an understudied endemic marine Isopod - *Isocladus armatus*

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Natural Science

At Massey University, Albany, New Zealand

William Samuel Pearman

2019

Abstract

The study of populations and the adaptive significance of traits is a major theme in molecular ecology literature. In this thesis I present three lines of research that contribute to the understanding the molecular ecology of a species of New Zealand endemic marine isopod - *Isocladus armatus* (family: *Sphaeromatidae*). The goal of this thesis is to develop and utilize a framework to better understand the genomics of marine isopods from a range of genomic perspectives.

The first primary chapter aims to assess two ways of enriching mitochondrial DNA from whole genome DNA, and to assemble this species mitochondrial genome. My research indicates that an atypical mitochondrial genome structure, widespread across *Isopoda* - but previously thought absent within *Sphaeromatidae*, is present within *I. armatus* suggesting that this trait has been maintained for an order of magnitude longer than previous estimates. The second primary chapter aims to describe and understand the genetic structure of populations for 8 locations around New Zealand, to understand connectivity and dispersal for *I. armatus*. Using a panel of 8,020 loci, I find high gene flow on a small spatial scale, while populations on a larger spatial scale exhibit a pattern of Isolation-By-Distance. Additionally, gene flow over one well known biogeographic barrier was much higher than between any other populations on a similar spatial scale, suggesting this barrier may not exhibit a strong effect on this species. Thus, my research indicates a need to revisit and study the way biogeographic barriers affect species with different life histories.

The final primary chapter aims to understand the genetic basis for colour polymorphism in *I. armatus*, with the intention of understanding the adaptive significance and selective mechanism behind this trait. I use genome wide association approaches with a panel of 20,000 loci to answer these questions. I found that loci associated with Colour Polymorphism exhibited signatures of disruptive selection, contrary to initial hypothesis where I expected balancing selection to main colour polymorphism. I propose that substrate heterogeneity in *Isocladus armatus'* habitat results in microhabitats, each of which imposes a selective pressure benefiting a specific morph type. The size of these microhabitats is so small that high levels of interbreeding between these microhabitats, and thus between morphs, results in the maintenance of polymorphism across the population.

Acknowledgements

Firstly, I would like to thank my partner, Maria, who has been incredibly supportive, patient, and especially tolerant of me ruining many trips to the beach by suddenly disappearing to look for isopods.

Secondly, I'd like to thank my main supervisors - Dr. Nikki Freed, Professor James Dale, and Dr. Sarah Wells, for being patient and encouraging me to pursue new ideas and avenues of research in my thesis. I'd also like to thank Drs Olin Silander, Libby Liggins, David Aguirre, and Vanessa Arranz. Everyone has provided insightful and useful suggestions throughout this project and helped me think of better ways to test my ideas. Thank you also for helping me with statistical analyses and interpretations, as well as for pointing me in the direction of literature I might have otherwise overlooked.

I'd like to thank my family, especially my parents, Dayle and Barry, and my siblings, Rebecca and Ben, for supporting me in doing my masters and helping me out, and especially for going out of their way to get me extra supplies when I was collecting isopods in the middle of nowhere.

Additionally, I'd like to thank my flatmates Ash Sargent and Justine Waterson, who provided many useful suggestions. I'd also like to thank Niel Bruce from the Queensland Museum for taking the time to talk with me about isopod taxonomy, and helping to resolve some of the issues I'd found during my studies. I'd also like to thank the team at the Auckland War Memorial Museum for letting me visit and go through their collections, without them I would most likely have collected the wrong species! Another thank you is owed to Eric Thorstensen for helping me out with access to homogenizers and his many useful suggestions regarding mitochondrial DNA enrichment.

Table of Contents

This thesis is arranged as a series of proposed publications, and as such there is significant duplication of content between some chapters (particularly between methods), as well as with the introductory and summary chapters. Additionally, the introduction and summary chapters are intentionally brief, as each main chapter has its own introduction and discussion.

List of Acronyms, Tables, and Figures

Acronyms

PCR - Polymerase Chain Reaction
MDA - Multiple Displacement Amplification
AMOVA - Analysis of Molecular Variance
CP - Colour Polymorphism
RDA - Redundancy Analysis
WGS - Whole Genome Sequencing
ECE - East Cape Eddy
ECC - East Cape Current
qPCR - Quantitative Polymerase Chain Reaction
CDF - Cumulative Distribution Function
DNA - Deoxyribonucleic Acid
SDS - Sodium Dodecyl Sulfate
EDTA - Ethylenediaminetetraacetic Acid
RFU - Relative Fluorescence Units
BP - Basepairs
PLD - Pelagic Larval Duration

MAF - Minor Allele Frequency
MAC - Minor Allele Count
GB - Gigabase
GBS - Genotyping By Sequencing
RE - Restriction Enzyme
ORI - origin of replication
Indel - Insertion or deletion
RADseq - Restriction site Associated Digest sequencing
MCMC – Markov Chain Monte Carlo

List of Tables

List of Figures

Chapter 1 - Introduction

Isopods are an order of crustacea with an almost global distribution (the only landmass without isopods is terrestrial Antarctica). They are found in freshwater, terrestrial, and marine environments (Poore & Bruce, 2012). Isopods are widely encountered and perhaps best known in terrestrial systems as pill bugs or woodlice. However, isopods are particularly abundant in marine environments (both deep sea and intertidal) where they exhibit a range of sizes and life histories, ranging from 0.3 mm to greater than 36 cm (Poore & Bruce, 2012).

Crustaceans, and isopods in particular, are of great interest to molecular biologists because of their wide diversity of evolutionary histories. Indeed, many crustacean traits have evolved multiple times within both *Crustacea* and *Isopoda.*(Morrison et al., 2002). Isopods are similar in this regard, yet are also extremely morphologically diverse (Poore & Bruce, 2012). As a result, they provide the potential for major insights into evolution due to the diversity of evolutionary processes behind key traits such as colouration (Wells and Dale, 2018; Shuster et al., 2014) or genome architecture (Doublet et al., 2012). In addition, isopods are widely distributed and encountered, which makes them a relatively accessible taxon for study.

Some marine species exhibit parasitic lifestyles, such as *Cymothoa exigua*, which enters a fish's mouth through the gills, and severs the blood vessels in the tongue, causing atrophication and enabling this isopod to act as a pseudo-tongue (Brusca & Gilligan, 1983). However, most marine isopods are detritivores and scavengers feeding on available decaying plant and animal matter (Poore & Bruce, 2012). In these environments (especially intertidal zones), many isopods exhibit cryptic or disruptive colouration to aid in predator avoidance (Hultgren & Mittelstaedt, 2015; Merilaita, 1998). Such colouration will function towards avoidance of key predators such as birds (Weiser & Powell, 2011), crabs (Woods, 1993), anemones (Jansen, 1968), and fish (Jansen, 1968; Jormalainen, Merilaita, & Tuomi, 1995).

All species lack a larval stage and have direct developing young. Eggs are brooded in the female's marsupium, and hatch into a juvenile phase (manca) which resembles an adult except it lacks the last pair of pereopods.

The South-Pacific, and New Zealand in particular, is a hotspot for marine isopod biodiversity (Poore & Bruce, 2012). Despite this, relatively little research has been conducted into the genetics of marine isopods in New Zealand, and even less so within the family *Sphaeromatidae*. Additionally, the most comprehensive taxonomic work of New Zealand Sphaeromatid isopods to date was published in 1977 (Hurley & Jansen, 1977), and much of this work may require revision in light of the development of more sophisticated molecular and morphological analyses (Niel Bruce, Pers. Comm.). To date, there is only a single publication on a New Zealand sphaeromatid isopod using molecular methods (Wells & Dale, 2018). Molecular methods have the potential to resolve many of the issues based around morphological analyses (in terms of classification), as well as helping to understand population connectivity and structure on a broad scale. Additionally, these methods can be used to identify genomic elements underpinning physical traits, such as colouration (Stinchcombe & Hoekstra, 2008; Wells & Dale, 2018).

DNA sequencing of mitochondria and genomic DNA allow assessment of population structure

Deeper insights into isopod evolution can be yielded through the sequencing and assembly of mitochondrial genomes. From a phylogenetic perspective this can be useful in order to reconstruct phylogenetic trees with greater resolution than standard approaches (Sullivan, Platt, Bradley, & Ray, 2017). Specifically, mitochondrial genomes contain more information for phylogenetic inference as a whole, rather than a few specific autosomal loci (Duchêne, Archer, Vilstrup, Caballero, & Morin, 2011). However, the application of this genomic data within *Isopoda* for phylogenetic inference is hindered by the lack of available sequenced isopod mitochondrial genomes, particularly within sphaeromatid isopods. While the majority of available mitochondrial genomes are found within the suborders *Asellota* and *Oniscidea*, only a single mitochondrial genome is available for species within the suborder *Sphaeromatidea*.

The use of mitochondrial genomes for isopod research is further hampered by the fact that some mitochondrial genomes across *Isopoda* exhibit an atypical structure that is not well understood outside of the suborder *Oniscidea.* This structure is, at present, thought to be absent within the family *Sphaeromatidae* (Doublet et al., 2012). This absence has been hypothesized to be due to a reversion to a ‘typical’ structure, with the atypical form proposed as the ancestral state for *Isopoda*.

This atypical structure (see figure 1.1 for a diagram) consists of two chromosomes - a linear 14.4Kb monomer, and a 28Kb dimer made up of two copies of the monomer fused together as inverted repeats (Doublet et al., 2012; Peccoud et al., 2017). This structure is further complicated by the fact that on the dimer, there are three loci that are each able to code for two tRNAs depending on which side of the dimer is in question.

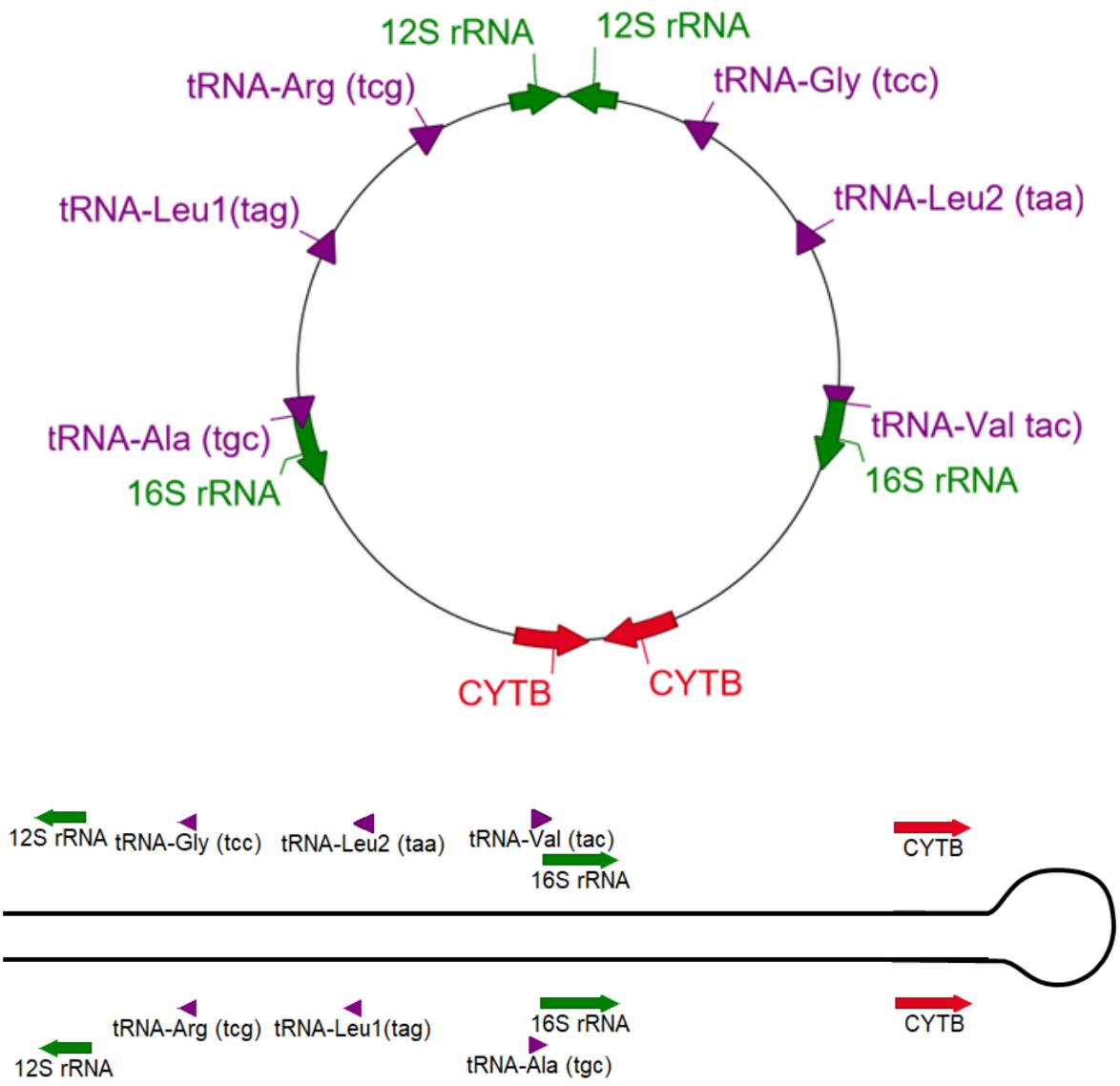


Figure 1.1 The currently proposed atypical structure ancestral to *Isopoda*, the key features are a large 28kb circular chromosome consisting of two inverted repeats, and a linear chromosome with a similar structure and a hairpin at the end. The arrows on the plasmid map point to sites are different between each side of the chromosome. Text inside of brackets associated with tRNA's indicates the codon sequence (i.e. tRNA-Ala (tgc)).

In addition to using mitochondrial sequences for understanding evolution, recent advances in reduced representation DNA sequencing have enabled genome level profiling of many individuals in a population. For example, recent studies have employed Restriction site Associated DNA Sequencing (RADseq) and Genotyping-By-Sequencing (GBS) techniques to resolve genetic connectivity among closely related populations (Elshire et al., 2011; Kahnt et al., 2018; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012; Vendrami et al., 2017). These approaches can be particularly useful as they have high resolution and make use of loci across the entire genome - enabling greater precision in inference of population structure (Morin, Martien, & Taylor, 2009).

The use of these genetic and genomic techniques can be beneficial in marine systems, where gene flow can be difficult to assess for some species using traditional ecological approaches (such as mark-recapture studies, tagging of birds or using GPS trackers (Lutcavage et al., 2000)). In particular, marine species (and often biphasic species, which are defined by having a juvenile larval stage) may be hard to study due to the often large number of offspring produced that incur high mortality (r-selected species). This can prevent, or hinder attempts to understand population connectivity in the absence of genetic data (Thorrold et al., 2002). Additionally, for biphasic species that undergo dispersal during the larval form, it is often difficult, if not impossible, to track these larvae because they are incredibly small relative to the size of the ocean (Levin, 2006). This has resulted in many studies employing genetic/genomic approaches for biphasic species, leading to a relative paucity of information for monophasic species, also referred to as direct developers (these are species that have a single life stage rather than having a larval stage). For example, a recent review of population genetic studies on benthic intertidal invertebrates in New Zealand showed only 10 out of 56 studies were focused on direct developers (Arranz Martinez, 2017). However, in both cases population genomics can be useful to detect extremely fine grain population structure (as been previously observed within *Isopoda* (Piertney & Carvalho, 1994)).

Reduced representation sequencing can link phenotype to genotype

The use of reduced representation genome sequencing can be used not only for population genomics but can also be used for correlating phenotypic traits to the underlying genotype. The abundant molecular markers known as SNPs (Single Nucleotide Polymorphisms) generated from reduced representation sequencing can be used to infer both population structure, and identify loci that may be associated with certain traits (Iles, 2008). This approach is widely used to understand the genetic basis of diseases in humans such as

Alzheimer's disease (Bertram & Tanzi, 2009), however this approach has been extended to traits of interest in evolutionary studies, such as colour polymorphism (Kim et al., 2019; Stinchcombe & Hoekstra, 2008; Wells & Dale, 2018). Colour polymorphism (CP) is the co-existence of at least two discrete colour morphs within a population, for which the presence cannot be attributed to recurrent genetic mutations, and where each morph has a genetic basis (Huxley, 1955). CP is found widely within eukaryotes, and has been the subject of significant attention in birds (Galeotti, Rubolini, Dunn, & Fasola, 2003), insects (Nosil, 2007) and crustaceans (Detto, Hemmi, & Backwell, 2008; Nokelainen, Stevens, & Caro, 2018). Within *Isopoda*, as well as many other marine invertebrates, colouration has been proposed to play an important role in predator avoidance. This is the case for *Pentidotea wosnesenskii*, where colouration has been shown to enable background matching and thus greater camouflage (Hultgren & Mittelstaedt, 2015). Alternatively, in the isopod *Idotea baltica*, disruptive colouration is thought to play a role in obscuring the shape of the organism to aid in predator avoidance (S. Merilaita, 1998). Outside of *Isopoda*, CP has been hypothesized to be naturally selected for background matching in three species of crab (Nokelainen et al., 2018; Palma & Steneck, 2001; M. Stevens, Lown, & Wood, 2014).

As mentioned above, SNP data from population genomic studies (such as RADseq and GBS approaches) has been used to identify the selective mechanism and genetic basis for CP (Kim et al., 2019; Wells & Dale, 2018). This was done by identifying alleles for which their presence was correlated with the expression of different phenotypes, such as colouration (Kim et al., 2019; Wells & Dale, 2018). The selective mechanisms behind colouration can also be identified using an Fst outlier approach (Foll & Gaggiotti, 2008). This approach identifies loci that exhibit greater than expected levels of differentiation (based on a measure of genetic differentiation called Fst) with respect to phenotype (relative to the distribution of Fst for neutral loci), in order to identify whether these loci are being acted upon by selection (Narum & Hess, 2011). The specific selective mechanisms can be difficult to determine, as available tools (e.g BayeScan) are only able to distinguish between disruptive/divergent selection, and balancing selection.

Understanding the selective mechanisms for colouration is particularly interesting because there are four primary hypotheses that have been proposed to explain the maintenance of CP within species (Galeotti et al., 2003; T. E. White & Kemp, 2016). The first is sexual selection, where processes such as disassortative mating (Huynh, Maney, & Thomas, 2011) or mate recognition (Detto et al., 2008) can act to maintain multiple phenotypes within a population.

The second hypothesis is balancing selection, where multiple alleles are maintained within a population for which the frequencies cannot be attributed solely to genetic drift. A primary example, and one considered to be a major mechanism for CP maintenance, is negative frequency dependent selection (Galeotti et al., 2003; T. E. White & Kemp, 2016). This occurs when the rarity of an allele confers greater fitness - such that if an individual has a rare phenotype then it may have greater fitness than the more common phenotypes.

The third hypothesis is that divergent selection may maintain CP when it is mediated by gene flow between populations. Divergent selection has the potential to maintain CP as it results in selection for different phenotypes in different populations. In the absence of gene flow this could lead to speciation, but when gene flow is high enough to overcome the strength of selection, it could result in the maintenance of multiple phenotypes within the same population, as well as account for the presence of intermediate morphs (Nosil, 2007). Finally, the fourth hypothesis is that a similar process can occur but within a population instead of between them. This mechanism, known as disruptive selection, can act to maintain CP but in this case the scale of gene flow is between morph types within a population (Rueffler, Van Dooren, Leimar, & Abrams, 2006).

This thesis examines three aspects of the molecular ecology of *Isocladus armatus*, a marine isopod endemic to New Zealand. *I. armatus* is an abundant sphaeromatid isopod found commonly around New Zealand in rocky shores and intertidal zones. Both within and among populations, this species exhibits extremely high levels of colour polymorphism. While colouration in this species varies across a continuum, many discrete morphs can be identified. Despite this massive variation, little work has been conducted to understand the evolution and maintenance, as well as the genetics of CP in this species. Jansen (1971) observed a relationship between temperature and salinity with morph frequency over a small spatial scale, and proposed these factors as potentially important in CP maintenance (Jansen, 1968). Recent work by Wells and Dale (2018) has indicated a strong genetic basis for CP in this species, but suggested more genetic work on a small spatial scale was required to identify the selective mechanisms and genetic basis of specific morphs. Additionally, this work indicated relatively high gene flow across a small spatial scale, but was unable to draw conclusions about connectivity across a larger range. *I. armatus* has a strong swimming ability (Morton & Miller, 1973) which has the potential to reduce gene flow between populations by resisting tidal displacement (Jansen, 1968), or to increase gene flow by enabling dispersal across greater distances (Wells & Dale, 2018). Finally, being a sphaeromatid isopod, *I. armatus* is thought to have a normal mitochondrial structure but this has not been tested directly. Determining the mitochondrial structure therefore has the

potential to help understand the evolution and paraphyly of atypical mitochondrial structure within *Isopoda*.

Specific Aims

Because *I. armatus* is highly abundant and widely distributed, it is a useful candidate for answering a multitude of questions regarding colouration, mitochondrial genetics, and population connectivity for direct developers in New Zealand. The first chapter of this thesis aims to assess two ways of enriching DNA samples for mitochondrial DNA with the goal of generating a mitochondrial genome for *Isocladus armatus*, and to use this genome to understand the evolution of atypical mitochondrial structure within *Isopoda*. The second chapter is a population genomic analysis of 9 populations of *I. armatus*, aiming to understand population connectivity on both large and small spatial scales. Finally, the third chapter provides an initial investigation into the genomic basis for colour polymorphism in *I. armatus*, using the same panel of SNPs used in the second chapter.

Chapter 2 – Long read sequencing reveals a novel mitochondrial genome structure for *Isocladus armatus*

Abstract

Mitochondrial genomics is widely used as a means of inferring evolutionary histories and phylogenies of a broad range of taxa. Sequencing and assembling mitochondrial genomes often relies on implicit assumptions regarding the size and structure of the mitochondrial genome. If these assumptions are not met, inferences regarding the mitochondrial genome (such as size, gene loss, or duplication) may be affected. In this study I use a marine isopod, *Isocladus armatus* (family: *Sphaeromatidae*) as a study species, to compare two approaches of enriching whole genome DNA samples (Multiple Displacement Amplification, and differential centrifugation) for mitochondrial DNA for long-read nanopore sequencing. I found either no increase, or a decrease, in abundance of mitochondrial DNA from both approaches for enrichment relative to the abundance of mitochondrial DNA obtained from shotgun sequencing. In addition, I use this nanopore sequencing to sequence and assemble the mitochondrial genome for *I. armatus*, this revealed a highly atypical mitochondrial genome that, while widespread across *Isopoda*, was thought absent in *Sphaeromatidae*. This research provides some evidence that this structure may be ancestral to *Isopoda* and potentially an order of magnitude older (400 million rather than 40 million years) than previous estimates.

2.1. Introduction

Mitochondria are an essential organelle of the eukaryotic cell, due to their role in energy metabolism. Current research proposes that mitochondria evolved through the incorporation of an early alpha-proteobacteria into a eukaryote ancestor (Gray, 2012; Roger, Muñoz-Gómez, & Kamikawa, 2017). This is thought to represent one of the most significant events in the evolution of energy metabolism (Okie Jordan G., Smith Val H., & Martin-Cereceda Mercedes, 2016) and it has resulted in animal cells possessing both nuclear and mitochondrial genomes. As a result, mitochondria inarguably possess characteristics of bacteria in terms of replication and evolution. This has meant that while mitochondrial and

nuclear genomes tend to undergo the same evolutionary trajectory, specific characteristics of mitochondria mean they can be particularly useful for inferring relationships between populations or species. These characteristics include the fact that 1) mitochondria tend to be overrepresented in DNA samples relative to nuclear genomic DNA due to their high abundance in a cell (the amount varies with tissue type and species, but in mammals tends to be between 80 and 2000 per cell) (Cole, 2016; Robin & Wong, 1988), 2) mitochondria are well studied and large amounts of data are available, which enables easier contextualization of results, 3) they are relatively conserved in terms of structure, and thus are relatively easy to study without much prior species specific information (Taanman, 1999; E. C. Yang et al., 2015), and finally 4) they are relatively constrained evolutionarily compared to the nuclear genome, due to their role in coding for critically important metabolic processes.

Mitochondria play a fundamental role in eukaryote cells because they synthesize large amounts of ATP, a molecule required for many biochemical processes. Any mutations that reduce ATP production are therefore deleterious, and in many cases potentially lethal, this means that mutations in mitochondrial genes are highly constrained. Additionally, because mitochondrial genomes replicate independently of the nuclear genome and are generally uni-parentally inherited, they tend to not undergo recombination (with some exceptions (Cotton, 2001; Luo et al., 2018)). These characteristics result in a highly conserved mitochondrial genome (in terms of both structure, and at the nucleotide level) meaning that mitochondria can be studied under a relatively simple evolutionary model. However, there are instances where this mitochondrial structure is not necessarily conserved, and organisms have evolved other ways to retain the essential role of mitochondria without the 'typical' structure (Doublet et al., 2012; Liu et al., 2017; Stampar et al., 2019; Wei et al., 2012).

2.1.1 Mitochondrial Structure across Eukaryotes

Within Bilateria (animals with bilateral symmetry), mitochondrial genomes tend to be circular in structure and contain 37 genes (13 protein-coding, two rRNAs, and 22 tRNAs, within a conserved arrangement). However, as mentioned above, this structure is not universal. Outside of Bilateria, non-typical mitochondrial genomes are relatively common, such as a linear mitochondrial chromosome within the class of marine invertebrates *Anthozoa* (Kayal et al., 2012; Stampar et al., 2019). But even within Bilateria, unusual mitochondrial arrangements are occasionally found in some lineages. For example, booklice (order: *Psocoptera*) possess a multipartite mitochondrial genome consisting of two circular

chromosomes (Wei et al., 2012). While thrips (order: *Thysanoptera*) also possess a multipartite mitochondria, but also exhibit massive size asymmetry (0.9kb and 14kb chromosomes) (Liu et al., 2017). Additionally, lineages with atypical structure are not united evolutionarily, and many closely related species still possess a 'typical' mitochondria.

2.1.2 Mitochondrial Structure within *Isopoda*

Current research suggests that isopods, in particular the suborder *Oniscidea*, are a lineage with an atypical mitochondrial structure. Specifically, they exhibit a multipartite mitochondrial genome that consists of both a linear and circular chromosome - each composed of similar copies of the main 'unit'. The linear chromosome consists of one copy of the mitochondrial unit, while the circular chromosome consists of two copies of the unit fused together as inverted repeats (Fig 2.1) (Peccoud et al., 2017). Each unit retains a relatively 'typical' gene arrangement. The terms used to describe these structures vary across the literature. As such, I will use the term 'unit' to refer to the fundamental mitochondrial unit; this being approximately 50% of the dimer, or the full length of the linear chromosome. 'Monomer' will be used to refer to the linear chromosome that consists of the unit as a linear piece of DNA, while 'dimer' will refer to the circular chromosome, which consists of two copies of the 'unit' fused together as inverted repeats. This unusual structure (see Fig. 2.1) has been proposed to be an ancestral state within *Isopoda* due to its presence across divergent lineages (such as *Asellota* and *Oniscidea*) (Fig. 2.2) (Doublet et al., 2012).

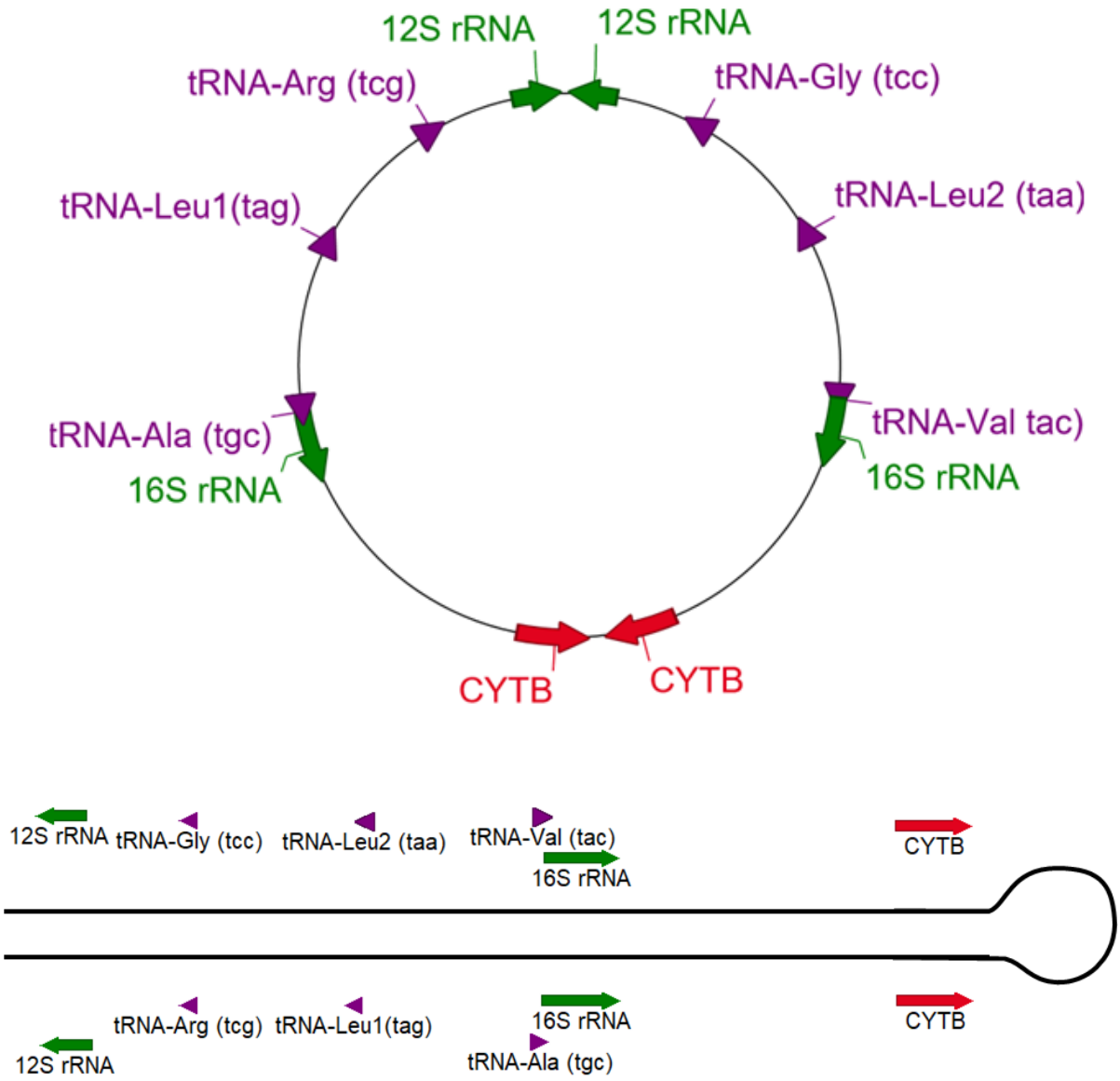


Figure 2.1 The currently proposed atypical structure ancestral to *Isopoda*, the key features are a large 28kb circular chromosome consisting of two inverted repeats, and a linear chromosome with a similar structure and a hairpin at the end. The arrows on the plasmid map point to sites are different between each side of the chromosome. Text inside of brackets associated with tRNA's indicates the codon sequence (i.e. tRNA-Ala (tgc)).

The ancestral state hypothesis, however, is partially contradicted by early research indicating its absence within *Sphaeroma serratum* and presumably the *Sphaeromatidae* as a whole (Doublet et al., 2012). In Figure 2.2 I summarize the current state of knowledge regarding the presence of this atypical structure across *Isopoda*. Three studies suggest that some *Sphaeromatid* isopods exhibit 'typical' mitochondrial genomes. These studies are limited to *Sphaeroma serratum* (Doublet et al., 2012; Kilpert, Held, & Podsiadlowski, 2012)

and *Sphaeroma terebrans* (Yang, Gao, Yan, Chen, & Liu, 2019) - and the mitochondrial genome for the latter species has not been made public.

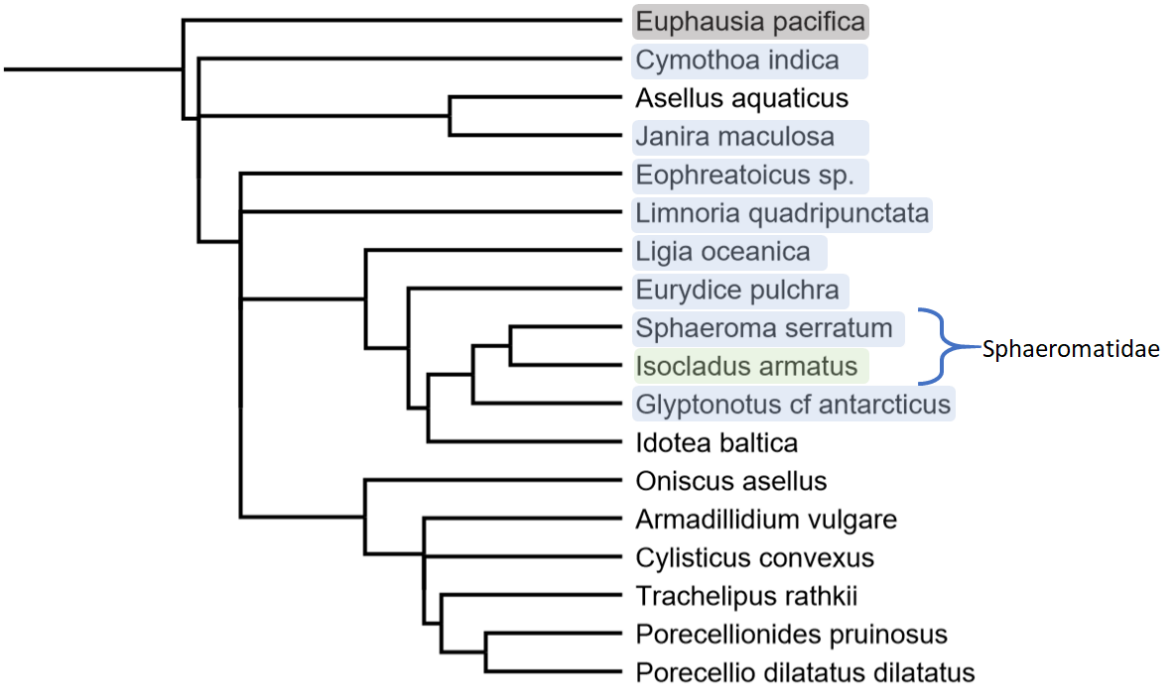


Figure 2.2 - Modified phylogenetic tree from Zou et al. (2018) - branch lengths are not to scale. *Euphausia pacifica* is an outgroup with a 'typical' mitochondrial genome (grey). Species highlighted in blue are isopods that possess a 'typical' mitochondrial genome, all uncoloured species have atypical structure, and Isocladus armatus (green) is our species of interest and has been manually inserted into this tree.

Within *Oniscidea* there are a few nucleotide changes between each unit of the dimer. These include three single nucleotide substitutions in three different tRNAs. These substitutions mean that the three tRNA loci are each able to code for two tRNAs (Peccoud et al., 2017). Specifically, the tRNA loci on one side of the dimer are able to code for a different tRNA at the corresponding loci on the other side of the dimer (Fig. 2.1). I refer to these types as either SNPs (when in reference to the mitochondrial unit) or as heteroplasmic sites (when in reference to a dimer), following the convention of Doublet et al. (2013).

These differences suggest that the monomer may potentially be the result of self-renaturation of a single stranded dimer during replication (Doublet, Helleu, Raimond, Souty-Grosset, & Marcadé, 2013). This is supported by recent research that has indicated that the monomer has non-complementary bases at the SNP sites, and that the CYTB end of the

monomer has a hairpin structure that should prevent monomer replication (Peccoud et al., 2017).

2.1.3 Approaches to Obtaining Complete Mitochondrial Genomes

Mitochondrial genomes can yield insights into the evolution of an organism, or group of organisms from two angles. The first is, as previously discussed, the actual structure of the genome and arrangement of genes within the mitochondria. The second is the use of mitochondrial genes in reconstructing the evolutionary history of a group of species based on phylogenetic analysis of the mitochondrial genome. Mitochondrial genomes can be highly informative for phylogenetic analysis because their evolution and replication is relatively simple, and is reasonably well understood.

While the majority of studies investigating phylogenetics from a mitochondrial perspective have employed PCR, recent work has employed various other methods to acquire the entire mitochondrial genome without PCR (Macher, Zizka, Weigand, & Leese, 2018; Nunez & Oleksiak, 2016; Wolff, Shearman, Brooks, & Ballard, 2012). One widely used method is to use a plasmid purification process, which is based on the ability of small, circular DNA to more easily renature than larger linear fragments of DNA. 'Typical' mitochondrial DNA is thus isolated as it is small and circular. However, this process is not necessarily applicable if there is a possibility of non-circular mitochondrial chromosomes. An alternative method, genome-skimming, is a relatively recent method whereby mitochondrial sequences are identified and extracted from whole genome sequencing (WGS) datasets (Johri et al., 2019; Lin, Xiang, Sampurna, & Hsiao, 2017). This method can be particularly effective because within sequencing data, mitochondrial genomes tend to be overrepresented relative to the nuclear genome. Therefore, while WGS datasets may not have enough data to assemble a whole genome, there may be enough data to assemble a mitochondrial genome. Multiple tools have been developed specifically for this purpose, however they rely on a 'typical' bilaterian mitochondrial structure, or at least some reference data of either the same species or a closely related one (Al-Nakeeb, Petersen, & Sicheritz-Pontén, 2017; Dierckxsens, Mardulyn, & Smits, 2017; Hahn, Bachmann, & Chevreux, 2013; Meng, Li, Yang, & Liu, 2019). These software tend to be designed to work with Illumina sequencing data, which while highly accurate, outputs only very short sequences (maximum of 300 bp). Consequently, some software may struggle to assemble a mitochondrial genome for species where there may be major structural rearrangements or variations relative to the reference

mitochondria supplied, as short sequences may not best capture the structure of the mitochondrial genome. For example, the tool MitoBIM, makes use of a reference gene or mitochondrial genome, which is used to identify initial mitochondrial reads and assemble these into a draft genome. This process is repeated, with each iteration incorporating the previously identified mitochondrial reads into the new draft assembly. Alternatively, in cases where *de novo* assembly of mitochondria is being used, it may be difficult to assemble the mitochondria or assess the quality of the assembly without prior expectations. One potential solution to this is to massively increase the coverage of the mitochondria to increase the likelihood of successful assembly. However, when using WGS data, these are not always applicable, especially in cases like *Isopoda*, where the structure of the mitochondrial genome may be impossible to resolve from just Illumina sequences. This would be due to difficulty in identifying reads from each side of the dimer due to the degree of repetitiveness.

Some issues relating to mitochondrial assembly (such as insufficient coverage) may be unavoidable when working with pre-existing datasets, but can be avoided when generating data. This can be done by either increasing sequencing throughput, or by enriching for mitochondrial DNA. One way to enrich for mitochondrial DNA could be to process the tissue in such a way that the integrity of the mitochondria is maintained, while the cell itself is lysed. Following this with a series of differential centrifugations has been shown to enrich for mitochondria. This approach was employed by Macher et al. (2018), who observed a 55-fold increase in concentration of mitochondrial DNA. Shotgun sequencing of such an enriched sample will increase coverage of the mitochondrial DNA and facilitate better assemblies.

An alternative tactic to enrich for mitochondrial DNA, is to use a Multiple Displacement Amplification (MDA) reaction. This makes use of rolling circle amplification alongside mitochondrial specific primers (such as COX1 and 16S). The reaction amplifies DNA using Phi29 polymerase at 30°C, which continually amplifies DNA in a circular fashion (Wolff et al., 2012). Hypothetically, the structure of the mitochondrial genome would be maintained throughout the amplification, meaning shotgun sequencing could be applied to this product. However, in cases like that of *Isopoda*, such a technique may be problematic because an MDA reaction would result in pieces of DNA that exhibit duplicated regions (due to the nature of rolling circle amplification). Therefore, in cases where mitochondrial genomes are already duplicated - it may be difficult to identify the correct structure/degree of duplication using an MDA approach.

Newer methods are becoming available to perform long-read sequencing with preferential sequencing of certain loci. One approach uses whole genomic DNA with a Cas9-mediated

enrichment (Gilpatrick et al., 2019). This is done by first de-phosphorylating the ends of the DNA, which reduces the ligation efficiency of the sequencing adaptors to the ends of the DNA. The Cas9 enzyme, along with guide RNAs (designed for the loci of interest) are then introduced into the reaction. This cleaves the DNA, and results in phosphorylated open ends of DNA at the loci of interest. Sequencing adaptors, which preferentially ligate to phosphorylated DNA (which are now found at the loci of interest due to DNA cleavage), can then be introduced into the reaction for ligation. Recently this technique was applied within human research to study methylation of cancer associated genes (Gilpatrick et al., 2019), and could be feasibly applied to mitochondrial genomics.

The biggest problems with many of these techniques, is they are potentially very costly, and they also require prior knowledge of the mitochondrial genome structure. For example, inferring the structure of a multi-partite mitochondrial genome with a single copy of the 16S rRNA gene in one of the chromosomes (such as within Order: *Psocoptera*) could be difficult when using an MDA or Cas9 enrichment. This is because the use of the 16S gene (or any other widely used barcoding gene, i.e COX1) as a source to design the guide RNA or the primers for the MDA would result in massive enrichment of one chromosome relative to the other. The resulting inference could therefore artefactually suggest massive loss of mitochondrial genes.

Despite their wide usage in biology, the majority of approaches tend to make use of a few mitochondrial genes amplified through Polymerase Chain Reaction (PCR). For *Isopoda* this is particularly prevalent, where relatively few sequenced mitochondrial genomes exist across the order; the majority of which are found within a single suborder (*Asellota*) (Benson et al., 2013). This paucity of mitochondrial data can hamper phylogenetic inference between closely related species, when relying on one or a few genes (Gee, 2003; Philippe et al., 2011). A key example of this is in the family *Sphaeromatidae* - a ubiquitously distributed family of marine isopods (Wetzer, Bruce, & Pérez-Losada, 2018). A recent study attempted to resolve the phylogenetic relationships of species within this family using a mitochondrial gene (16S rRNA) and a nuclear gene (18s rRNA) (Wetzer et al., 2018). This approach worked well for the majority of species, but some relationships were incongruent between each gene - specifically within the *Exosphaeroma* clade (where *Exosphaeroma* includes *Isocladus*), where *Exosphaeroma* as a genus was not widely supported and included two other genera within it (Wetzer et al., 2018). In situations like this, where there is incongruence between trees, or relationships are not clear - increasing the amount of data may help resolve the issue. In this case, a complete mitochondrial genome could provide greater phylogenetic resolution than two genes.

2.1.4 Paucity of Mitochondrial Genomes for *Isopoda*

The isopod family *Sphaeromatidae* is relatively poorly represented on a mitochondrial genomic scale compared to many other isopods. In the NCBI nucleotide database, there is only a single mitochondrial genome for a sphaeromatid isopod - *Sphaeroma serratum*. While there are 946 mitochondrial gene accessions for various genes across this family. By comparison, there are 62 mitochondrial genomes found for all of *Isopoda*, with almost half belonging to the suborder *Asellota* and just over a quarter to *Oniscidea*.

2.1.5 Objectives

The goals of this study are to 1) investigate the use of methods for mitochondrial enrichment that may aid in genome assembly and, 2) assemble the mitochondrial genome of a New Zealand sphaeromatid isopod - *Isocladus armatus*. Specifically, I will compare MDA and a differential centrifugation approach to whole genome sequencing for mitochondrial assembly. These results will help us determine the presence or absence of the 'atypical' mitochondrial genome in this species and enable further understanding into the evolution of this structure. Additionally, sequencing the mitochondrial genome will facilitate future work to elucidate the relationships between species within the *Exosphaeroma* clade (i.e. the *Sphaeromatid* clade containing *Isocladus armatus*).

2.2 Methods

2.2.1 DNA Extraction

Due to extremely high nuclease levels in isopods, DNA was extracted using a modified protocol for the Qiagen DNEasy Blood and Tissue kit, modified from (Wells & Dale, 2018). Cephala were removed from each sample of *Isocladus armatus* using an ethanol and flame sterilized blade, before being homogenized in a 65°C solution of 22 µl 20% SDS and 178 µl of 0.5M EDTA (pH 8). 10 µl of Proteinase K was then added and the solution was incubated at 65°C at 800 rcf for 24 hours. After 24 hours, 400 µl of Qiagen Buffer ATL was added, alongside another 15 µl aliquot of Proteinase K and incubated again for 24 hours. After this, 400 µl of Buffer AL and 400 µl of 100% Ethanol was added to each sample and vortexed for 20 seconds. The solutions were passed through the spin column in two steps due to the large volume. The column was then washed according to the manufacturer's protocol, before a three step elution took place. Each elution consisted of 50 µl of nuclease-free water

incubating on the column for 15 minutes at room temperature before centrifugation at 8,000 rcf, resulting in a total of 150 µl of eluate.

Samples were concentrated in an Eppendorf Vacuum Centrifuge at 30°C, to reduce total volume to approximately 50 µl. At this point the concentration was determined by using 2 µl of DNA in a Qubit fluorometer assay.

2.2.2 Enrichment

Multiple Displacement Amplification

Multiple Displacement Amplification (MDA) was performed using the REPLI-g Mitochondrial DNA Kit (Qiagen catalog 151023). I followed a modified protocol based on Wolff et al. (2012). 100 ng of DNA was added to a PCR tube and the volume adjusted, using nuclease free water, to 20 µl. An amplification mix was made consisting of 1µl of each primer (at a concentration of 10 µM) and 27 µl of the REPLI-g Reaction Buffer. The primers included in this reaction were 16S2 and Crust-16Sf (Table 2.1). This mix was then added to the DNA and mixed. The reaction was briefly spun down and incubated at 75°C for 5 minutes before proceeding. The samples were then allowed to cool to room temperature, 1 µl of REPLI-g Midi Polymerase was added and samples were then mixed and spun down. The reactions were incubated for 16 hours at 33°C, then heated to 65°C for 3 minutes to deactivate the polymerase.

A reaction was classified as successful if I observed significant increases in concentration of double stranded DNA (as measured by a qubit assay), and the observation of a high molecular weight band on a 1% agarose gel.

Testing for specific enrichment of mtDNA after Multiple Displacement Amplification

To quantify if mitochondrial DNA had been successfully enriched, I performed a qPCR assay using the primers described in Table 2.1. Because MDA takes place at a low temperature, there is the possibility that even very specific primers may cause non-specific amplification. Therefore, I performed a qPCR on the enriched samples to assess whether or not mitochondrial DNA was successfully enriched using Crust-16Sr and 16S2 primers (Table 2.1).

Primer Design and qPCR

Primers were selected from the literature based on previous use within *Isopoda*. Specificity of primers was then determined by aligning the primers to the mitochondrial genome of *Sphaeroma serratum* in the software Geneious 9.1 (Drummond et al., 2011). Primers were ordered from Integrated DNA Technologies (IDT), with the last 3 nucleotides of the 3` end being modified to incorporate a phosphorothioate backbone (this prevents endonucleases present in MDA reactions from destroying single stranded primers).

These primers are described in Table 2.1. Additional checking of specificity of the 16S primers was conducted through PCR and then subsequent Sanger sequencing (performed by Macrogen inc). The PCR reaction was performed using DreamTaq hotstart master mix, following the manufacturer's protocol, with a 5 minute hotstart step at 95°C, 40 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 2 minutes, and then a final extension at 72°C for 5 minutes.

Table 2.1 Description of primers used in this study.

Primer Name	**Region**	**Orientation**	**Sequence (5` - 3`)**	**Reference**
Crust-16Sf	16S	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
16S2	16S	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002

A range of 10 fold serial dilutions was carried out to dilute the DNA to have a range of concentrations from 4 ng/µl down to $4x10^{-6}$ ng/ul. For these reactions I used 0.5 µl each of 10 µM forward and reverse primer, 1 µl of template DNA, 5 µl of polymerase master mix, and 3 µl of ddH2O. The reaction conditions were 1 cycle at 95°C for 7.5 minutes, and 30 cycles of 10 seconds at 95°C, then 30 seconds at 60°C and were done using the PikoReal Real-Time PCR system.

Finally, in reactions where mitochondrial DNA appeared to have been successfully enriched using MDA, I used nanopore DNA sequencing to sequence the enriched sample. Amplified DNA that had been enriched using MDA was cleaned up using SPRI paramagnetic beads

(AmPURE XP beads, Beckman Coulter) to remove enzymes and primers from the MDA reaction. Then 80 µl of AMPure XP beads were added to the DNA and mixed through pipetting (1.8:1 ratio of beads to amplified DNA), incubated for 5 minutes at room temperature, and then placed on a magnet. The supernatant was removed and discarded and 500 µl of fresh 70% ethanol was added and then removed. This step was repeated with 200 µl of fresh 70% ethanol which was also removed immediately after addition. The beads were allowed to dry for 30 seconds, and then the tubes were removed from the magnet. 25.5 µl of nuclease free water was then added, and the beads were mixed back into solution through flicking. This was then incubated for 10 minutes at 37°C, and then spun down briefly. The beads were pelleted on the magnet and the supernatant removed into a clean DNA Lo Bind Tube. 3 µl of NEBuffer 2 and 1.5 µl of T7 Endonuclease 1 were then added, the sample was then mixed through pipetting and incubated at 37°C for 15 minutes. After this 20 µl of nuclease free water and 35 µl of AMPure XP beads were added and then incubated on a Hula mixer for 20 minutes at room temperature. The sample was then spun down and placed on a magnetic rack and incubated for 1 minute. The supernatant was removed and discarded. With the sample still on the rack, I added 200 µl of 70% ethanol and then immediately removed. This step was then repeated. The beads were allowed to dry for 30 seconds and 10 µl of nuclease free water was added. The sample was removed from the magnetic rack, flicked to resuspend the beads, and incubated at 37°C for 15 minutes. The beads were pelleted on the magnet, and the colorless supernatant was removed and carried forward into the library preparation step for sequencing. See section 2.2.3 for details.

Differential Centrifugation

Fifteen samples of *Isocladus armatus* were collected from Browns Bay, Auckland, with no preference to colour morph or sex, but solely based on being >8mm. The cephala were removed to minimize DNA from endogenous bacteria. The cephala were pooled, and DNA extracted by homogenizing them with a Potter-Elvehjem homogenizer in 8.5 mls of a pre-chilled buffer consisting of 0.25 M sucrose, 10 mM EDTA, and 30 mM Tris-HCl (pH 7.5) (Tamura & Aotsuka, 1988). The homogenate was split into four 2 ml centrifuge tubes and centrifuged at 1000 g for 1 minute at 4°C, and the supernatant was then transferred to clean 2 ml tubes and centrifuged again, this was repeated for a total of 3 centrifugations. The supernatant of the last centrifugation was then transferred to new tubes and centrifuged for 15 minutes at 14,700 g at 4°C again. This protocol was modified from Macher et al. (2018).

The enriched mitochondrial pellets were then extracted according to the Promega Wizard Genomic DNA Purification Kit protocol for mouse tails, with the following modification: the

lysis was carried out at 65°C (as opposed to 56°C), and the DNA was rehydrated in 25 µl, giving a total of 100 µl of enriched mitochondrial DNA (25 µl per tube).

The presence of purple proteins or pigments co-precipitating with DNA, indicated the need for further purification steps. The Qiagen Gel Extraction kit was used with the following modifications and was performed without running the DNA on a gel. 25 µl of DNA was added to 300 µl of Buffer QG and 100 µl of Isopropanol and mixed by flicking. The solution was then passed through a Qiagen spin column, and washed with 750 µl of Buffer PE, and centrifuged. The DNA was finally eluted in a 3 step elution, using 50 µl of nuclease-free water at each step, with centrifuged preceded by a 5 minute room temperature incubation. This was repeated for each tube in order to process the full 100 µl of mitochondrial enriched DNA)

Testing for specific enrichment of mitochondrial DNA after differential centrifugation

In order to assess enrichment of mitochondrial DNA using the differential centrifugation method, I ran samples on a 1% agarose gel for 90 minutes at 90 mV. I considered enrichment successful if at least two DNA bands were present on a gel - as I expected one band to be residual nuclear DNA, and one band from the mitochondria.

Sequencing of Enriched Samples

For both MDA and differential centrifugation samples, DNA was sequenced using an Oxford Nanopore MinION, with a Rev D flow cell. Because relatively low read numbers are required to assemble a mitochondria if a sample was successfully enriched, initial sequencing was done using a pre-used and washed flow cell. Samples were prepared for sequencing using the SQK-RBK004, or rapid barcoding kit, and where possible I used barcodes that had not previously been used with the flow cell. Flow cells were sequenced for 48 hours, and the reads base-called using Guppy 3.1.5 and de-multiplexed using qcat 1.1.0.

I considered enrichment successful if mitochondrial reads were at least an order of magnitude higher in proportion than the whole genome sequencing. Mitochondrial reads were identified by aligning reads to the mitochondria assembled in 3.3, using minimap2 (H. Li, 2018). Relevant reads were extracted from sequenced reads by 1) de-multiplexing and selecting reads from the used barcodes, and 2) if the barcodes had been used previously on the flow cell, I used minimap2 to align the reads against the reference genome for the

previous barcode. Reads that failed to align were then extracted and treated as relevant reads.

2.2.3 Whole Genome Sequencing

Using the same DNA extraction protocol as for the MDA reaction above, outlined in section 2.2.1, I sequenced whole DNA extractions of the same individual (BBFD2), using both the Rapid Barcoding Kit (SQK-RBK004) and the Ligation Sequencing Kit with barcoding expansion kit (SQK-LSK109). Sequencing was performed with MinION Rev D flow cells, using MinKNOW v19.05.0. The resulting data was then basecalled using Guppy 3.1.5. The base-called files were then de-multiplexed, barcodes removed, and chimeric reads separated using qcat 1.1.0.

2.2.4 Mitochondrial Assembly and Annotation

The mitochondrial genome was assembled using Flye 2.4.2 (Kolmogorov, Yuan, Lin, & Pevzner, 2019), from whole genome shotgun nanopore data (section 2.2.3). The plasmid and metagenome assembly options were used, alongside an estimated genome size of 2 GB.

The resulting contigs were then mapped to the *Sphaeroma serratum* partial mitochondrial genome and the mapped contig was then considered the 'mitochondrial' bait. Because of the potential for major structural variations, as well as potentially very high nucleotide divergence between *S. serratum* and *I. armatus,* this initial contig may not be reliable. Instead I used this initial contig in a procedure similar to that described by Hahn et al. (2013). This method uses the initial contig and aligns all other reads to it. I then extracted the mapping reads and used them downstream as our mitochondrial read pool.

In order to assemble the unit, I used Flye with the mitochondrial read pool already identified, with an estimated genome size of 14kb and a minimum overlap of 8000bp. The reasoning behind this was that a very conservative minimum overlap would resolve either the dimer or the unit with confidence. This assembly was then annotated using MITOS2 (Bernt et al., 2013) with modified settings to allow for increased overlaps in genes (modified setting was increasing the maximum overlap between features to 100 bp, from 50 bp), as isopod mitochondria have previously been reported to have significant compaction due to overlapping genes. This entire approach was also repeated to assemble the dimer, however the requirement for a minimum overlap was dropped to 1000bp.

In order to assess the accuracy of the unit, I aligned all Illumina reads obtained from DArTseq (see chapter 2) to this unit. Illumina reads are highly accurate and enable us to assess the quality of our contig and to identify whether errors are the result of mismatched bases, or indels. To do this, I used Geneious 9.1 to identify variants that could appear as as SNPs or indels.

Assessment of Self-Similarity

In order to assess whether the mitochondria exhibited atypical structure, I mapped all reads back onto the circular mitochondrial chromosome using minimap 2.11, to identify all mitochondrial reads. These reads were then extracted using samtools 1.9 (H. Li et al., 2009), and assessed for self-similarity using the fastx-rlength.pl script from David Eccles bioinformatics script repository (http://dx.doi.org/10.5281/zenodo.556966). This tool identifies single sequences that contain both a k-mer and its reverse complement, which accordingly indicates that the read contains two regions with self-complementarity. K was set to 17, and a minimum reverse k-mer count of 10 was implemented, and a maximum of 1800. The resulting selection of reads was manually curated in Geneious 9.11 (Drummond et al., 2011), removing reads that exhibited a self-self dotplot indicating highly repetitive erroneous sequences.

The curated selection of reads were then aligned back to the mitochondrial genome, and the position of alignments were then used to assess the likelihood of an atypical mitochondrial genome structure. This was done by identifying reads mapping across the junction between genome units. If atypical structure is present, I would expect to see two peaks in coverage of inverted reads across these junctions.

Additionally, I split the assembled mitochondrial dimer into two halves, each consisting of one copy of the ‘unit’. The reverse complement of the one sequence was then assessed for similarity against the other, using the EMBOSS dottup program using a window size of 10 (Rice, Longden, & Bleasby, 2000).

Identifying Presence of the Monomer

If a linear chromosome is present, I would expect it to be difficult to identify based on sequencing data alone, as it would be virtually identical to the dimer.

To test this hypothesis, I used Tombo 1.5 to generate the average standard sequence signal per position for both junctions. Previous research has suggested that a hairpin forms at the origin of replication, in the CYTB junction (Peccoud et al., 2017). Thus I would expect to see evidence for a hairpin in the sequence signal within the CYTB junction.

Additionally, I plotted a histogram of the distribution of sequencing speeds for our pool of self-similar reads, and a pool of random non-mitochondrial reads. I tested for a difference between the distributions of speed for a random sample of 10,000 of non-mitochondrial reads, and the distributions of speed for the self-similar mitochondrial reads (identified above in section 2.4.1). This was done using an alternative hypothesis that the Cumulative Distribution function of the random reads was greater than that of the mitochondrial self-similar reads. I conducted this test in R 3.5.2, using a Kolgorov-Smirnov test.

Structural Inference

Because both sides of the dimer were almost identical (with the exception of a low coverage region), I used the assembly of the mitochondrial unit with Medaka 0.7.1 to identify variants within this unit, with the hypothesis that the differences between each unit of the dimer should appear as SNPs. Additionally, because of inconsistent placement of the origin of replication, I confirmed its position through a GC-Skew analysis. This was done by calculating GC-Skew using GenSkew ("GenSkew - visualization of nucleotide skew in genome sequences," n.d.), using a step size of 20, and a window size of 100. This method is able to predict the origin of replication (*ori*) based on shift points in GC skew (Arakawa, Suzuki, & Tomita, 2009).

2.3. Results

2.3.1 Enrichment

Initial assessments for mitochondrial enrichment appeared positive for MDA and differential centrifugation. Mitochondrial-enriched DNA from three collections of *I. armatus*, from Browns Bay, Auckland, using multiple strand displacement was assayed using qPCR (Fig. 2.3) and sequencing, while enrichment using differential centrifugation was assayed through gel electrophoresis and sequencing.

qPCR

qPCR showed significant enrichment of 16S rDNA for both samples that had been subjected to MDA relative to genomic DNA. As expected, the amount of 16S rDNA decreased with increasing dilutions, with all dilutions indicating more 16S rDNA in MDA reactions than in control (non-MDA treated) genomic DNA (Fig. 2.3).

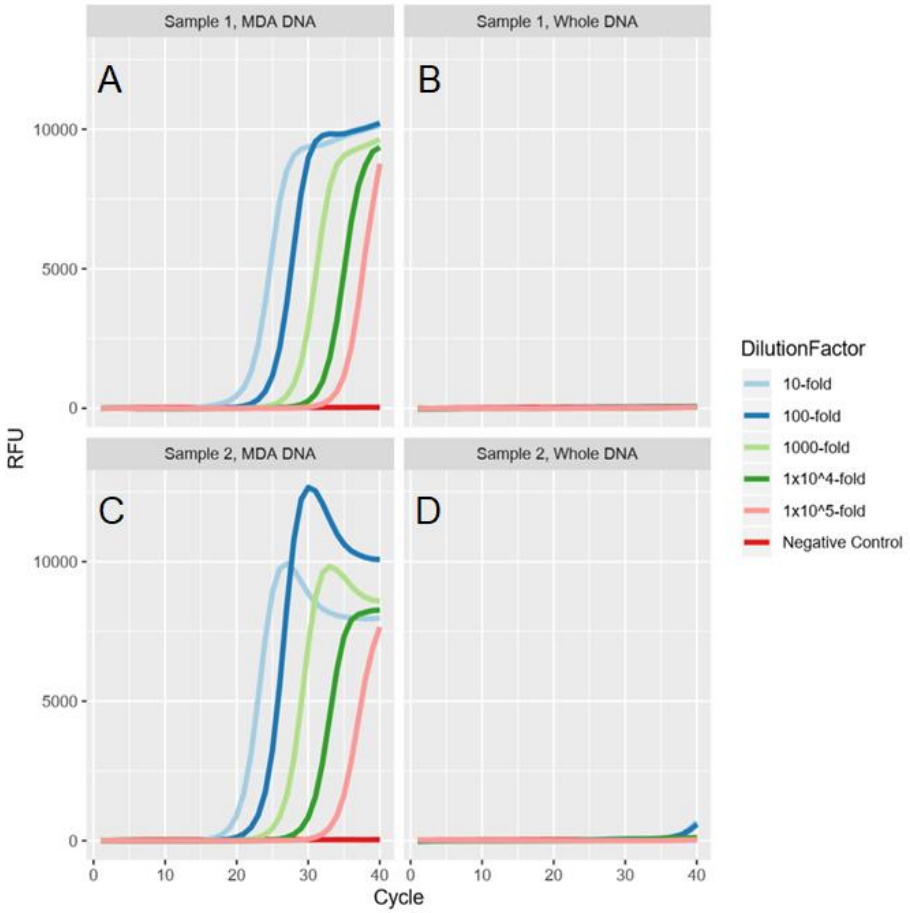


Figure 2.3 qPCR amplification curves for each sample and the controls. Plots A and B represent the same sample DNA, but plot A has been subject to multiple displacement amplification (MDA), while B has not. Similarly plots C and D represent the same sample, but C has been the subject of MDA. RFU (on the Y axis) means Relative Fluorescence Units, the higher this value the greater the amount of double stranded DNA in the sample.

qPCR results are also summarised in Table 2.2, where the negative controls all showed no amplification for 16S rDNA, and the Cq values for the amplified samples were all much lower than those of the whole DNA, indicating successful enrichment. Cq values represent the quantitation cycle, or the cycle at which fluorescence is detected, the lower this value, the more double stranded DNA is present in the sample.

Table 2.2 Cq values for each qPCR reaction of non-amplified and amplified (using Multiple Displacement Amplification), alongside their subsequent dilutions (D1-D5). D1 is a 10-fold dilution D2, is a 100-fold dilution, D3 is a 1000-fold dilution, D4 is a $1x10^4$ fold dilution, and D5 is a $1x10^5$ fold dilution. Unamplified samples represent the results of the sample without the Multiple Displacement Reaction, while Amplified samples are the same samples but post amplification. Letters in brackets in the sample IDs refer to the samples in Figure 2.3)

	Negative Control	D1	D2	D3	D4	D5
Sample 1, Unamplified (A)	-	-	-	-	-	-
Sample 1, Amplified (B)	-	18.24	22.24	25.75	29.52	32.27
Sample 2, Unamplified (C)	-	37.48	37.62	-	-	-
Sample 2, Amplified (D)	-	17.73	20.62	23.74	27.71	31.66

MDA based enrichment appeared initially successful as I saw increased abundance of mitochondrial DNA (as based on presence of 16S amplicons) relative to the original sample. This was supported by consistent amplification across serial dilutions of the MDA-enriched DNA but not the original sample (Fig. 2.3). Additionally, the Cq values for amplified DNA were much lower than for the unamplified DNA (Table 2.2).

However, sequencing (see methods section 2.2.3) of sample 1 indicated very little mitochondrial DNA, as I only observed 10 mitochondrial reads - representing 0.007% of all of the sequences (Table 2.3). Analysis of all sequences indicated that there were 164 18S rDNA sequences, potentially the result of mis-amplification during the MDA.

Differential Centrifugation

Differential centrifugation appeared initially successful as I observed two fluorescent bands on an agarose gel (see Figure 2.4). This is the result that would be expected for successful enrichment. However, subsequent nanopore sequencing did not support this result and it appeared as though this enrichment failed (Table 2.3), as only 0.017% of reads appeared mitochondrial in origin. Additionally, repeated attempts at enrichment and extractions

showed limited reproducibility, with only a few of the attempted extractions appearing successful (as indicated through gel electrophoresis).

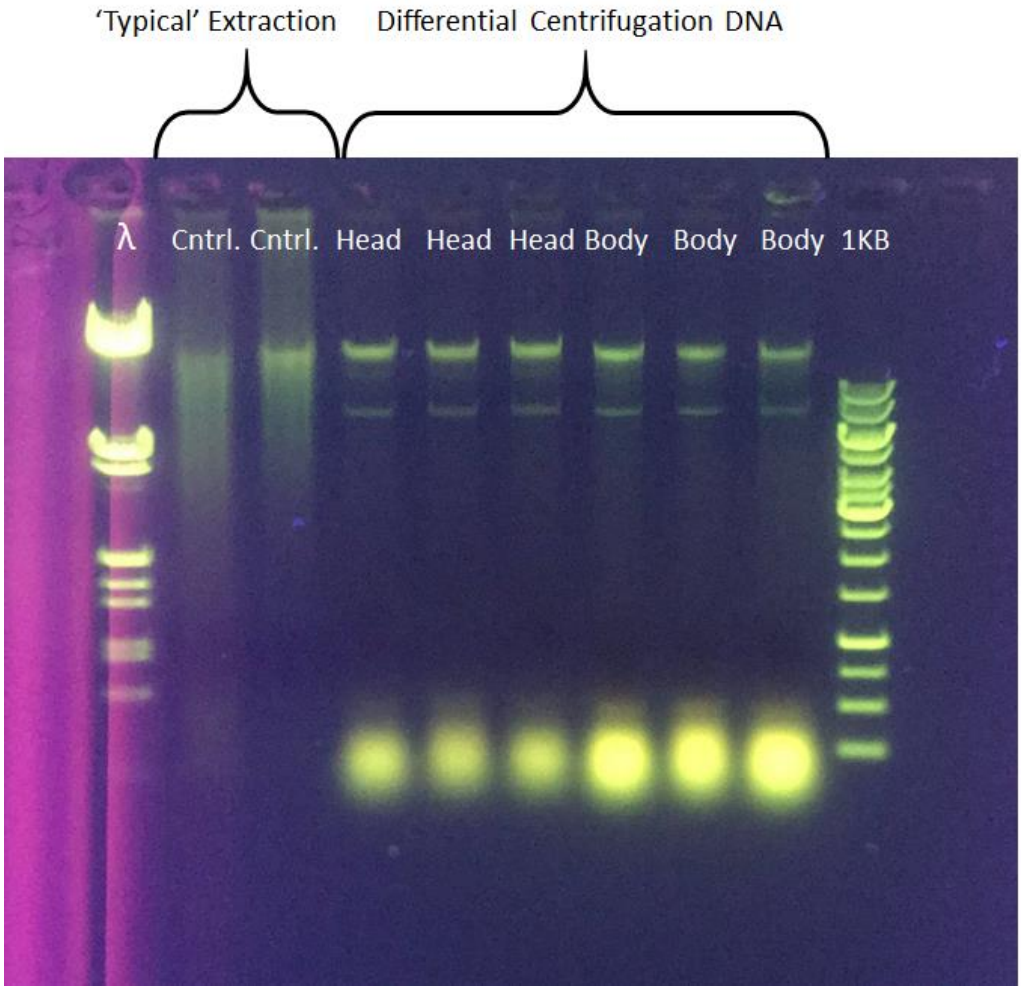


Figure 2.4 - Differential centrifugation DNA extraction results of *Isocladus armatus*. Typical indicates the extraction method described in section 2.2 (modified qiagen DNA extraction protocol), while 'Head' and 'Body' indicate differential centrifugation DNA extractions performed on either the cephala or body of *Isocladus armatus*. The left ladder is the lambda ladder, while the right ladder is the Thermofisher 1kb+ ladder.

Nanopore sequencing indicated that both methods of enrichment, differential centrifugation and multiple strand displacement, appeared to fail. Differential centrifugation returned a mitochondrial proportion within the same order of magnitude as the whole genome sequencing, while the MDA reaction returned a mitochondrial proportion two orders of magnitude below the shotgun sequencing. These patterns suggest a reduction, rather than enrichment, in abundance of mitochondrial DNA.

Table 2.3 - Assessment of enrichment success using the number and proportion of sequences obtained for each mitochondrial extraction process. Relevant reads are classified as reads from the sequencing reaction that most likely originate from the sample of interest (as opposed to reads from samples previously used on the flow cell).

Method of Enrichment	Number of Reads	Number of Relevant Reads	Number of Mitochondrial Reads	Proportion Mitochondrial reads
Differential Centrifugation	29,187	29,187	51	0.0017
Multiple Displacement Amplification	1,040,000	142,632	10	0.00007
Whole Genome Shotgun Sequencing	2,215,224	2,215,224	3032	0.0013

2.3.2 Assembly Results

Dimer Assembly

There were 467 mitochondrial reads within the initial read pool, with a median length of 2,893bp, min length of 209bp, and max length of 19,526bp. Using Flye, an assembly was generated from this initial read pool. Using this assembly, I identified a total of 3,032 putative mitochondrial reads. This read pool had a minimum length of 88bp, a maximum length of 19,526bp, and a median length of 1,418bp. *De novo* assembly of these reads yielded a single contig that was 28,766bp in length.

Within this assembly there was extensive variation in coverage, with the highest coverage found over the second half of the monomer. A maximum coverage of 331 and a minimum coverage of 2 was observed (Fig. 2.5).

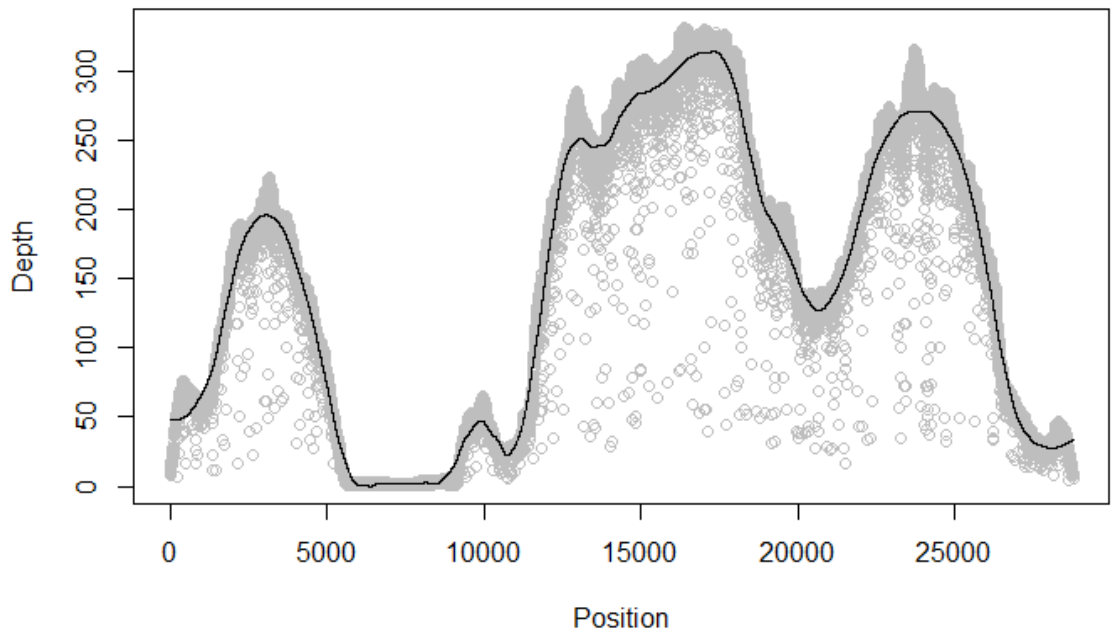


Figure 2.5. Coverage plot of *de novo* assembly of *Isocladus armatus* mitochondria. Gray dots indicate depth at each specific position, while the black line indicates the smoothed (loess, span of 0.1) coverage. Position indicates distance (in base pairs) away from the centre of the 12S junction

This assembly consisted of two units found as inverted repeats, as shown in Figure 2.6. The MITOS annotations of the genes indicate duplication and presence as inverted repeats, with junctions at the 12S and CYTB regions.

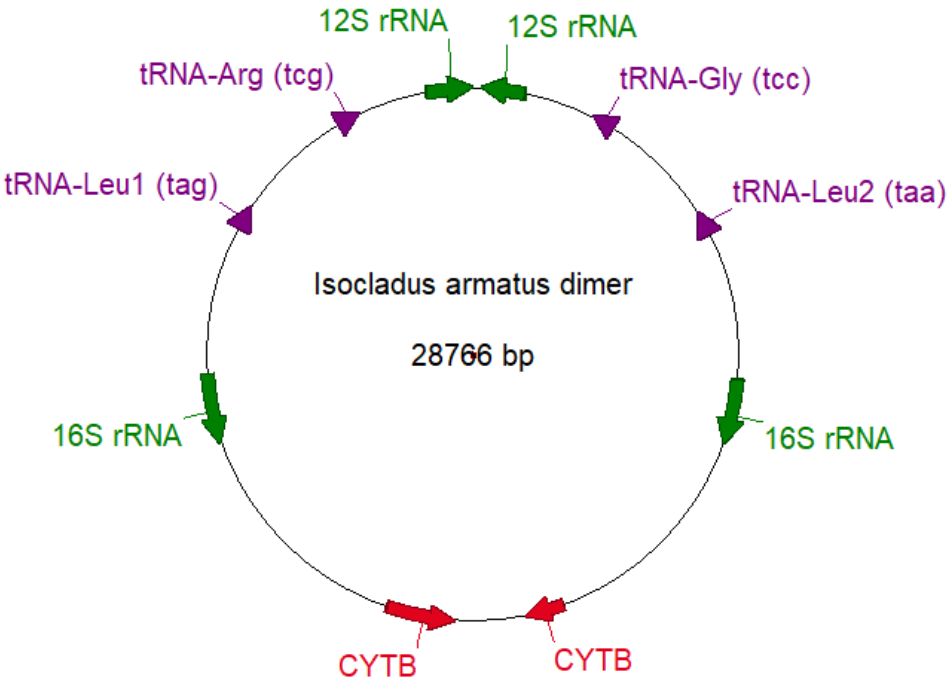


Figure 2.6 Annotation of rRNAs, junctions, and heteroplasmic tRNAs in *Isocladus armatus*. Directionality is indicated by direction of arrows, these annotations suggest duplication of the units as inverted repeats.

I hypothesize that the dimer is likely misassembled around the CYTB junction as I observed that this region is hampered by very low coverage relative to the rest of the dimer, and the annotations in this region are not congruent with those observed on the raw reads that map across the junction (Supp. Fig. 2-4). I predicted that the origin of replication (*ori*) occurs within this junction because this is where it is found in the raw reads, as well as in other species of isopod (Doublet et al., 2013). However, instead I found an *ori* 600 bp downstream outside the junction and between copies of the ND5 and ND4 genes (Supp. Fig. 1).

I find that each side of the dimer is almost identical to the other (Fig. 2.7), there are a relatively large number of mismatches between sides of the dimer, likely resulting from assembly errors, or potentially from neutral substitutions present on one side of the dimer. However, some of these mismatches are associated with indels in homopolymeric regions (regions where tandem repeats of the same 1 or 2 bases occur), and heteroplasmic sites associated with changes in tRNA. This is shown above in Figure 2.6, where two of the three heteroplasmic sites identified by Peccoud et al (2017) are found.

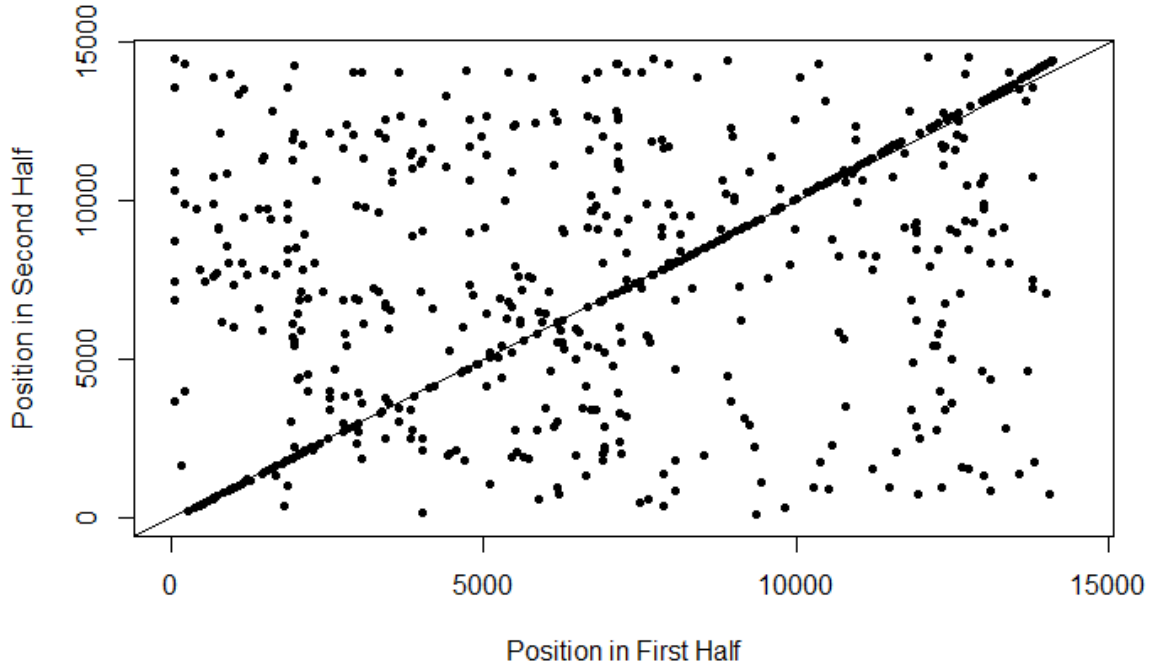


Figure 2.7 Dotplot of each half of the mitochondrial dimer, with the second half reverse complemented to assess for self-similarity. A 1:1 relationship would indicate a perfect match, (as indicated by the black line). The large amount of variation in the relationship is likely to be the result of assembly errors arising the high error rates of nanopore sequencing. Alternatively, it is possible that these are real differences due to neutral substitutions present on only one side.

Unit Assembly

I assembled a linear contig that was 14,435 bp long, with a mean coverage of 299x (Fig. 2.8). The 16S region of this assembly was 98.9% identical to the reference 16S sequence of *Isocladus armatus* (differences appear associated with 5 homopolymeric regions). This linear unit was almost identical to one side of the dimer, with differences largely appearing to be errors as a result of indels (insertions or deletions of one or more bases) in homopolymeric regions, as well as mismatches due to the likely misassembly of the dimer.

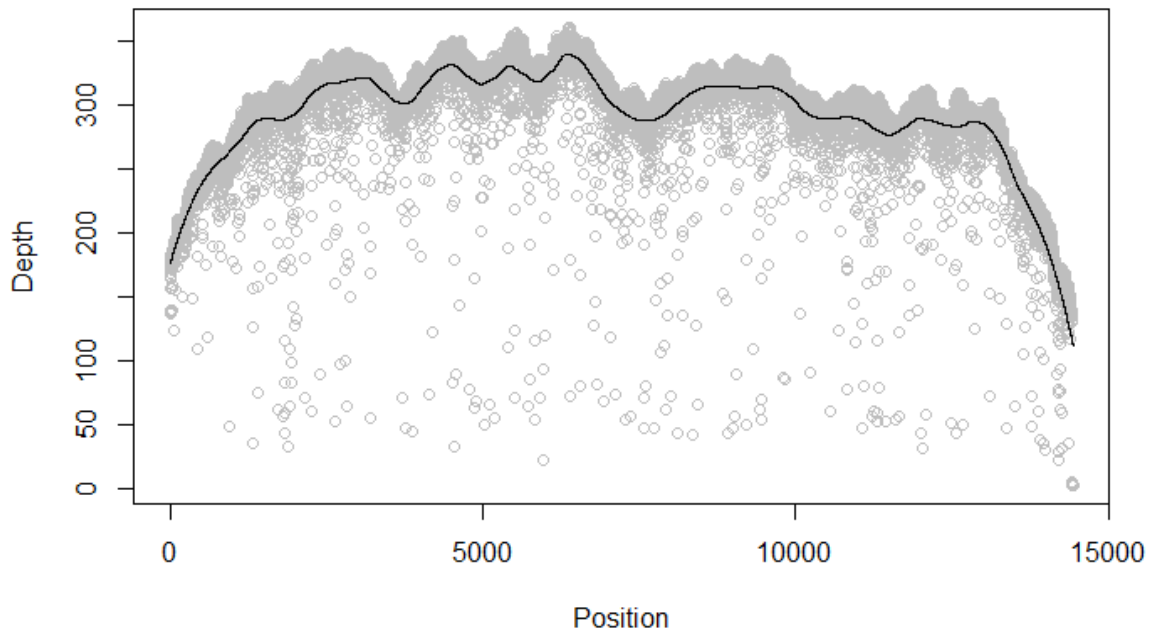


Figure 2.8 Coverage plot of the mitochondrial unit assembled through de-novo assembly with minimum overlap of 8000bp. Black line indicates smoothed coverage (loess, span of 0.1) , grey points indicate coverage at each specific position. Position represents distance (in base pairs) from the start of the 12S gene.

The cumulative GC skew graph (Fig. 2.9) showed an increase in GC skew across the monomer. Using this information, GenSkew predicted that the origin of replication for this assembly was at position 14,197 based on the point of lowest cumulative GC skew.

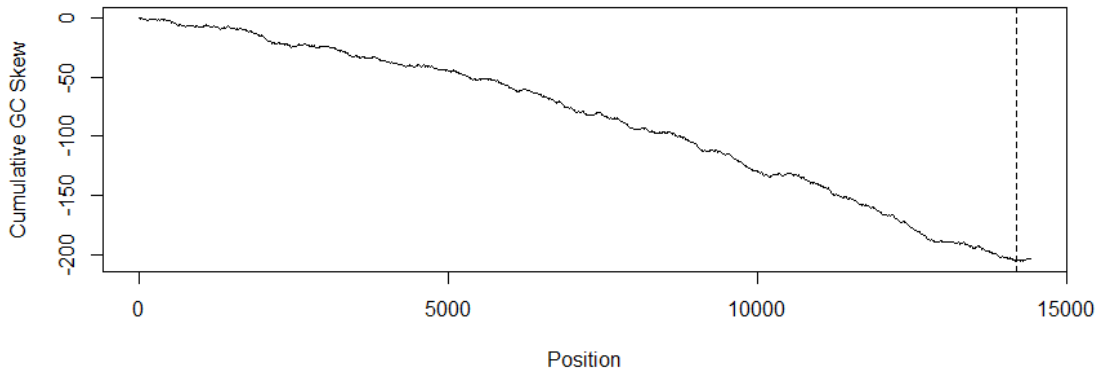


Figure 2.9 - Cumulative GC skew for the mitochondrial unit, using a step size of 20, and a window size of 100. The dashed line indicates the proposed position for the origin of replication (14,197), while position represents distance from the start of 12S gene (in base pairs)

Assembly Errors

I observed internal stop codons present in the COX3, COX1, ND4L, ND3, ND2, ND6, and ND5 genes. Additionally, I observed numerous translational exceptions associated with these genes. This results in many of these genes appearing as duplicated or split in the annotation. This is likely due to the presence of many erroneous indels, resulting in what appear to be frameshifts, but what are instead likely the result of errors in the assembly. However, overall the annotations for the mitochondrial unit are relatively consistent with a 'typical' mitochondrial gene arrangement (Fig. 2.10).

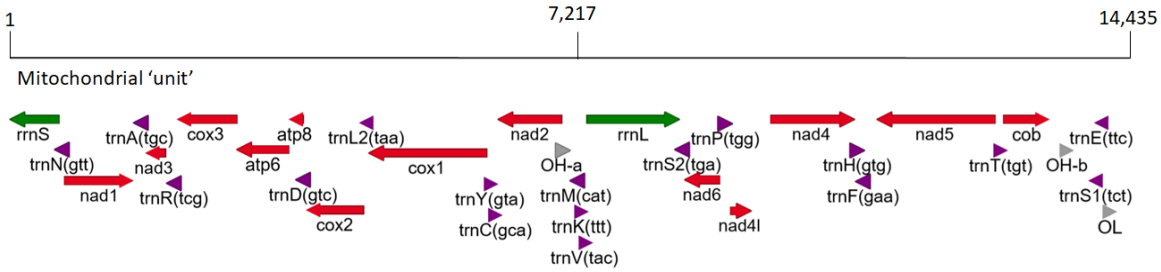


Figure 2.10 - Annotation of the mitochondria 'unit'. Heteroplasmic site information is not incorporated into this figure. The top line indicates the distance from the start of the 12S gene (in base pairs)

The DArTSeq data had high coverage distributed over approximately half of the mitochondrial unit, and zero coverage over the rest. Geneious conservatively identified 22 potential errors, 16 of which were indels associated with homopolymeric regions. The 6 other errors were SNPs - 1 of which was associated with the previously identified SNPs (see section 3.3.1). The other SNPs were present at low frequency and may be the result of real SNPs due to population differentiation - as I pooled all DArTSeq data for this analysis. This is not surprising given that DArTSeq intentionally targets polymorphic regions.

Missing Features

The following tRNAs were absent from this assembly of the unit - tRNA-Trp, tRNA-Gly, tRNA-Leu1, and tRNA-Ile. However, tRNA-Leu1 and tRNA-Gly are present on the dimer assembly, and are encoded by a heteroplasmic site (these appear as SNPs in this assembly; see section 3.3.1), while tRNA-Trp is located in the 12S junction on the dimer. As a result, the only feature that appears to be truly absent is tRNA-Ile. Additionally, this tRNA is also absent in the three annotated raw reads, suggesting it is likely absent from the mitochondrial genome rather than a result of misassembly (Supp. Fig. 2-4).

2.3.3 Evidence for Atypical Structure

Variant Identification

Medaka identified 3 SNPs present in the unit. These SNPs had relatively high support in terms of depth, with the first two SNPs present at an almost 50:50 ratio, while the third was present at a 1:3 ratio (Table 2.4).
The first two SNPs have previously been observed within *Isopoda* and each results in a change in tRNA (Peccoud et al., 2017), while the third SNP has an unknown function and has not been previously reported.

Table 2.4 - Position and function of identified variants within the mitochondrial unit, * indicates opposite strand

Position	**Reference: Alternative**	**Function**	**Depth (number of reads)**	**Frequency ratio**
2120	C:G	tRNA-Gly->tRNA-Arg	C:154,G:135	C:0.53, G:0.47
4608	T:C*	tRNA-Leu1>tRNA-Leu2	T:182,C:133	T:0.58,C:0.42
14,411	A:G	Unknown	A:46,G:118	A:0.28,G:0.72

Dimer

I found reads that were longer than the expected size of a ‘typical’ mitochondria, and at least 148 reads that spanned the proposed junctions between copies of the unit in the dimer. Additionally, for self-similar reads (sequences containing a k-mer and its reverse complement), I saw two peaks in coverage: one over each junction, with lower coverage found over the rest of the mitochondria (Fig 2.11). These peaks in coverage corresponded to a max coverage of 73, and 75 for the CYTB and 12S junctions respectively. A total of 277 reads containing inverted k-mer counts greater than 10, and less than 1800 were found and used in this analysis.

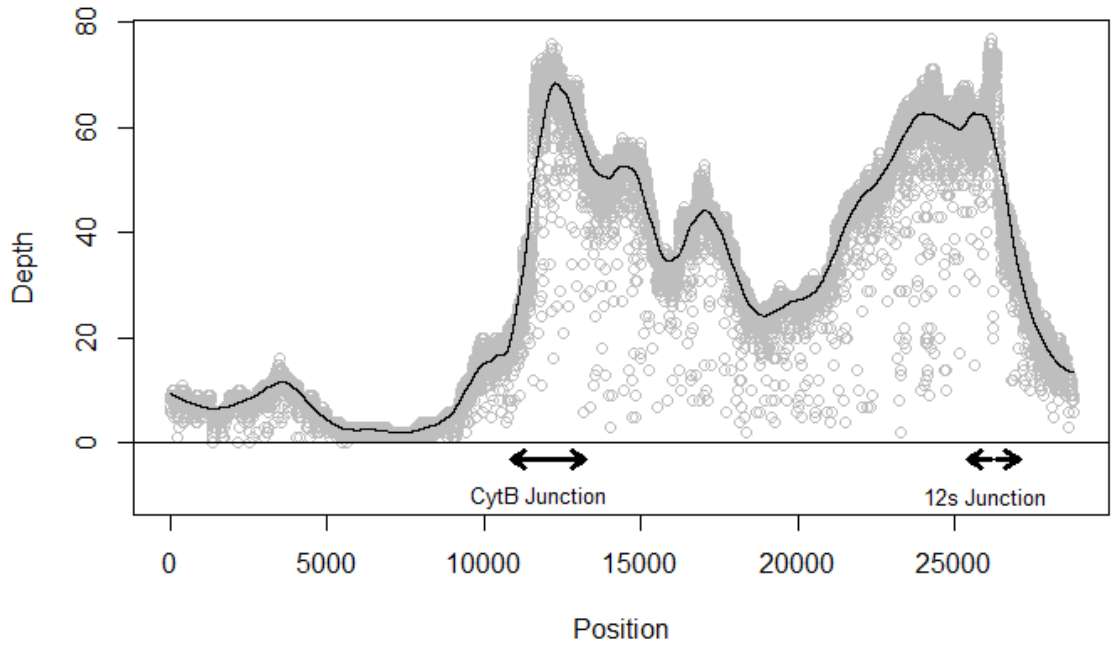


Figure 2.11. Coverage of reads containing inverted K-mers across the *de-novo* assembly of the mitochondrial dimer. Arrows indicate the position of the junctions between copies of the mitochondrial unit.

Monomer

The Kolgorov-Smirnov Test tests for a difference between distributions, this identified a difference in the distribution of sequencing speeds between the mitochondrial reads containing inverted K-mers, and the sequencing speed of non-mitochondrial DNA. This result suggests that these mitochondrial reads have a lower CDF (Cumulative Distribution Function) than the random non-complementary reads ($D^{+} = 0.26$, $p < 0.001$). I observed a normal unimodal distribution for speed with non-reverse complementary reads, while I observed a bimodal distribution for reverse complementary reads. For sequencing speed distributions I saw a peak at approximately 360 events per second in both non-reverse complementary and reverse complementary sequences, while in the reverse complementary reads I saw a second peak at 330 events per second (Fig 2.12) .

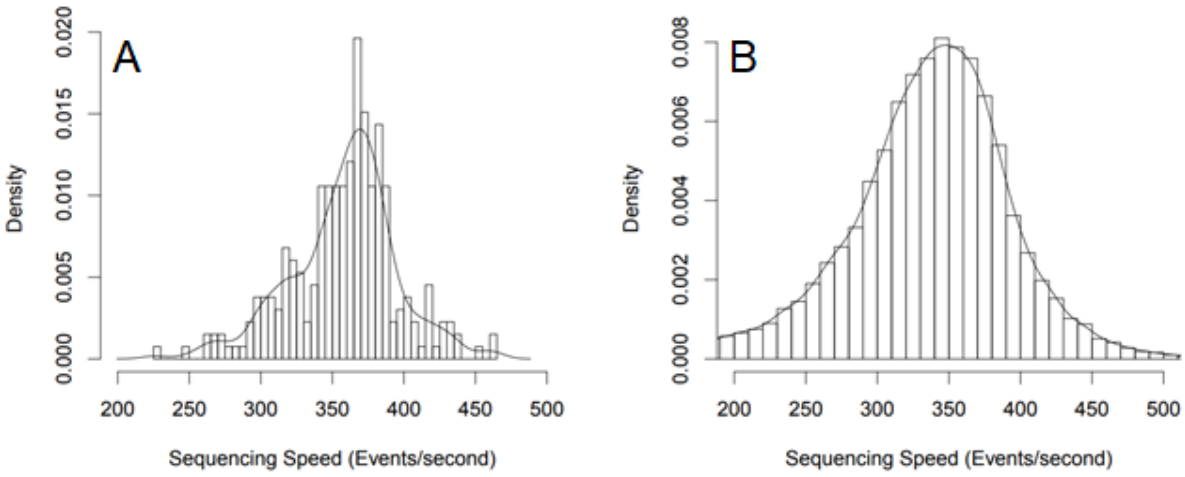


Figure 2.12 Distributions of sequencing speeds for reverse complementary mitochondrial reads (A) and non-reverse complementary sequences (B)

Additionally, I found evidence for regular stalls within the CYTB junction, but not the 12S junction. This suggests that a hairpin may be present in the CYTB junction, but not the 12S junction - consistent with our prior expectations. This is supported by a reduction in signal extremes observed at the beginning of the CYTB junction (Fig 2.13). This pattern has previously been suggested as the result of sequencing stalls arising from hairpins during sequencing (White, Pellefigues, Ronchese, Lamiable, & Eccles, 2017).

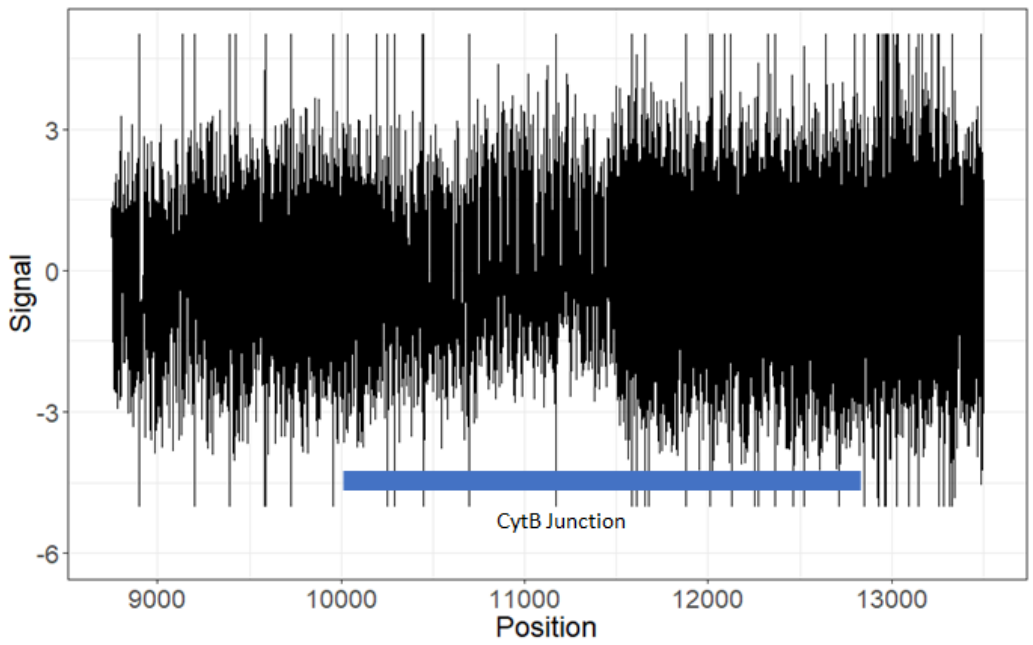


Figure 2.13 Standardized sequencing signal generated from the nanopore aligned to the CYTB junction and surrounding regions. A large drop in signal variance is observed at the beginning of the junction associated with stalling in the pore.

2.4. Discussion

2.4.1 Failure to Enrich Mitochondrial DNA

Despite initial results suggesting otherwise, I found that two different methods to enrich for mitochondrial DNA failed. These methods were a Multiple Displacement Amplification (MDA) approach, and a differential centrifugation approach. qPCR results using template DNA from MDA enriched samples produced stronger amplification relative to the non-enriched samples. However, sequencing of an MDA enriched sample indicated less mitochondrial DNA was actually present in these samples than in the original. In fact, many of these sequences originated from 18s rRNA within the nuclear genome. Although the primers (Table 2.1) used in the MDA reaction target 16S rDNA, the failure to amplify mitochondrial DNA could be the result of nonspecific primer binding due to the very low temperature (33°C) at which the reaction took place. If these primers bind non-specifically, I would also expect the qPCR to indicate positive results, because I used the same primers for both qPCR and MDA. However, these primers appear to be relatively specific because I have successfully used them to amplify and sequence the 16S gene at an annealing temperature lower than that used in the qPCR. As a result, it is not entirely clear why the MDA failed to enrich for mitochondria.

Differential centrifugation appeared to make little difference to mitochondrial concentrations, despite early results suggesting successful enrichment. The early results suggested enrichment of mitochondria, because of the presence of two bands on a gel, and I interpreted this as a mitochondrial and nuclear band. While the band was not at a size consistent with expectations regarding mitochondria (Fig. 2.4), this could be attributed to supercoiled mitochondrial DNA moving faster through a gel. However, with the mitochondrial structure relatively well resolved through Whole Genome Sequencing, the results remain unclear because I would have anticipated a 14.4kb sized band on a gel as a result of the linear monomer. Although these results were promising, they were not easily replicable, and they required pooling of individuals. Additionally, I cannot rule out the possibility that the

observed bands were due to the presence of a bacterial endosymbiont (such as Wolbachia) or plasmid extracted in conjunction with the isopod DNA.

2.4.2 Accuracy of Assembly

The assembly of the mitochondrial genome as a dimer was inconsistent in coverage and placement of the origin of replication in relation to the position of the CYTB junction (both raw reads and the assemblies of the dimer and unit had differing placements of the *ori*). This region was also associated with low coverage. Therefore, I avoided making substantial inferences on the structure of the mitochondria from the dimer, except in regions of high coverage.

I found that the assembly of the mitochondrial unit was generally accurate relative to both Illumina reads, and a reference 16S sequence, and had consistently high coverage. All identified errors were indels associated with homopolymeric regions. This is an expected result for nanopore assemblies which can often be minimized through additional rounds of polishing (Tyler et al., 2018). However, I found that additional polishing, while reducing the number of indels, also resulted in the removal of some features such as some tRNAs. This is potentially the result of uneven polishing, resulting in what appear to be frameshifts causing movements of the tRNA codons.

The gene arrangement for this mitochondrial genome was also very similar to those previously observed within isopods (Peccoud et al., 2017). However, I found a few notable exceptions. I found that two tRNAs were located at the end of the unit, or within the CYTB junction. The difficulty in amplifying and sequencing the junctions could explain their absence in some previous studies (Doublet et al., 2013; Kilpert et al., 2012).

Although my mitochondrial annotation suggests an *ori* located between the ND2 and 16S genes (Fig 2.10), I suggest instead it is likely a result of misassembly or misannotation. This is because I also found that two *ori*'s were annotated right at the end of the unit, next to the CYTB gene which was the expected location, as this area has previously been reported as the location of the control region in mitochondria for *Isopoda*. This is supported by the GC-Skew analysis, which proposes position 14,197 as the most likely site for the *ori* based on being the lowest point of cumulative GC skew due to be being rich in AT content, or low in GC content which is expected for *oris* (Doublet et al., 2013; Sahyoun, Bernt, Stadler, & Tout, 2014).

2.4.3 Evidence for Atypical Structure

I find evidence that *Isocladus armatus* exhibits an atypical mitochondrial structure that consists of a circular 28kb chromosome and potentially a linear chromosome (see section 2.3.3 for details). This has been observed in other lineages of isopods, but previously thought to be absent within *Sphaeromatidae*. My results were contrary to my initial expectations, as *Sphaeroma serratum* (the most widely studied species within *Sphaeromatidae*) - exhibits typical metazoan mitochondrial structure. However, it appears that the *I. armatus* mitochondrial genome consists of a ~28kb circular dimer made up of two copies of a 'typical' mitochondrial genome, fused together as inverted repeats. There is additional, limited evidence that a linear chromosome consisting of a single copy of a typical mitochondrial genome is also present. The junctions between the units in the dimer are between copies of the CYTB gene, and between copies of the 12S rRNA genes (see Figure 2.6).

Previous research suggests that the linear monomer is non-functional and is the result of self-renaturing of the dimer during replication - resulting in the presence of a monomer (Peccoud et al., 2017). This research predicts that this linear monomer has a hairpin at one end as a result of this self-renaturation. Peccoud et al (2017) confirmed this through PacBio sequencing, by using the presence of a hairpin as a natural SMRT Bell (SMRT Bells are a synthetic hairpin used for sequencing with PacBio that are added to the ends of double stranded DNA).

While techniques have not yet been developed to enable hairpin detection with Nanopore sequencing, I hypothesized that a hairpin should leave an identifiable signal in the raw nanopore sequencing signal data. I predicted that a hairpin would take longer to pass through a sequencing pore, and thus the sequencing speed should show a bimodal distribution - where dimeric reads move faster, while monomeric reads move slower. Additionally, this would show a signal in the sequencing trace (White et al., 2017) - where I expected to see a reduction in variance of the signal within the CYTB junction, as well as evidence for the presence of stalls in this region. This is what I observed, suggesting the presence of both a linear and circular chromosome. While this weakly suggests a linear chromosome is present (or at least a hairpin within the dimer), I cannot make many inferences based off this analysis and instead suggest it warrants further investigation.

I did not expect to observe atypical structure in *Isocladus armatus* due to previous reports that indicated the absence of atypical structure in another Sphaeromatid, *Sphaeroma*

serratum. Despite this, my results are consistent with those that would be expected for a species with atypical structure. I hypothesized, upon initial evidence of this atypical structure, that a large number of reads containing inverted repeats, along with the presence of known heteroplasmies would constitute strong evidence for the proposed structure. In particular, the presence of the following factors strongly supports the likelihood of the proposed atypical structure. 1) a 28kb draft genome that consists of two inverted repeats, 2) More than 148 reads spanning the junctions between units on the dimer (specific number depends on position within junctions), and finally 3) I find evidence that the same SNPs or heteroplasmies found in other isopods possessing atypical structure are present in this species. These three pieces of evidence strongly support an atypical, circular 28Kb mitochondrial structure.

However, it is important to note that there are alternative explanations that could partially undermine the strength of support for atypical mitochondrial structure. The first is if a large number of mitochondria were undergoing replication immediately prior to DNA extraction - in this case we may see sequences that appear to contain inverted repeats. Secondly, if during the preparation of DNA for nanopore sequencing, chimeric DNA strands formed - this could also explain these observations as these occur when two strands of DNA are ligated together during adapter ligation. Chimeric reads are unlikely as they should be removed during the sequencing quality control prior to assembly, and the junctions I observed contain unique sequences rather than consisting entirely of the ends of each unit fused together. Additionally, because more than 150 reads were identified that mapped across the junction, I suggest that it is very unlikely that this result is due to DNA replication prior to extraction.

While our results suggest a 28kb dimeric structure, I cannot completely resolve the structure of the mitochondrial genome due to low coverage in some regions. Because the two units are more or less identical, I was able to collapse the dimer into a single contig that represents the fundamental mitochondrial unit. By identifying SNPs present in this contig I was able to infer the differences between each side of the dimer. This method detected two SNPs or heteroplasmies that would indicate differences between sides of the dimer, while the third SNP was located near the end of the contig.

The first SNP causes a change from tRNA-Gly to tRNA-Arg, while the second SNP results in a change from tRNA-Leu1 to tRNA-Leu2. It remains to be determined whether the third SNP is functional, or whether it is even truly present - as it is located very near the end of the mitochondria and the Illumina reads do not map this section. Additionally, the two functional SNPs are present at an approximately 50:50 ratio, while this SNP exists at a 30:70 ratio. The

two tRNA SNPs have previously been observed within *Isopoda* at similar ratios to those I observed (Chandler, Badawi, Moumen, Grève, & Cordaux, 2015), while the third SNP has not previously been reported. Alternatively, the low ratio of the third SNP could occur if it is present on the linear monomer (which may exist as ⅓ of the copies of the mitochondrial unit), rather than the dimer.

The first two tRNA SNPs have been reported within *Oniscidea*, alongside another SNP which causes the tRNA-Val loci to also code for tRNA-Ala on the other side of the dimer (Doublet, Souty-Grosset, Bouchon, Cordaux, & Marcadé, 2008; Peccoud et al., 2017). However, other non-functional SNPs have been reported within *Oniscidea*, at lower ratios than 50:50. This means I cannot rule out our third SNP as being truly present. The presence of the two tRNA SNPs within both *Oniscidea* and *Sphaeromatidae* is most likely the result of maintenance due to selection, as these SNPs encode for vital genes that should result in balancing selection. This explanation is also the most parsimonious, because it requires the fewest steps to achieve the same outcome relative to a convergent evolution explanation. This also suggests that these sites have been maintained across very highly diverged isopod lineages and, given the divergence times of families, suggests maintenance across hundreds of millions of years (Lins, Ho, Wilson, & Lo, 2012).

2.4.4 Structural Differences to Oniscidea

In *Oniscidea*, there are conflicting reports on the size of the junctions between the genome units in the dimer. One recent paper suggests that the gap between the 12S genes range from 34 to 42 base pairs, and for the CYTB genes, it ranges from 0 to 3 bp (Peccoud et al., 2017). While another paper suggests that the CYTB junction should span at least 249 base pairs, as this junction contains the mitochondrial control region (Doublet et al., 2013). Doublet et al (2013) suggests that this may be reconciled by the presence of an additional transient or rare alternative structure that may not be detected when pooling multiple samples, as done by Peccoud et al (2017). Our results contrast with both results and suggests that the size of these junctions within *I. armatus* are approximately 144 base pairs for the 12S junction and approximately 919 base pairs for the CYTB junction.

Regardless, our results contrast quite strongly with observations made by both Doublet et al. (2012) and Peccoud et al. (2017) in that the junctions within *Isocladus armatus* are not only much larger than those reported by previous studies, but also contain genes beyond those of the ORI - specifically containing 3 tRNAs. While tRNA-Trp has previously been reported as potentially within this junction, tRNA-Ser1 and tRNA-Glu have not.

Interestingly, in *I. armatus* two copies of tRNA-Ser are present, with one found in the same location as in other species - between the ND2 and ND6 genes. Unsurprisingly the structure of this region is very similar across species, with the greatest similarity to *S. serratum*. The key differences I observe are the presence of an ORI at the end of the ND2 gene (this is probably erroneous), but also that the tRNA-Lys locus in *I. armatus* is at the same position as the tRNA-Glu locus in *S. serratum*. Instead the tRNA-Glu locus is for *I. armatus* is found within the CYTB junction.

2.4.5 Is atypical structure an ancestral state of *Isopoda?*

I find that atypical structure in *Isocladus armatus* presents in almost the same way as *Oniscidea* despite these two lineages being highly diverged. The key difference in the way this structure presents is the absence of the tRNA-Val/tRNA-Ala heteroplasmy. In order to deconstruct the evolution of these traits, it is important to consider that the heteroplasmies are not synonymous with the presence of 'atypical' structure. Because I observe atypical structure within *I. armatus* it is most likely that atypical structure evolved first in an ancestral isopod, followed by 'reversion' to a 'typical' metazoan structure within *S. serratum*.

Atypical mitochondrial structure is widespread throughout *Isopoda*, but is curiously absent in some species despite close relatives processing this atypical structure (such as family:*Ligiidae* and family:*Janiridae*) (Doublet et al., 2012; Kilpert et al., 2012). Given that this structure appears paraphyletic within *Isopoda*, it is difficult to explain its evolutionary origins. There are two hypotheses proposed that could explain the distribution of this structure across *Isopoda*. The first hypothesis proposes that this atypical structure appeared multiple times independently across lineages, while the second hypothesis suggests that this structure evolved once early on in isopod evolution and was followed by repeated reversions among lineages back to a 'typical' metazoan mitochondrial structure (Doublet et al., 2012). Our results support the latter hypothesis, as the presence of atypical structure within the Sphaeromatid *Isocladus armatus* suggests *S. serratum* may have reverted back to a typical structure. Given I observed two heteroplasmies in *I. armatus* that are present in other isopod lineages, it is also likely that these two heteroplasmies are at least as old as the most recent common ancestor between *Sphaeromatidea* and *Oniscidea*. This would then mean a reversion in structure for *S. serratum* would also require the re-acquiring of the two tRNA genes coded for by the heteroplasmies in the atypical structure within *I. armatus*. Alternatively, the absence of atypical structure in *S. serratum* and presence in *I. armatus*

could be explained by convergent evolution of mitochondria within two lineages of *Isopoda* as proposed by (Doublet et al., 2012). I suggest that this is highly unlikely, because it is highly improbable that the same atypical structure evolved twice in two distantly related lineages of *Isopoda*, and that both groups would exhibit the same variants associated with this atypical structure.

While atypical structure alone would not be sufficient to argue that this state is ancestral, the presence of two of the same heteroplasmies within two highly diverged lineages suggests that this structure is likely an order of magnitude older than previous estimates (at 400 million years, rather than 30 million) (Doublet et al., 2008; Lins et al., 2012).

2.4.6 Conclusion

I aimed to assess two means of enriching DNA samples for mitochondrial DNA. I did this by comparing differential centrifugation and multiple displacement amplification to the efficacy of whole genome sequencing for mitochondrial genome sequencing. Additionally, I aimed to sequence and assemble the complete mitochondrial genome for a New Zealand marine isopod, *Isocladus armatus*. This work aimed to enable future work to understand phylogeny within this group and to help understand the evolution of atypical mitochondrial genomes within *Isopoda*.

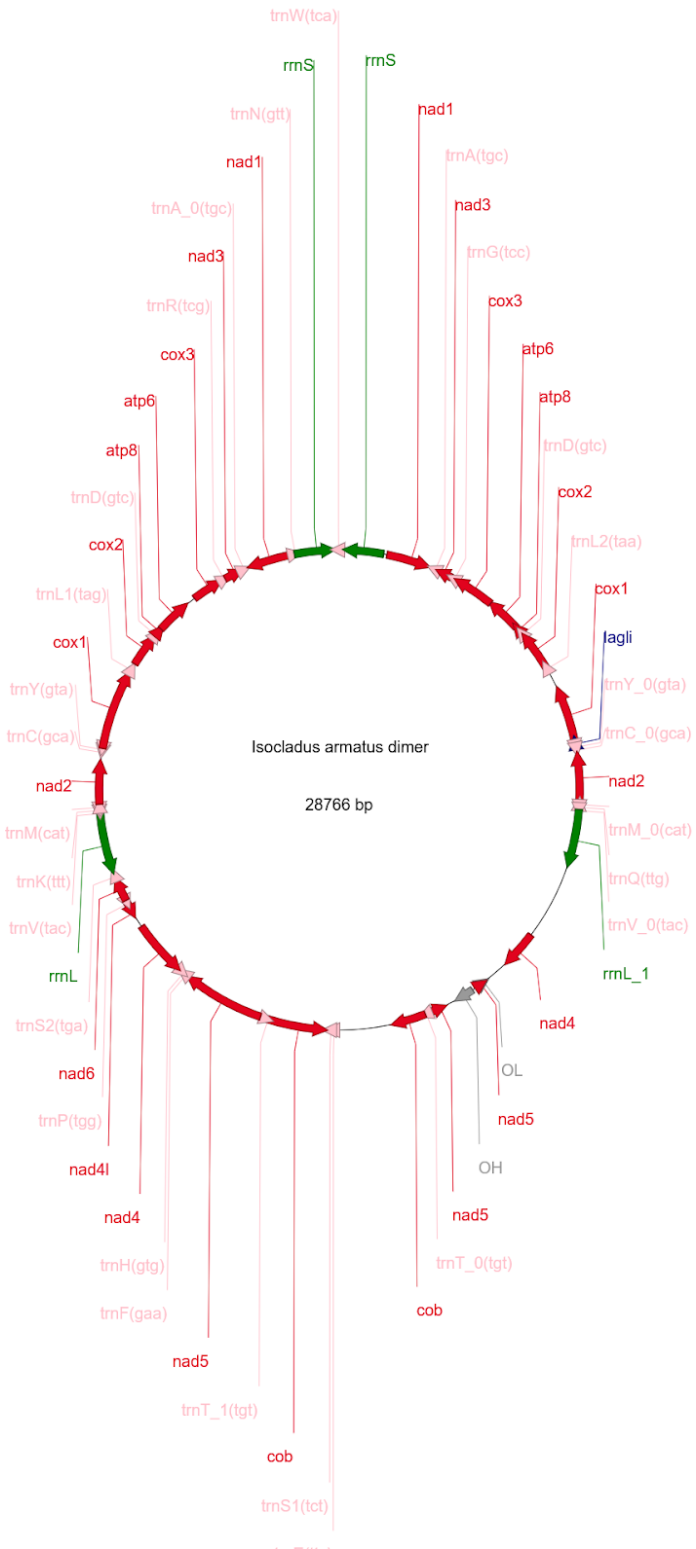
I find that previously established methods for enriching DNA samples for mitochondria failed on *Isocladus armatus*, despite initial results suggesting that they worked. Additionally, I show that long-read sequencing is able to accurately resolve the mitochondrial genome structure, even in the presence of duplications and inverted repeats. Finally, I found that *Isocladus armatus* possesses an atypical mitochondrial structure similar to what is found in many other isopods, such as *Armadillidium vulgare* (Doublet et al., 2012). Importantly, two of the three known heteroplasmic sites for this structure in *A. vulgare* are shared with *I. armatus*, while the third site is not heteroplasmic. I suggest that this atypical structure and shared heteroplasmies are ancestral to at least the closest common ancestor between *Sphaeromatidae* and *Oniscidea*, and suggest that the lack of atypical structure in *S. serratum* may be the result of reversion from atypical structure rather than *I. armatus* possessing this structure through convergent evolution. These results support the suggestions of Doublet et al (2012) who propose atypical structure as ancestral within *Isopoda*.

2.4.7 Future Directions

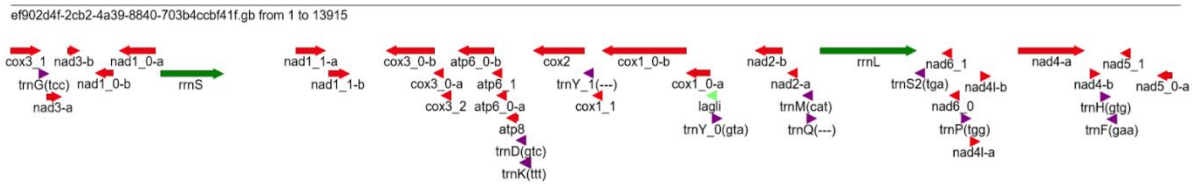
The structure of the mitochondrial genome should be further corroborated through PacBio sequencing, southern blots, and better characterization nanopore signal associated with hairpins. Specifically, PacBio sequencing should be conducted following the analyses performed by Peccoud et al (2017) in order to better characterize the chromosomes and the structure of the junctions. Southern blots should be conducted to ascertain the presence of both a 28Kb circular and 14Kb linear chromosome. Finally, further characterization of the nanopore sequencing signal for naturally occurring hairpins should be performed. This could be performed by obtaining synthetic oligonucleotides with and without predicted hairpins, as well as an oligo that contains the same sequence as that of the CYTB junction in *I. armatus*. Further sequencing of these would help in the characterization of a) the presence of the hairpin in the linear chromosome, and b) the DNA sequencing signal of hairpins.

Future work should involve the addition of Illumina sequencing to this data (which has higher accuracy than nanopore sequencing) in order to correct erroneous indels found in this assembly. This work should also aim to generate mitochondrial genomes for species throughout *Sphaeromatidae*. This should involve work within the genus *Sphaeroma*, in order to understand whether atypical structure has been secondarily lost, or gained repeatedly across *Isopoda*. Given that I did not observe the well documented tRNA-Ala/Val heteroplasmy in *I. armatus*, further work aiming to understand atypical mitochondrial structure should also include a range of species across various evolutionary lineages. Finally, any additional work should aim to generate mitochondrial genomes for species within the *Exosphaeroma* clade and attempt to resolve the relationships between species. This is particularly important within *Isocladus*, as there is evidence (see chapter 2) that *Isocladus armatus* may represent a cryptic species complex.

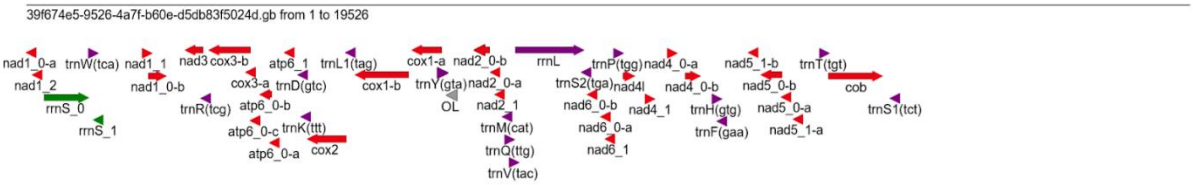
Supplementary



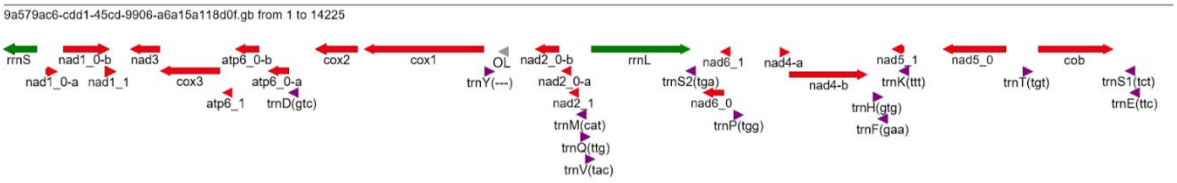
Supp. Fig. 1 Annotations of the entire dimer - because the annotations appeared inconsistent or inaccurate over regions of low coverage, I primarily rely on annotations of the mitochondrial unit rather than these annotations.



Supp. Fig 2. Annotations from mitos of a raw mitochondrial read (ef902d4f-2cb2-4a39-8840-703b4ccbf41f)



Supp. Fig 3. Annotations from mitos of a raw mitochondrial read (39f674e5-9526-4a7f-b60e-d5db83f5024d)



Supp. Fig 4. Annotations from mitos of a raw mitochondrial read (9a579ac6-cdd1-45cd-9906-a6a15a118d0f)

Chapter 3 - Dispersal and Population Genomics of *Isocladus armatus*

Abstract

Population connectivity in marine organisms is widely studied, and relatively well understood for biphasic species. However, comparatively little study has been conducted on direct developers in New Zealand. While most research suggests that direct developers should exhibit lower gene flow between populations than biphasic species, this is not always true and there are numerous cases where direct developers exhibit higher connectivity that otherwise expected. In this study, I use a panel of 8,020 SNPs to describe population structure and gene flow for a direct developer, the endemic marine isopod *Isocladus armatus*, over a range of spatial scales. I find evidence to suggest that on a small spatial scale, gene flow between locations is extremely high, suggestive of an island model of migration, or a metapopulation. However, on a larger spatial scale, populations exhibit a very strong pattern of Isolation-By-Distance. The intersection of the sampling range by a well-known biogeographic barrier at East Cape, provides the opportunity to understand how this barrier can influence direct developers. Interestingly, this research indicates that *I. armatus* has unusually high migration over this barrier, suggesting a low influence of this barrier for population connectivity in *I. armatus*. My study suggests a need to revisit and further study the way biogeographic barriers affect species with different life histories.

3.1 Introduction

The highly dynamic nature of marine environments can influence how marine populations are structured and connected (Temunović et al., 2012). Understanding how the marine environment influences population structure and connectivity is important and can be useful for fisheries and conservation management strategies. However, while a significant amount of research has been dedicated to understanding how population structure is shaped by environment and life history within terrestrial environments, less is known about how these variables affect population structure in marine systems. Furthermore, species within the marine environment display a wide range of life histories and often have major differences with terrestrial organisms. We know little about how differences in life history traits can affect population structure at large scales. This is partially because marine populations and

demographics can be difficult to study using conventional ecological techniques due to the completely different environmental medium to the terrestrial environment.

Life history influences gene flow and study approaches

One example of the difficulty of studying population structure in marine environments is that directly tracking specific individuals movements (e.g. by physically 'tagging' every individual within a population) is not easy (although it has been done with larger marine species such as tuna (Lutcavage et al., 2000), sharks (Queiroz et al., 2016), and sea turtles (Schofield et al., 2007)). Because dispersal is often sporadic, it tends to be impractical or impossible to detect for many marine species using these conventional methods in marine environments. Additionally, applying these methods to marine species (and often biphasic species; these are defined by having a juvenile larval stage) may be difficult due to the often large number of offspring produced, which incur high mortality (r-selected species). This can hinder attempts to understand population connectivity in the absence of genetic data (Thorrold et al., 2002). Because biphasic species may undergo dispersal during the larval stage, it is often difficult, if not impossible, to track these larvae because they are incredibly small relative to the size of the ocean (Levin, 2006).

These species contrast to direct developers which have a single life phase where juveniles tend to resemble immature or miniature forms of the adult. The type of life cycle of a species has been shown to have a significant effect on population connectivity and dispersal of species, where direct-developers tend to show greater population structure than their biphasic counterparts (Puritz et al., 2017). For many biphasic species, the adult form tends to have limited movement over large spatial scales, and instead dispersal is generally the result of movement of the larval form (Puritz et al., 2017). Dispersal during the larval phase may represent the most important mechanism behind population connectivity in sessile species such as oysters, because the larval forms tend to be mobile and displaced by ocean currents.

In contrast, direct developers tend not to display such a degree of mobility and may have instead have much lower dispersal potential than their biphasic counterparts. Some direct developers have evolved mechanisms to increase dispersal potential, such as some molluscs which secrete mucous threads to increase buoyancy and enable flotation within the water column (Martel & Chia, 1991; Prezant & Chalermwat, 1984). Other direct developers are thought to rely on rafting for dispersal, such as attachment to floating kelp rafts (Nikula, Fraser, Spencer, & Waters, 2010). Rafting is a stochastic process whereby individuals

inhabiting oceanic debris (i.e algal mats, driftwood, even plastic) move to new locations as a result of the movement of the debris.

The difference in dispersal potential between biphasic species and direct developers has led to the hypothesis that biphasic species should exhibit higher population connectivity (and thus lower population structure) than their direct developing counterparts (Ayre, Minchinton, & Perrin, 2009). While this is often the case, some species buck this trend (Ayre et al., 2009; Puritz et al., 2017). For example, Ayre et al (2009) examined the population structure of ten intertidal species across a biogeographic barrier in Southeast Australia. Gene flow for six of the eight biphasic species was significantly affected by the presence of this barrier, while the two direct-developing species were unaffected. This is contrary to predictions based on the expected relationship between larval dispersal and genetic heterogeneity. These cases are not uncommon within intertidal organisms, suggesting that population structure is not entirely dependent on life history. As a result, it is important to study and understand direct developing species to further understand the relationship between life history and population connectivity.

Models of Gene Flow

Processes such as rafting, as well as other mechanisms of dispersal, can result in populations exhibiting patterns of connectivity consistent with two well-studied models of migration, island and stepping stone models of migration. The island model was proposed by Wright (1931) and posits all populations in question exchange migrants between each other (see Figure 3.1A for an example). Over time this should lead to convergence to a unified allele frequency across all populations in the model. The stepping stone model was proposed by Kimura and Weiss (1964). In this case dispersal occurs only between adjacent populations, which leads to decreased similarity between populations further apart than those closer together (see Figure 3.1B for an example).

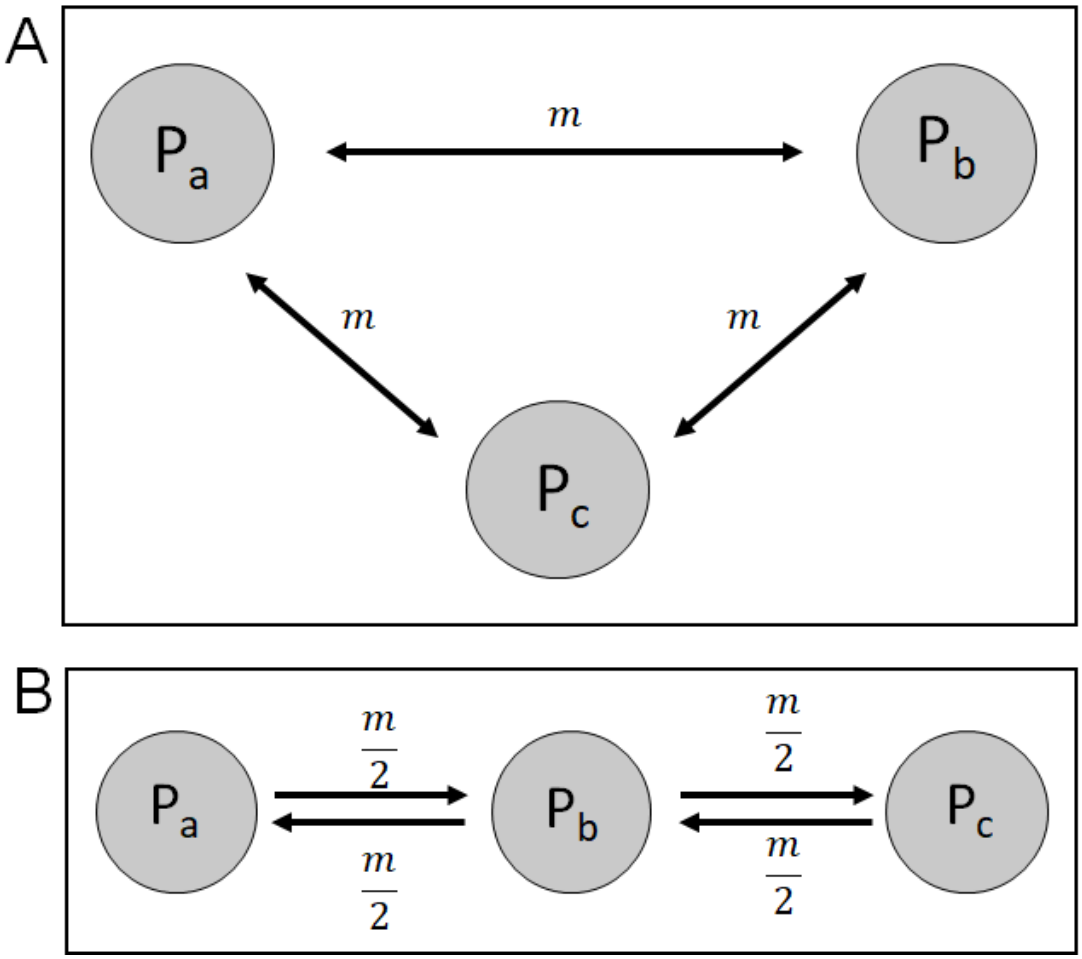


Figure 3.1 Models of migration, A is the Island model proposed by Sewall Wright in 1931, while B is the stepping stone model proposed by Kimura and Weiss in 1964. In both cases *P* represents population, while the subscript letter denotes the specific population, *m* indicates the migration rate between each population. In the case of B this represents the overall unified migration rate.

Dispersal or migration as measured by direct observation is not necessarily an accurate measure of gene flow if the migrating individuals do not reproduce, because reproduction is necessary for contemporary genetic connectivity. However, population genetics, the study of how populations interact is based on genetic information (either theoretical or based on real molecular data) can provide a practical alternative to direct observation by enabling indirect estimation of these parameters. This method of genetic analysis uses information on allele frequencies among populations to understand how they are connected; where more similar frequencies of alleles are associated with greater genetic similarity. Consequently, population genetic analysis can provide detailed estimates of migration, dispersal, and population structure, without relying on direct observation. However, population genetics can struggle to distinguish contemporary and historical gene flow. This is an issue more problematic for marine organisms, as many approaches make assumptions about constant

and equal migration, effective population sizes, and gene flow and genetic drift being in equilibrium for the populations being analysed - assumptions which are often unmet in marine species (Hellberg, 2009). However, the development of more sophisticated analyses have relaxed this assumption (Mussmann, Douglas, Chafin, & Douglas, 2019; Wilson & Rannala, 2003).

Population Genetics in *Isopoda*

Marine isopods offer a highly tractable opportunity to investigate dispersal potential for direct-developing species because they are reasonably well studied and present a high amount variation in population structure at different spatial scales (Table 3.1). For example, the species *Jaera albifrons* exhibits significant population structure at a scale of 5m (Piertney & Carvalho, 1994), while for other species the distance between sampling localities can be >11kms before structure is detected (Mariella Baratti, Goti, & Messana, 2005; M. Baratti, Filippelli, & Messana, 2011; Wells & Dale, 2018).

Table 3.1 Species of intertidal isopods for which population genetic data is available along with the minimum reported distance at which significant population structure is detected. mtDNA is DNA from the mitochondria, while SNPs are Single Nucleotide Polymorphisms.

Species	**Distance at which significant population structure is detected**	**Marker Type**	**Reference**
Isocladus armatus	>11kms	SNPs	(Wells & Dale, 2018)
Idotea baltica	>26kms	Allozyme	(Sami Merilaita, 2001)
Jaera albifrons	5m	Allozyme	(Piertney & Carvalho, 1994)
Exocirolana braziliensis	1km	Allozyme;mtDNA	(Lessios & Weinberg, 1993; Sponer & Lessios, 2009)
Austridotea lacustris	11km	mtDNA	(McGaughran, Hogg, Stevens, Lindsay Chadderton, & Winterbourn, 2006)

Additionally, some species of isopod exhibit transoceanic distributions, despite completely lacking a larval phase. One example is *Sphaeroma terebrans*, an isopod found in mangroves that is distributed across the Atlantic and Indian Oceans. The high dispersal rate and wide

distribution of this wood-boring isopod is likely the result of a reliance on rafting for dispersal (M. Baratti et al., 2011). These cases are interesting as they present two contrasting pictures of dispersal within isopods. On one hand, microgeographic population (<1 Km (Lessios & Weinberg, 1993; Sponer & Lessios, 2009)) structure is found in some species - suggesting relatively low dispersal capacity, while on the other hand some species exhibit connectivity over much larger spatial scales (>20 Kms (Sami Merilaita, 2001)). These cases represent an interesting scenario where life history may not necessarily be a good predictor of population connectivity. This is evidenced by *Septemserolis septemcarinata*, a benthic sub-Antarctic shallow-water isopod. Leese et al. (2010) conducted a population genetic study on this species looking at populations on three sub-Antarctic islands. This work found that despite significant geographic barriers to dispersal, models that indicated gene flow were more likely than a model indicating no gene flow. Furthermore, the most likely model of gene flow was congruent with the direction of the Antarctic Circumpolar Current, suggesting that ocean currents may play an important role in dispersal for both biphasic and direct-developing organisms. The unexpectedly high connectivity between two of these populations suggests that there is long distance current dispersal between these barriers as a result of rafting on algal mats. In New Zealand, this is thought to be a major dispersal mechanism for the isopod species *Limnoria stephenseni*. A comparison of phylogeographic diversity for this species and its macroalgal host showed concordant patterns, suggesting *L. stephenseni* is dispersed through rafting on this macroalgae (Nikula et al., 2010).

Within a New Zealand context, the majority of population genetics studies in marine environments are focused on species with a biphasic life cycle, with less attention paid to direct developing species. A recent review of population genetic studies on benthic intertidal invertebrates in New Zealand showed only 10 out of 56 studies were focused on direct developers (Arranz Martinez, 2017). However, multiple species of New Zealand isopods have exhibited greater than anticipated population connectivity despite their life history (McGaughran et al., 2006; Wells & Dale, 2018). These results reinforce the idea that life history may not be a good predictor of population connectivity for isopods. For example, species from the freshwater-estuarine genus of isopods, *Austridotea*, have shown high levels of population connectivity over a distance of 11km, with most divergence between populations detected on much larger spatial scales (McGaughran et al., 2006).

These above patterns were supported by Wells and Dale (2018), who suggest that the marine isopod *Isocladus armatus* may have a greater dispersal potential than expected. However, it is unclear how this may function over larger spatial scales, especially where known biogeographic breaks intersect the range of populations being studied. Wells and

Dale (2018) showed that there was no significant population divergence between two sites separated by 11km of coastline. On a much larger scale (1000kms), however, there was negligible connectivity and the populations were considered entirely isolated. The lack of population structure on a local scale is potentially indicative of a highly mobile species. However, it is not clear how the biogeographic barriers and geographic distance between the populations studied by Wells and Dale (2018) may affect this species over intermediate spatial scales. Increasing sampling to encompass a wider range of distances between populations can provide greater resolution of population connectivity within this species.

Specific Objectives

In this study I aim to quantify population differentiation and connectivity between populations of the isopod *Isocladus armatus* over a range of spatial scales. I integrate these results with data from a previous analysis (Wells and Dale 2018) with the objective of understanding the physical range of populations of this species, and to understand the role of geography (both distance and seascape features) that could influence gene flow and dispersal.

3.2 Methods

3.2.1 Sample Collection and Processing

Sample Collection

Specimens of *Isocladus armatus* were collected in June 2018, from around the North Island, New Zealand, from locations where *I. armatus* has previously been recorded (Fig. 3.2) (Hurley & Jansen, 1977; “iNaturalist.org,” n.d.). These sites were at Stanmore Bay, Browns Bay, Opito Bay (Coromandel), Mt Maunganui, Māhia Peninsula, and Wellington. At each site I attempted to collect a minimum of 32 individuals, up to a maximum of 48, distributed as equally as possible amongst morphotype (see chapter 4) and sex. Where possible, I collected specimens larger than 5mm in order to ensure sufficient DNA could be extracted for sequencing. At each site, the maximum distance between two samples was approximately 30 m. Samples were stored at -80°C in 100% ethanol until extraction. For details on samples collected and processed in 2015 (Kaikōura, Hatfields Beach, and one set of Stanmore Bay samples) see Wells and Dale (2018).

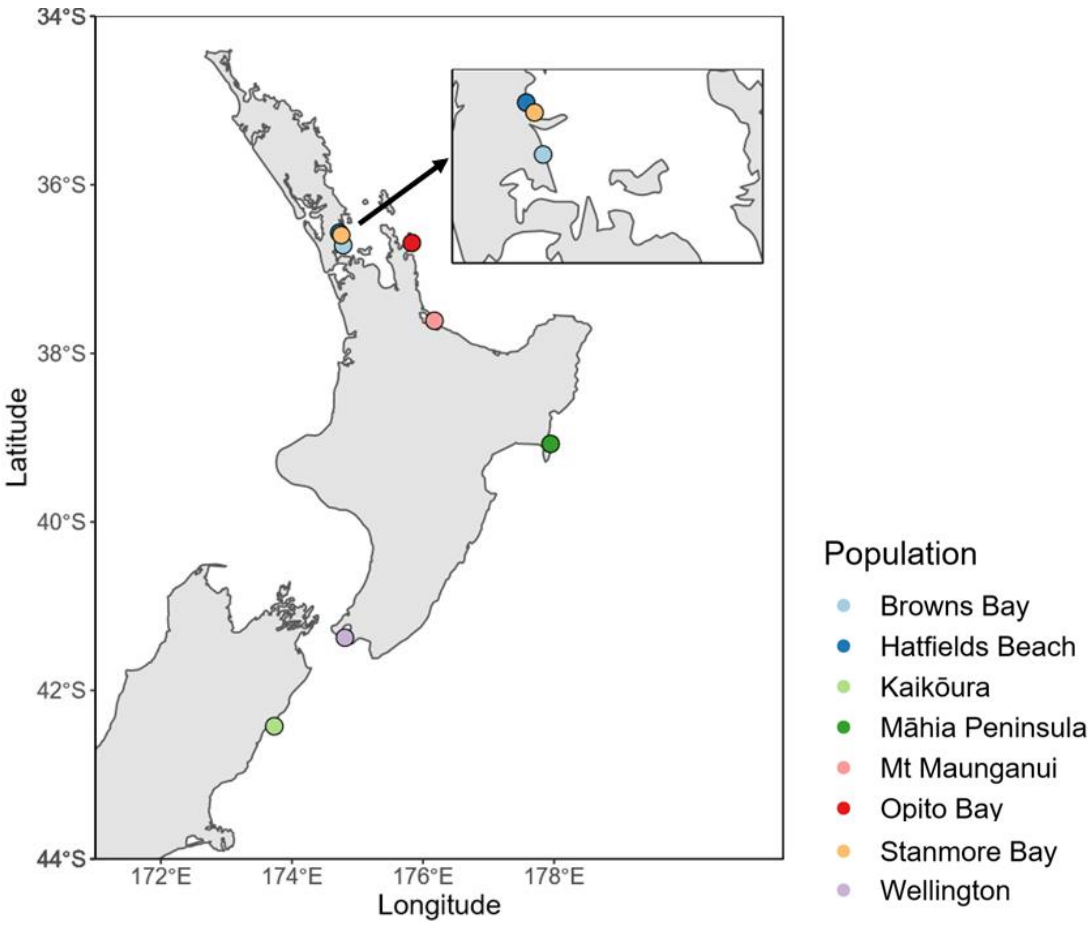


Figure 3.2 Map of sampling localities, coloured dots indicate location. Inset in top right indicates the populations within the Auckland region.

Where >32 samples were collected, I chose samples for DNA extraction and subsequent sequencing based on the size and degree of morphotype ambiguity (where less ambiguous morphotypes were chosen preferentially over the more ambiguous ones (see chapter 4 for details)). Details of the specific number of each morphotype and sex from each location are detailed in Table 3.2.

Table 3.2 Number of samples from each location grouped by sex and colour morphotype. All sites are from 2018 except where explicitly stated.

Site	Sex	Striped	Unpatterned	Variegated	White	Other	Sum
Hatfields Beach 2015	Female	3	6	4	2	1	16
Hatfields 2015	Male	0	4	4	4	3	15
Stanmore Bay 2015	Female	4	4	4	4	0	16
Stanmore Bay 2015	Male	4	4	4	4	0	16
Stanmore Bay	Female	4	4	4	4	0	16
Stanmore Bay	Male	4	4	4	4	0	16
Browns Bay	Female	4	4	4	4	0	16
Browns Bay	Male	4	4	4	4	0	16
Māhia Peninsula	Female	0	6	4	3	0	13
Māhia Peninsula	Male	2	9	5	0	0	16
Mt Maunganui	Female	1	3	3	4	0	11
Mt Maunganui	Male	4	4	4	3	0	15
Opito Bay	Female	4	6	1	2	0	13
Opito Bay	Male	3	6	4	3	0	16
Wellington	Female	0	5	5	5	0	15
Wellington	Male	0	5	5	5	0	15
Kaikōura 2015	Female	1	7	4	3	0	15
Kaikōura 2015	Male	1	5	4	0	5	15

DNA Extraction

DNA was extracted using a modified protocol for the Qiagen DNEasy Blood and Tissue kit. Samples were first rinsed with ddH2O to remove residual ethanol and then the cephala were

removed using an ethanol and flame sterilized scalpel. The cephala were then homogenized in a 65°C solution of 22 µl 20% SDS and 178 µl of 0.5M EDTA (pH 8) to denature fast-acting nucleases (Wells & Dale 2018). 10 µl of Proteinase K (20mg/mL, >600mAU/mL) was then added and the solution was incubated at 65°C at 800 rcf for 24 hours. After 24 hours, 400 µl of Qiagen Buffer ATL was added with an additional 15 µl aliquot of Proteinase K, and incubated again for 24 hours.

After this, 400 µl of Buffer AL and 400 µl of 100% Ethanol was added to each sample and vortexed for 20 seconds. The solutions were passed through a spin column in two steps due to the large volume. The column was then washed according to the manufacturer's protocol, before a three step elution took place. Each elution consisted of 50 µl of nuclease-free water incubating on the column for 15 minutes before centrifugation for one minute at 7,000 rcf. This resulted in a total of 150 µl of eluate.

DNA samples were concentrated in an Eppendorf Vacuum Centrifuge, at 30°C, to reduce total volume to approximately 50 µl. At this point the concentration was determined by using 2 µl of sample in a Qubit fluorometer assay.

The concentrations were then adjusted through the addition of nuclease-free water to be between 10 and 50 ng/µl before being submitted for DArTseq.

DNA Sequencing and GBS Method

DNA samples were processed by Diversity Arrays Technology Ltd (Canberra, Australia) using a genome complexity reduction approach, a form of Genotyping-By-Sequencing. In this case, samples were processed with the same approach as employed by Wells and Dale (2018), so as to ensure datasets could be used together, as the DArTseq approach is optimized for each species.

The DArTseq approach digests high molecular weight DNA using restriction enzymes (in this case, the enzymes were PstI and SphI). Adapters with sequences complementary to the restriction enzyme overhangs are ligated to the digested DNA. The samples are then amplified, using primers hybridizing to the adapters, to amplify fragments containing the adapters. The amplified products are then sequenced on an Illumina HiSeq 2500. This approach is explained in more detail in; (Elshire et al., 2011; Kilian et al., 2012; Wells & Dale, 2018).

The resulting sequences are demultiplexed based on individual barcodes, then filtered to remove reads with mean Q-scores less than 25. DArT proprietary bioinformatics pipelines were implemented to first remove low quality sequences, and stringently assess barcode accuracy to ensure samples are correctly assigned to the appropriate individual. The reads are then compared against both GenBank viral and bacterial databases, as well as an internal Diversity Arrays database to identify potential contaminants.

The filtered data is then processed into a secondary pipeline, which performs SNP calling. SNPs are only called if both homozygous and heterozygous genotypes can be identified. This approach incorporates technical replicates from the library preparation step, to ensure consistency in SNP calls across replicates.

SNP Filtering and Data Processing

The dataset provided by DiversityArrays was analyzed together with the data from Wells and Dale (2018). To ensure datasets were comparable I filtered each dataset separately based on the conditions described below, and then extracted loci shared in both datasets for the remainder of the analyses. This was because the increased data output in the 2018 dataset (as a result of the inclusion of more individuals and populations) meant that almost twice as many SNPs were identified in this dataset, than in the 2015 dataset. By retaining shared SNPs the effect of this was minimized. One individual was excluded from all analyses (BBFA3), as this sample had a very high number of SNPs missing across all loci (93%), suggesting it may be a different species. Filters were conducted in the R packages; dartR (Gruber, Unmack, Berry, & Georges, 2018) and radiator (Gosselin, 2019).

A SNP call rate of ≥ 0.9 was implemented. This ensures that for each locus, a SNP genotype was identified in at least 90% of all individuals. Mean read depths of >5X and <50X were also implemented, while high read depths are often desired in other scenarios, excessively high coverage may suggest duplicated genome elements which could confound further analyses. On the other hand, low read depths (<5X) may reduce confidence in the accuracy of an allele identification.

SNPs with a low minor allele frequency (MAF) can influence the inference of population structure when using model based analyses (Linck & Battey, 2019). Linck and Battey (2019) observed that as MAF filters increased in stringency, higher levels of admixture were observed. This is partially attributed to the expectation that, under a coalescent model, low frequency alleles are relatively recent mutations within populations, and thus contain more

information about recent population history rather than long term history. Other research suggests that SNPs with low MAF are sometimes the result of technical errors (i.e. PCR, sequencing error, or bioinformatics processing errors), but are also often true aspects of populations (Roesti, Salzburger, & Berner, 2012). The authors suggest that in these cases, low MAF loci may be uninformative, and in fact may bias population structure inference. While it seems that both of these papers may be at odds with each other regarding the cause of the impacts of MAF on population statistics, they are both consistent in suggesting there is a confounding effect that can be avoided by their removal. Filtering on MAF in of itself can be problematic, as missing data is abundant across SNP datasets generated through reduced representation approaches. This can lead to loci with varying levels of missing data being filtered differently (Linck & Battey, 2019). As a result, I retained SNPs with a Minor Allele Count (MAC) ≥ 3, as suggested by Linck & Battey (2019) and Rochette et al. (2019) to ensure consistency across loci.

A filter on observed heterozygosity was also implemented, where loci with a $H_o > 0.5$ were removed. This is because high heterozygosities can be the result of polyploidy, or whole genome duplication events. In a scenario where loci have undergone duplication events and become fixed for each allele then we would expect to see excess heterozygosity (Hohenlohe, Amish, Catchen, Allendorf, & Luikart, 2011).

In cases where multiple SNPs were located on the same sequence, I retained the SNP that had the highest reproducibility (based on the proportion of technical replicates that resulted in the same allele being called), and in the event these were equal, then the SNP with the highest amount variation across individuals was retained. Removing these SNPs ensures that polymorphisms known to be tightly linked are removed, while removing the SNP with the lowest variation ensures that the most informative loci are retained.

Finally, loci that were deemed under selection were also removed. Non-neutral loci are expected to exhibit different allele frequency spectra to neutral loci. Generally, neutral loci are expected to have genomic signatures caused by, or at least consistent with, neutral processes in populations, while loci under selection are affected by selective pressures and thus may bias inference of population demographics (Luikart, England, Tallmon, Jordan, & Taberlet, 2003). Loci under selection (often termed outlier loci) were identified using BayeScan (Foll & Gaggiotti, 2008), with population as the grouping factor. Loci exhibiting a q-value (analogous to p-value but more stringent in the context of BayeScan (Foll, 2012)) of ≤ 0.05 were excluded from any further analyses.

While it is common practice to filter loci based on being in Hardy-Weinberg Equilibrium (HWE) within populations (Morin, Leduc, et al., 2009; Van Wyngaarden et al., 2017; Waples, 2015; Wells & Dale, 2018), this is not always the best practice. For example, STRUCTURE clusters individuals into a Hardy-Weinberg populations (Pritchard, Wen, & Falush, 2010), thus by implementing a filter on HWE for this type of analysis, an *a priori* structure would be imposed on the dataset because loci associated with (or responsible for) would be removed by definition. Additionally, the analyses that were performed (with the exception of calculation of Fst) did not have any reliance on populations being in HWE. Therefore, the only analyses for which a HWE filter was implemented, were the Fst and associated analyses (such as the mantel tests, redundancy analyses, and population statistics). Loci out of HWE were first identified using the R package *pegas* with Bonferroni correction for multiple testing, and removed using the dartR package.

3.2.2 Data Analysis

Basic Statistics

Inbreeding coefficients, observed heterozygosities, and expected heterozygosities were calculated and averaged across loci, using the R package *hierfstat* (Goudet, 2005).

Basic Fst statistics were calculated using StAMPP (Pembleton, Cogan, & Forster, 2013). Fst was used as the primary measure of genetic differentiation. While other measures such as *Gst* and *Josts D* have been developed in recent years to accommodate attributes of marker types such as microsatellites, Fst remains relatively robust for biallelic markers such as SNPs. Furthermore, SNPs have a lower mutation rate than other types of markers (such as microsatellites) and thus are more likely to conform to the assumptions for calculating Fst (Whitlock, 2011). Relatedness was calculated using Triadic Maximum Likelihood following Wells and Dale (2018). This was conducted using Coancestry (Wang, 2011), as implemented in the R package “related” (Pew, Muir, Wang, & Frasier, 2015).

Isolation-by-Distance was tested for by conducting two types of analysis in the R package vegan (Oksanen et al., 2010). I opted to perform two tests, as concerns have previously been raised about the suitability of solely using mantel tests as a tool for detecting isolation by distance (Kierepka & Latch, 2015; Meirmans, 2015). The first was a mantel test conducted testing between a Slatkin's linearized Fst matrix (linearized using 1/1-Fst (Rousset, 1997)) and an overwater distance matrix (overwater distance refers to the

minimum distance between two sites where the path taken is restricted to marine environments with a minimum and maximum depth). Overwater distance was calculated using the marmap (Pante & Simon-Bouhet, 2013) and fossil (Vavrek, 2011) R packages to calculate the minimum distance between populations around the coast, within a depth range of 150 m. The second test was a redundancy analysis (RDA) using the geographical coordinates as independent variables, and allele frequencies as the dependent variables (Meirmans, 2015). The amount of variance explained by the constrained variables is then equal to the variance explained by the spatial components. This was then converted to the total amount of genetic variation explained by multiplying by the total amount of genetic variation (as calculated through the basic statistics function in hierfstat).

Principal Component Analysis

Principal Component Analysis (PCA) is a widely used tool for visualizing and understanding the genetic structure of populations (Ma & Amos, 2012; Novembre & Stephens, 2008; Reich, Price, & Patterson, 2008). They are particularly useful as no assumptions are made about the data (i.e such as populations being in HWE), nor do they utilize any prior knowledge about population structure, additionally they have no reliance on a biological model and are computationally simple. PCA's were performed using the R package *adegenet* (Version 2.1.1) (Jombart & Ahmed, 2011). PCA's were conducted on all populations together, and also only on the Auckland populations. The latter analyses were conducted to help identify any population stratification based on morphotype.

In order to understand whether the principal components relate to geography, a Procrustes transformation of the first two principal components was performed using MCMCpack in R (Martin, Quinn, & Park, 2011). Procrustes transformations scale, stretch and rotate the PCA in order to minimize the differences between two matrices (in this case - principal components were compared to geographic coordinates). The scaled PCA coordinates were then superimposed on a map of New Zealand.

Structure Analysis

STRUCTURE analyses were implemented using STRUCTURE 2.3.4 (Falush, Stephens, & Pritchard, 2003) and performed with all populations (including the repeated samples of Stanmore Bay). For these analyses, an admixture model was assumed, Markov Chain Monte Carlo simulations were run with 100,000 iterations, and a burnin of 50,000. 10 replicates of each run were conducted, for each value of K from 2 through 9. Final population

inference was performed by consolidating the results for each level of *K*, in CLUMPP (Jakobsson & Rosenberg, 2007).

Identifying the optimal value of *K*, or genetic clusters in a dataset, has received significant attention (Evanno, Regnaut, & Goudet, 2005; Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015; Verity & Nichols, 2016; Waples & Gaggiotti, 2006). However, choosing *K* is not as simple as $K=N_{pop}$ because increasing K also increases the risk of overfitting the model and explaining structure that doesn't necessarily exist. On the other hand, setting *K* as too low has the potential to underfit the model and fail to identify real population structure. Additionally, multiple hierarchical levels of genetic structuring may exist within a species, and this information may not be captured by defining K (a good discussion about this can be found in (Meirmans, 2015)).

It is also important to recognize that the defining criterion for a population are relatively loose and may be contextually defined. For example, varying *K* in analyses of human populations shows distinct groups falling out with increasing values of *K* but the specific designation of these groups as populations is subjective (M. Hahn, 2019). For example, in human population genetics, populations of entirely admixed individuals exist, at which point we must make a decision as to whether we are interested in the contemporary population, or historical populations. For this reason, I chose to present a range of scenarios where *K* varies from 2 to 6. Additionally, early investigations using the Evanno method (Evanno et al., 2005) to identify optimal values of K failed and no optimal value was determined.

Analysis of Molecular Variance

An Analysis of Molecular Variance (AMOVA) was conducted using the R packages *ade4* (Dray, Dufour, & Others, 2007) and *poppr* (Kamvar, Tabima, & Grünwald, 2014). This approach is used to partition the variance in genetic data into within and among (or between) levels of a predefined hierarchy (levels could include individuals, populations, or large groups such as countries) (Excoffier, Smouse, & Quattro, 1992). I tested the significance of the AMOVA through the random permutation function in *pegas*, with 1000 permutations. I used a hierarchy (in ascending order) of sample, morphotype, population, and then by the observed north-south division found in my other analyses.

Migration Estimates

Estimates of migration rates between populations were calculated using BayesAss3-SNPs (Mussmann et al., 2019; Wilson & Rannala, 2003). This method estimates migration rates

based on Bayesian Markov Chain Monte Carlo (MCMC) inference. I used the BA3-Autotune software developed alongside BayesAss3-SNPs to identify optimal parameters for the analysis. I excluded the 2015 Stanmore Bay population for this analysis to avoid duplicating a population for migration estimates.
I ran this analysis with 1000 randomly sampled loci from section 3.4, and used the following parameters for the migration estimates: $5x10^6$ burn-in and $1x10^7$ iterations, allele-frequency mixing of 0.2125, migration mixing of 0.075, and inbreeding mixing of 0.075.

I filtered migration rates estimated to be close to zero by subtracting the standard deviation from the mean and dropping estimates with values less than 0.005. This approach removes any estimates of zero or insubstantial migration.

Estimates of Dispersal Distance

I estimated average dispersal distance following the methods described by Puebla et al. (2009), Rousset (1997) and Van Wyngaarden et al. (2017). Over an appropriate spatial scale, it has been shown that Slatkin's Fst can be estimated using the following calculation:

$$Fst_{slat} \approx \frac{A_1}{4D\sigma} + \frac{r}{4D\sigma^2}$$

Using this calculation, it is possible to estimate mean dispersal distance (σ) from A1, population density (D), the distance between sites, and the intercept of the regressions of Slatkin's Fst on distance. However, A1 is a constant that in my case is unknown and is dependent on dispersal function. Puebla et al. (2009) show that the slope of the IBD relationship can then be used as an estimate of (4D2)-1. As a result, it is possible to directly estimate σ if only D is known.

Thus, the following calculation was used to estimate σ :

$$Fst_{slope} \approx \frac{1}{4D\sigma^2}$$

This can be rearranged to estimate σ, through the following equation:

$$\sigma = \frac{1}{2\sqrt{D}\ \sqrt{Fst_{slope}}}$$

Fst_{slope} was estimated by performing a linear regression in R 3.5. This regression predicted genetic distance (Slatkin's Fst) using geographical distance (overwater).

D was then calculated following the methods described by Puebla et al. (2009), where I treated the north group (all populations excluding Wellington and Kaikōura) as a single population and calculated its effective population size using MIGRATE-N v 3.6. While it would not normally be appropriate to assume that these groups are a single population, the linearity of Isolation-By-Distance relationship (see 3.3.4) was strong enough to justify this assumption for this analysis, where they are treated as a single linearly distributed population. The effective population size of this group was then divided by overwater distance between the two points in my sample range that were furthest apart (Hatfields Beach, and Māhia Peninsula) in order to estimate population density over this range. Because no information is available on the depth range for *Isocladus armatus* I used the maximum depth observed for a closely related species - *Exosphaeroma gigas*, which was a depth of -150 m (Hurley & Jansen, 1977).

Because this model is only robust over an intermediate spatial scale, I determined the maximum spatial scale at which this model should hold by using the following calculation (Puebla et al., 2009):

$$\sigma - 0.2\sigma/(2\mu)^{1/2}$$

For this calculation I used a mutation rate (μ) of 2.8x10^-9 substitutions per site per generation (Amos, 2010; Keightley, Ness, Halligan, & Haddrill, 2014).

3.3 Results

3.3.1 SNP Filtering and Data Processing

Upon receipt of the SNP data from DaRTseq, I performed quality control on the loci to ensure consistency and appropriateness for population genomic inference. Filtering reduced the total number of loci from 78,927 to 8,020 (Table 3.3). With the most pronounced effects being the removal of secondary loci (these are SNPs that originate from a sequence with at least one other SNP on it). Initial investigations showed that many filters had very little or no quantitative effect on the results - this is likely because either they result in the removal of very few loci, or other filters would have removed the loci as well.

Table 3.3 SNP filtering results and the amount of SNPs remaining after each filter. Datasets were first filtered separately and shared SNPs between the datasets were retained for the combined dataset. This was performed to minimise batch effects as a result of samples being sequenced at two different times. Additionally, the increased data output in the 2018 dataset (as a result of the inclusion of more individuals and populations) meant that almost twice as many SNPs were identified in this dataset, than in the 2015 dataset. By retaining shared SNPs the effect of this is minimized.

Condition	Number of SNPs remaining
Initial number	78,927
Present in Wells and Dale 2018 and the current study, filtered separately	
Call Rate > 0.8	57,348
Depth >3 and <50	54,123
Replicability > 0.9	52,800
Monomorphic Loci	51,370
Secondary loci	31,848
Shared between datasets (N=10,036 SNPs)	
Neutral loci	9,620
Call Rate > 0.9	8,334
Depth >3 and <50	8,334
Replicability > 0.95	8,334
Monomorphic loci	8,334
Secondary loci	8,334
MAC > 3	8,037
Observed Heterozygosity < 0.5	8,020
Loci out of Hardy Weinberg Equilibrium within populations	4,552

3.3.2 Population Statistics

F statistics indicated almost no divergence between each pair of Auckland populations (Hatfields, Stanmore, and Browns Bay), nor was there any evidence for a difference

between time points at Stanmore Bay. Similarly pairs of populations closest geographically to each other also indicated divergence with gene flow - see Table 3.4. Additionally, we saw a sharp change in Fst between Māhia and Wellington.

Table 3.4 Population Fst values generated with STaMPP. Non-significant results are in bold, the only non-significant values are between Hatfields Beach and Stanmore Bay and for populations at Stanmore Bay between years. Non-zero P-values are included in brackets. All Fst values without an associated P value are significant at $P < 0.0001$

	Hatfields Beach 2015	**Stanmore Bay 2015**	**Stanmore Bay 2018**	**Browns Bay 2018**	**Opito Bay 2018**	**Mt Maunganui 2018**	**Māhia Peninsula 2018**	**Wellington 2018**
Stanmore Bay 2015	**0 (P = 0.486)**							
Stanmore Bay 2018	**0.001 (P = 0.062)**	**-0.001 (P = 0.936)**						
Browns Bay 2018	0.002	0.002	0.002					
Opito Bay 2018	0.031	0.029	0.03	0.032				
Mt Maunganui 2018	0.06	0.061	0.061	0.062	0.032			
Māhia Peninsula 2018	0.081	0.079	0.081	0.081	0.064	0.059		
Wellington 2018	0.143	1.39E-01	0.149	0.148	0.163	0.185	0.208	
Kaikōura 2015	0.148	0.144	0.153	0.152	0.166	0.19	0.208	0.11

All populations except Wellington and Kaikōura had relatively similar heterozygosities around 0.11 (Table 3.5), while these two other populations had slightly lower observed heterozygosities at around 0.1. Expected heterozygosities showed a similar pattern, with Kaikōura and Wellington both exhibiting slightly lower H_e(0.6) than the other populations (around 0.12), all values of these included 0 within one standard deviation of the mean suggesting these values were non-significant. Inbreeding coefficients across the populations

range from a maximum of 0.12, to a minimum of 0, like the other statistics, these values all include 0 within one standard deviation of the mean.

Table 3.5 Population statistics calculated using the filtered loci data. Fis is the inbreeding coefficient, R is the relatedness coefficient calculating using Triadic maximum likelihood (Pew et al., 2015; J. Wang, 2011), He and Ho represent expected and observed heterozygosities respectively, calculated using hierfstat (Goudet, 2005). The mean value is presented first, followed by the standard deviation.

Population	Fis	R	He	Ho
Hatfields Beach 2015	0.11 SD 0.28	0.008 SD 0.016	0.12 SD 0.14	0.1 SD 0.13
Stanmore Bay 2015	0.12 SD 0.29	0.0003 SD 0.002	0.12 SD 0.14	0.1 SD 0.13
Stanmore Bay 2018	0.06 SD 0.25	0.0002 SD 0.001	0.12 SD 0.14	0.11 SD 0.14
Browns Bay 2018	0.06 SD 0.23	0.0004 SD 0.002	0.12 SD 0.14	0.11 SD 0.14
Opito Bay 2018	0.06 SD 0.23	0.0007 SD 0.003	0.11 SD 0.14	0.1 SD 0.14
Mt. Maunganui 2018	0 SD 0.21	0.005 SD 0.02	0.11 SD 0.15	0.12 SD 0.18
Māhia Peninsula 2018	0.03 SD 0.21	0.002 SD 0.008	0.09 SD 0.15	0.09 SD 0.15
Wellington 2018	0.01 SD 0.22	0.002 SD 0.005	0.06 SD 0.13	0.06 SD 0.14
Kaikōura 2015	0.06 SD 0.24	0.003 SD 0.014	0.06 SD 0.13	0.05 SD 0.13

3.3.3 Principal Component Analysis

I next investigated population genetic structure using a principal component analysis of the SNPs. This PCA explained a large amount of the variance within the data, with PC1 accounting for 19.5%, PC2 for 3.65%, and PC3 for 2.11% (Fig. 3.3B). The majority of the variance explained by PC1 appeared to primarily delineate Wellington and Kaikōura from the rest of the populations (Fig 3.3A). As a result, I refer to the “Southern Group” as Wellington and Kaikōura, while the Northern group contains all the other sampled populations.

All populations from Auckland completely overlapped. This indicates that there was no obvious delimitation in terms of the frequencies of allele combinations between populations from Stanmore Bay in 2015 and 2018. In addition, Hatfields Beach, Stanmore Bay, and Browns Bay all appeared to be the same population. PC2 primarily delineated the northern populations, and revealed three cases of potentially recent migration between Mt Maunganui

and the Māhia Peninsula (Fig 3.3A) (one individual is obscured in Figure 3.3 by other points).

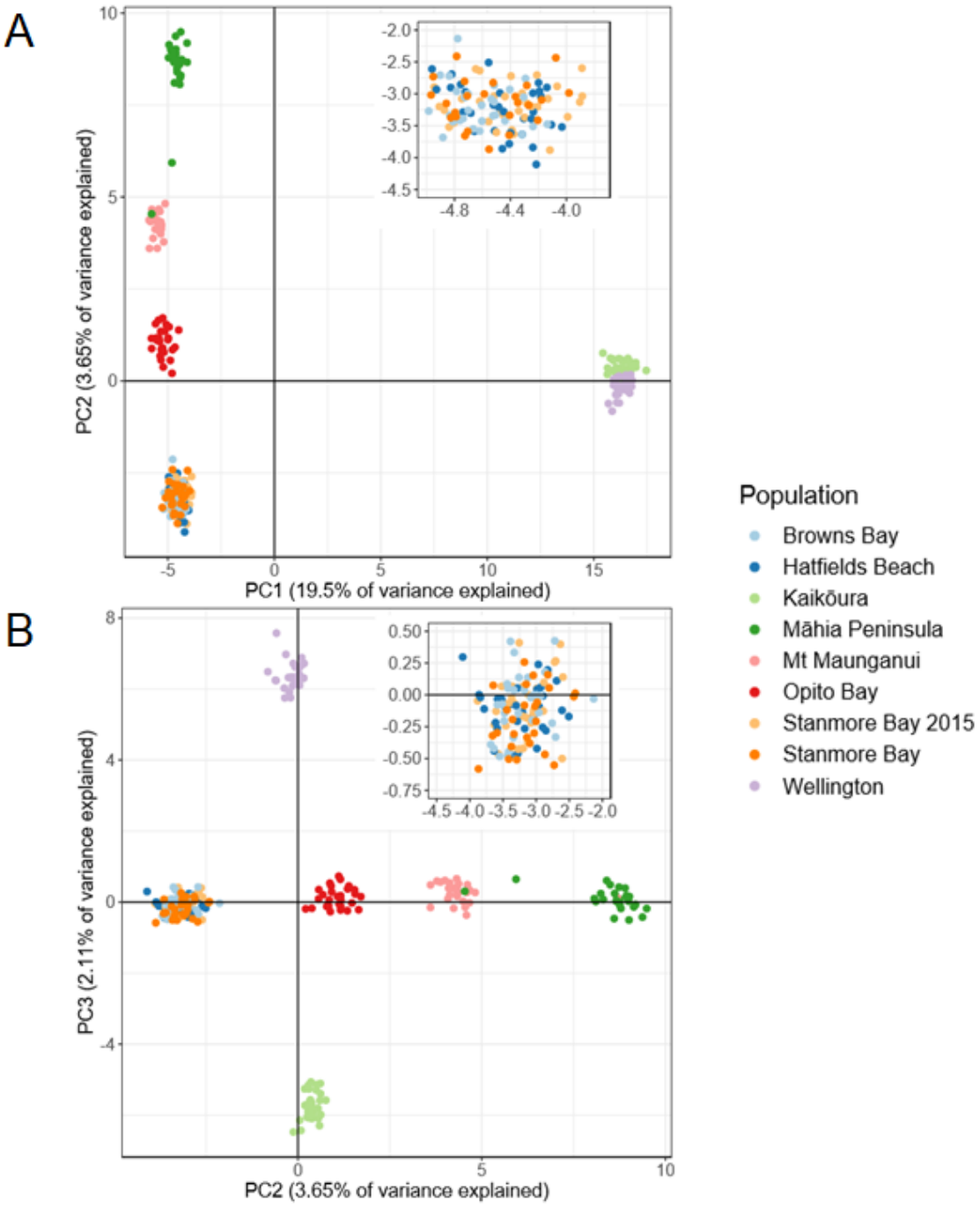


Figure 3.3 Principal Component Analysis (PCA) for the filtered loci coloured by population. PC1 appears to largely delimit the North and South groups of populations, while PC2 separates those in the northern group, and finally PC3 appears to separate the populations within the south group. Insets are focused on the Auckland populations to display any potential substructuring within them.

Additionally, the spacing of each genetic cluster appeared to be concordant with the locations in geographical space. Indeed, A Procrustes transformation of the PCA onto geographical space revealed that the genetic data was strikingly concordant with geography (Fig. 3.4).

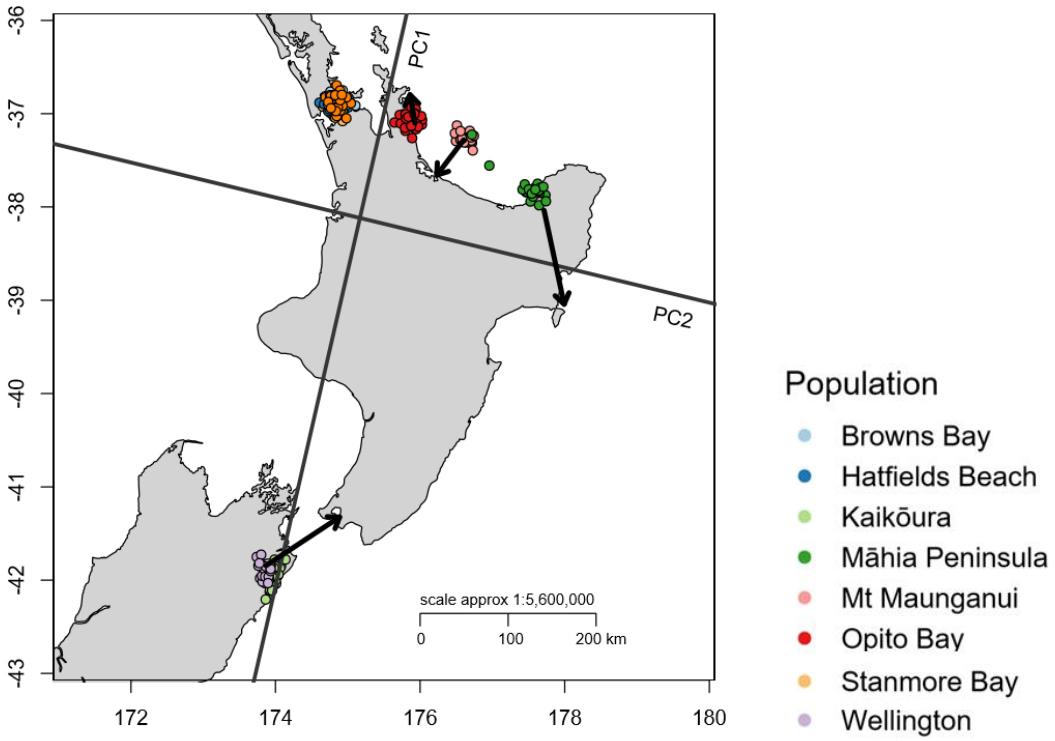


Figure 3.4 Procrustes transformation of PC1 vs PC2 superimposed on a map of the sampling range. Arrows point to where a population is found, while clusters indicate their location in Procrustes transformed space.

Finally, PCA of the Auckland populations based on colouration indicated no substructure or population stratification by morphotype (Figure 3.5)

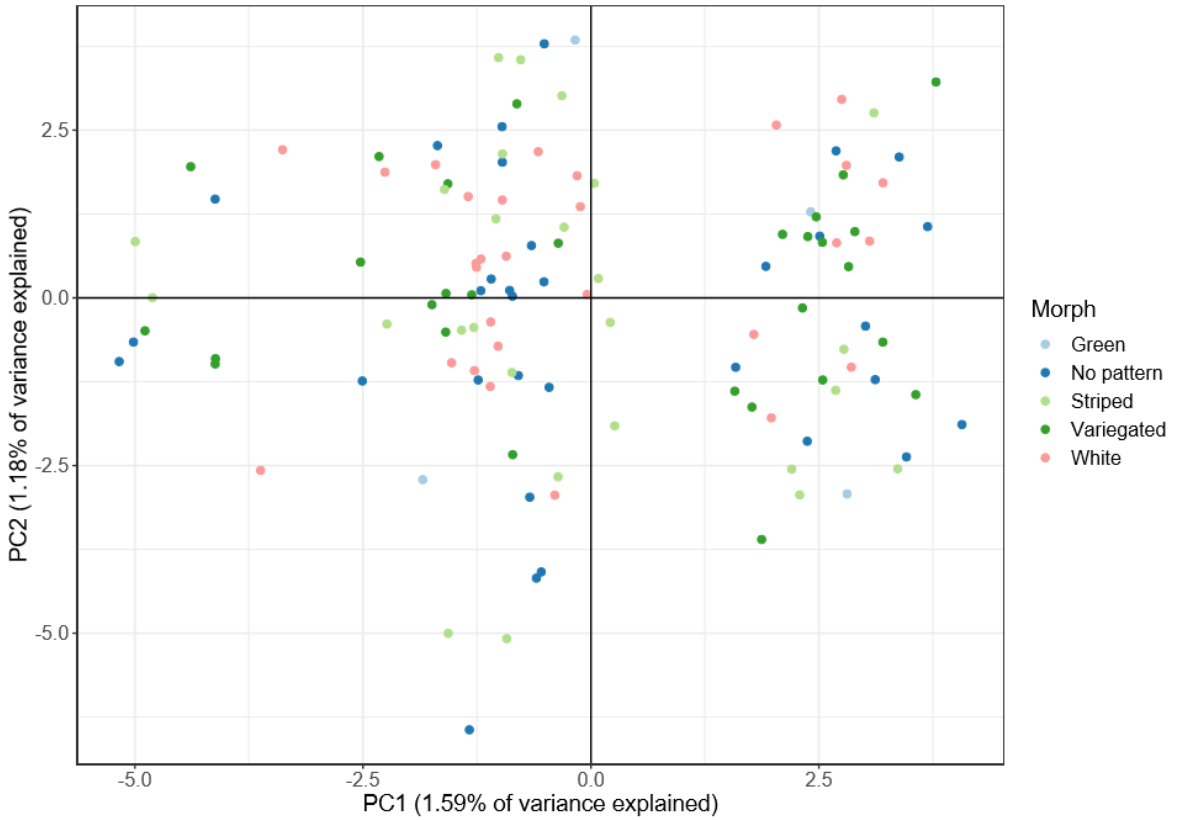


Figure 3.5 Principal Component Analysis of Auckland populations coloured by morphotype. No substructuring or stratifying effect of colouration is observed.

3.3.4. Isolation-By-Distance and Geographic Concordance

Testing for a geographical effect using a Mantel test and a redundancy analysis indicated that geography affected population structure. A mantel test showed a significant correlation between Slatkin's linearized Fst and geographic distance (r = 0.83, P = 0.001) (Fig. 3.6). However, this relationship was primarily driven by the populations pairs excluding Wellington or Kaikōura. All points greater than 0.15 linearized Fst are pairwise comparisons that include at least one of either Wellington or Kaikōura. An additional Mantel test was performed excluding Wellington and Kaikōura, this gave a Mantel statistic of 0.92 (P = 0.001), indicating a very strong correlation between genetic and overwater distance within the Northern group).

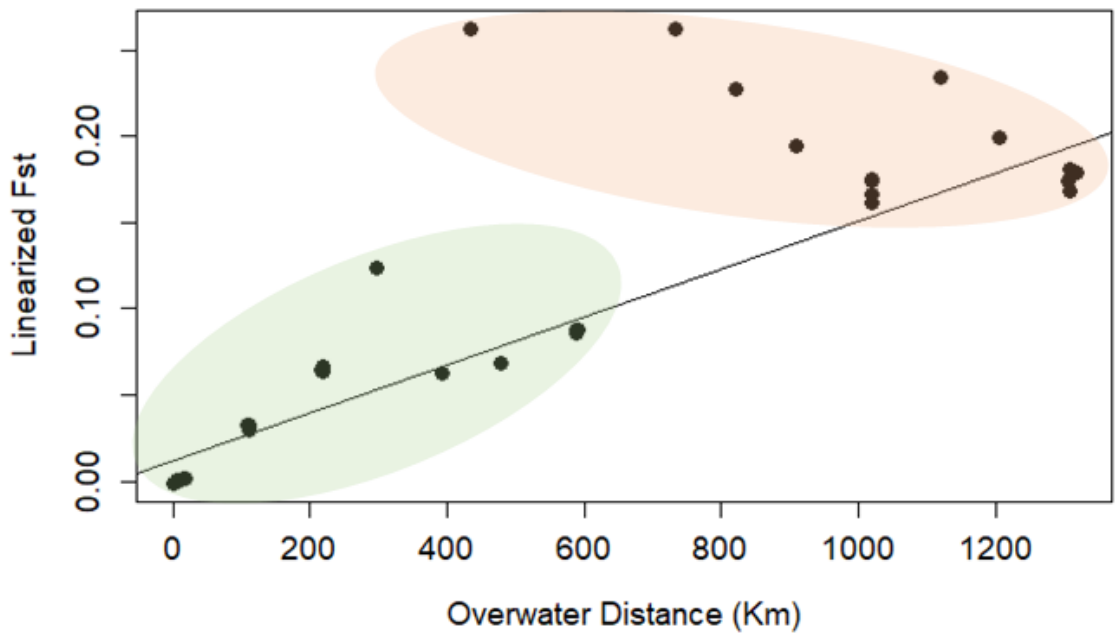


Figure 3.6 Correlation between overwater distance and Slatkin's linearized Fst. The green oval indicates pairs of populations within the North group, while the orange oval indicates pairs of populations that include the South group. These ovals are added post-hoc. The black line indicates the slope of relationship for just the North group.

A redundancy analysis (RDA) showed that 5.6% of the genetic variation was driven by the spatial arrangement (P = 0.001) (Table 3.6). While the majority of the variation was due to other untested parameters.

Table 3.6 Results of the redundancy analysis, constrained variance is due to the distance between populations, while unconstrained is from untested variables.

	Inertia	Proportion	Rank	Proportion of Genetic Variance Explained
Total variance	28.24	1		1
Constrained variance	18.32	0.65	2	0.056
Unconstrained variance	9.92	0.35	6	0.944

3.3.6 Structure Analysis

Following the Isolation-By-Distance and geographic analysis, I implemented a STRUCTURE analysis. This method implements a clustering approach, which probabilistically (in a Bayesian framework) assigns individuals to one or more populations under an admixture model which allows different proportions of an individual's genome to have arisen in different populations (Pritchard et al., 2010). STRUCTURE analyses showed similar results to the PCA, with all Auckland locations consistently clustering together, while Kaikōura and Wellington also always clustered together across all ranges of K (Fig. 3.7).
As K was increased, genetic clusters started to fall out between Opito Bay and Māhia Peninsula. When K=3, Māhia formed a distinct cluster with high admixture with Mt. Maunganui. When K was increased to 4 Māhia fell out as mostly distinct, and Mt. Maunganui and Opito Bay appeared distinct. At K = 4, as with K=5, genetic admixture was evident between adjacent populations with the exception were except Wellington and Māhia, while when K=6 this pattern fell out and suggested some admixture between Opito Bay and Wellington, likely the result of overfitting. At values of K > 6 I observed increasing noise, which obscured patterns within the data (see Supp. Fig 1).

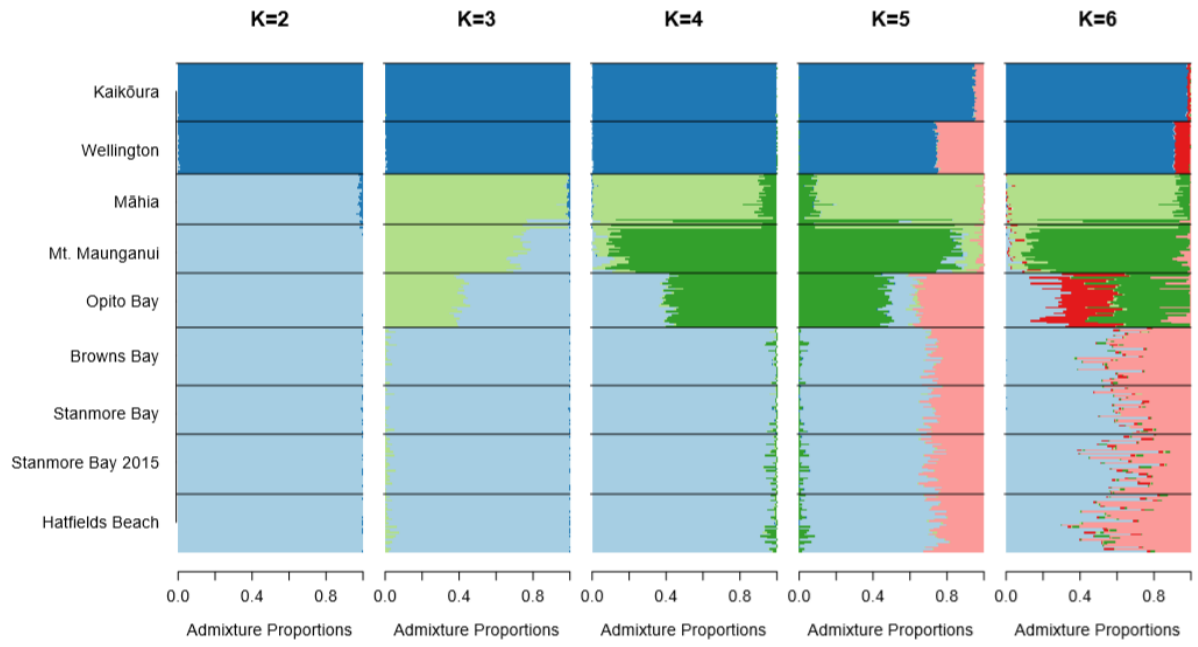


Figure 3.7 Admixture barplots based on Bayesian clustering analyses generated in STRUCTURE and consolidated in CLUMPP for a range of values of K.

3.3.7 Analysis of Molecular Variance

Analysis of Molecular Variance indicated a high degree of population structure, but no significant variation could be attributed to morphotype. These results (Table 3.7) suggested that the majority of genetic variation is between samples (56.23%), followed by regions (28.91%), while between population within region variance was just below 5%. Variation between morphotype within population was non-significant, indicating there was no evidence of population stratification by colour morphotype.

Table 3.7 Results of the Analysis of Molecular Variance (AMOVA), regions are defined as the north group (all populations excluding Kaikōura and Wellington) and south group (Kaikōura and Wellington).

Source of Variation	df	Percentage of Variation	P-Value
Within samples	261	56.23%	<0.001
Between samples within morphotype	252	10.1%	<0.001
Between morphotype within populations	33	0.098%	0.2
Between populations within region	7	4.65%	<0.001
Between regions	1	28.91%	0.028

3.3.8 Migration Analysis

Migration analysis supported the results obtained by previous methods, and revealed some population structure, but also suggested low migration between most populations.
These estimates were based on Bayesian coalescent analysis in BayesAss3 and indicate three major groups among the nine populations: Auckland based populations, East Coast populations, and the Southern population found in other analyses. These estimates indicated no migration in or out of the Southern population, but some negligible migration into Wellington from Kaikōura (Fig. 3.8 and Table 3.8).

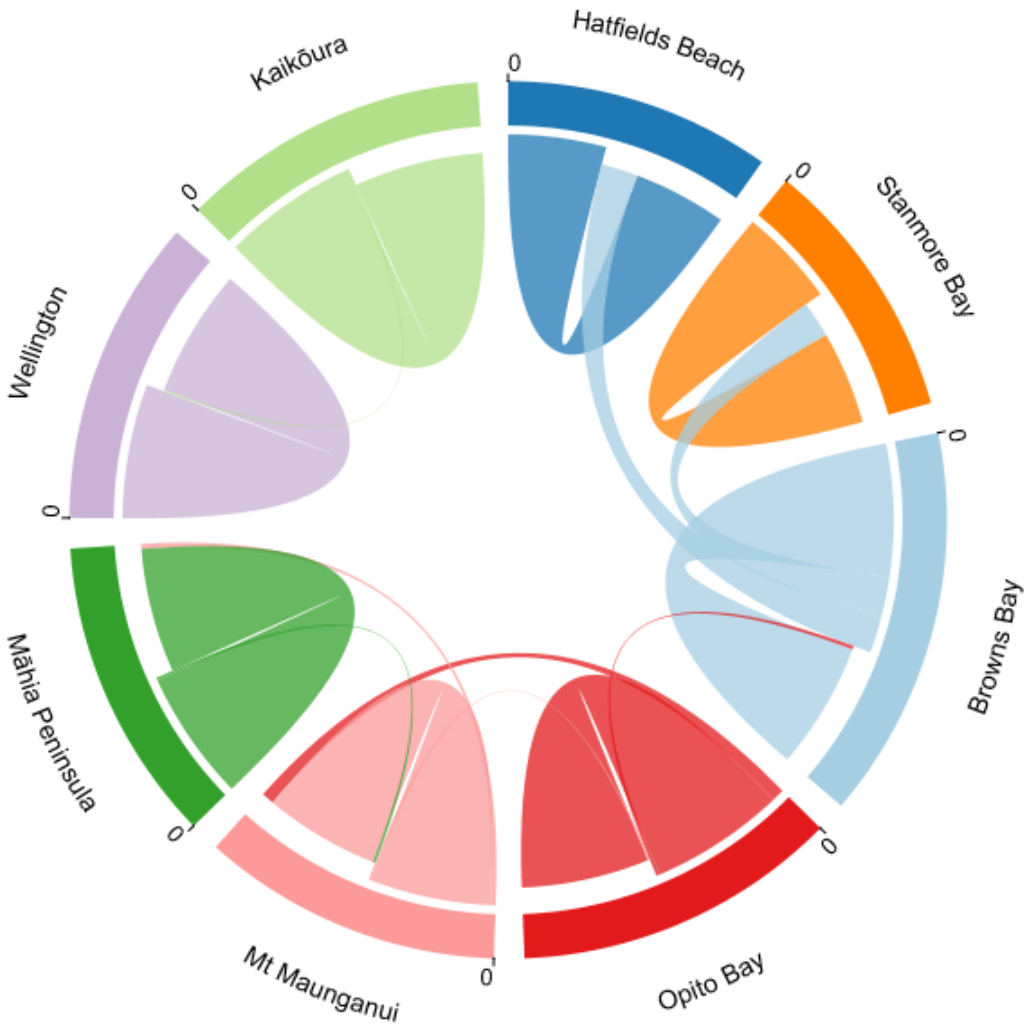


Figure 3.8 This chord diagram displays migration estimates between each population, these estimates represent the proportion of individuals in a population that are estimated to be migrants. The largest arc for each population indicates within population movement, such as a self-recruitment, while the smaller arcs are indicative of between population movement or migration. As a result, wider arcs indicate higher migration rates.

Within the Auckland populations I observed high migration at approximately 0.25. Among the populations along the East Coast, migration estimates were consistently higher among adjacent populations than between non-neighbouring populations, suggestive of a stepping stone model of migration (Kimura & Weiss, 1964). These migration rates were relatively low - generally between 0.01 and 0.02 (Table 3.8).

Table 3.8 - BayesAss migration estimates (as proportion of migrant individuals in population). Values in bold are those used in the final analyses and used to create figure 3.8. The upper triangle represents migrant from the population on the left into the population on the top.

	Hatfields Beach	**Stanmore Bay**	**Browns Bay**	**Opito Bay**	**Mt Maunganui**	**Māhia Peninsula**	**Wellington**	**Kaikōura**
Hatfields Beach	**0.6749**	0.0082	**0.2673**	0.0083	0.0082	0.0083	0.0082	0.0082
Stanmore Bay New	0.0092	**0.6761**	**0.2583**	0.0093	0.0093	0.0095	0.0095	0.0094
Browns Bay	**0.0083**	**0.0082**	**0.9166**	**0.025**	0.0084	0.0084	0.0082	0.0084
Opito Bay	0.0087	0.0087	0.0089	**0.9248**	**0.0138**	0.0088	0.0087	0.0088
Mt Maunganui	0.0096	0.0095	0.0095	0.0904	**0.8333**	**0.0192**	0.0094	0.0095
Māhia Peninsula	0.0094	0.0093	0.0093	0.0095	**0.0369**	**0.8978**	0.0092	0.0092
Wellington	0.0088	0.009	0.0091	0.0092	0.009	0.0091	**0.926**	**0.0108**
Kaikōura	0.0084	0.0083	0.0082	0.0083	0.0082	0.0085	0.0082	**0.9335**

3.3.8 Dispersal Distance Estimates

The average dispersal distance between an adult and its offspring was calculated using the methods outlined by Puebla et al. (2009). Using this approach, with the slope of the regression of Slatkin's linearized Fst against geographic distance ($2x10^{-4}$, $p = 1x10^{-12}$) and the methods outlined in section 3.2.2 - Estimates of Dispersal Distance, I estimated that the average dispersal distance was 116m.

I confirmed that the spatial scale was appropriate for this observation by calculating the appropriate spatial scale over which this model should be robust, which was determined to be 1240km. This is greater than the spatial scale in my model, which means the analysis for dispersal distance should be relatively robust.

3.4 Discussion

My results have provided new insights into the population structure and connectivity in *Isocladus armatus*, a marine isopod endemic to New Zealand. This is one of the first studies to employ a Genotyping-By-Sequencing (GBS) approach to a marine invertebrate in New Zealand. Genotyping-By-Sequencing has the capacity to detect population structure with high precision, which can be particularly useful for understanding structure on a relatively small spatial scale. I use a large panel of SNPs (8,020 loci) to understand population structure with high resolution over both large and small spatial scales.

Using this method, I was able to calculate relatedness estimates for individuals within each population. Using a triadic maximum likelihood method, I found that all estimates were not significantly different to zero. This was because the estimated relatedness minus the standard deviation was always less than zero. This suggests that *I. armatus* has very large effective population sizes, such that I would expect to see low inbreeding. Indeed, this observation is not surprising, because marine organisms often have very large effective population sizes, especially for species with high dispersal capacity (Hellberg, 2009).

3.4.1 Spatial Structure

Small Spatial Scale (<20km) and temporal variation

Within Auckland, the populations of *I. armatus* exhibited very little, or no, structure. This was on a very small spatial scale (distance between furthest populations was 20 km). However, a temporal aspect was also included in the dataset. Kaikōura, Hatfields Beach and one of the populations from Stanmore Bay were collected and processed in 2015, while Browns Bay, and a replicate Stanmore Bay sample set were collected in mid-2018. This extended sampling timescale has the possibility to influence my dataset and could potentially confound inferences made that incorporate Kaikōura into analyses. To test this, I retained both Stanmore Bay populations in initial analyses to see if there was any significant genetic differentiation between each timepoint. No differences between the 2015 and 2018 Stanmore Bay samples were observed, and thus both datasets were retained in further analyses.

Due to attempting to obtain equal numbers of each morphotype, our sampling design was not random within populations. Therefore, it is possible that morphotype could lead to some population subdivision which could influence estimates of population connectivity. However,

this does not appear to be the case as the AMOVA suggested no significant variation due to colour. Additionally, the PCA indicated no effect of colouration.

Across all Auckland populations I observed no significant population differentiation based on PCA, and STRUCTURE analyses, and only very low structure between Browns Bay, and Stanmore Bay and Hatfields Beach based on Fst analyses. However, the low structure indicated with the Fst analyses is in line with other studies of Auckland populations of marine organisms, such as *Evechinus chloroticus* (a native sea urchin) (Nagel, Sewell, & Lavery, 2015), as well as an invasive tunicate - *Styela clava* (Goldstien, Schiel, & Gemmell, 2010). However, these species both have a pelagic larval phase, which is expected to contribute to reduced genetic heterogeneity of populations. In contrast, I expected some population structure within *I. armatus* due to its direct development life history. This is because direct-developers, generally have low dispersal potential relative to biphasic species (Puritz et al., 2017).

The similarity between *I. armatus*, and those previously described species may suggest that at least within the Hauraki Gulf, a meta-population dynamic may exist. The many islands within the gulf may contribute to high levels of migration between samplings localities resulting in population structure consistent with an island model of migration, as opposed to the stepping stone model hypothesized for the populations on the east coast (see below).

Migration estimates for Auckland populations between BayesAss iterations were consistently high. Interestingly these estimates suggested migration from Browns Bay into Hatfields Beach and Stanmore Bay, with no migration between the latter two populations. This could be because Hatfields Beach and Stanmore Bay are the same population (as indicated by all other analyses), or because my sample size was not sufficient to detect extremely fine grain structure. Because BayesAss3-SNPs estimates the proportion of individuals in a population that are migrants, in two populations with the same allele frequencies (such as a population being incorrectly subdivided) BayesAss3-SNPs may struggle to identify migrant individuals. Because Browns Bay shows a small amount of differentiation this method may still work for this population. This result points to the importance of incorporating prior knowledge into these analyses and making use of multiple approaches to resolving population structure.

Large Spatial Scale (>20km)

PCA, STRUCTURE, and Isolation-By-Distance analyses supported the presence of a North-South split. This split indicated that the North group included all populations excluding

Kaikōura and Wellington, while these two populations formed the South group. Analyses of the North group indicated a strong pattern of Isolation-By-Distance, which was supported by the Mantel test and redundancy analysis - which both suggested an effect of distance on the genetic structure of these populations.

Principal Component Analyses supported this result, as I observed a strong relationship between PC distance and geographic distance - this is shown in Fig. 3.6, where geographically adjacent populations tend to be more genetically similar, but also that a relationship between the geographic space and the PC genetic space also exists (Fig. 3.4). The PCA separated out all populations (except those within Auckland) into separate clusters, with the second PC delineating populations in the north group, and third PC delineating populations in the south group (Fig. 3.3).

Isolation-By-Distance within the North group was strongly supported by both a Mantel test which tests for correlations between genetic and spatial distance (Fig. 3.6), and a redundancy analysis which attempts to partition the variation observed between constrained variables (in this case spatial coordinates) and unconstrained variables (factors that have not been tested). The Mantel test suggested a very strong correlation between genetic and overwater distance, however a redundancy analysis suggested that only 5.6% of the genetic variation in my dataset was due to the spatial structure. The AMOVA indicated that the largest proportion of variation (after variation within samples) was between the North-South break. Together, these results suggest that while distance plays a clear role in population structuring in *I. armatus*, other factors (such as Isolation-By-Environment) likely also play a role.

Because the highest migration rates along the East Coast were between adjacent populations, they indicate a pattern of Isolation-By-Distance characteristic of a stepping stone model of dispersal. However, these results also showed higher migration estimates with the direction of the prevailing ocean currents, suggesting a role of ocean currents in dispersal for *I. armatus* - such as rafting (Fig. 3.9).

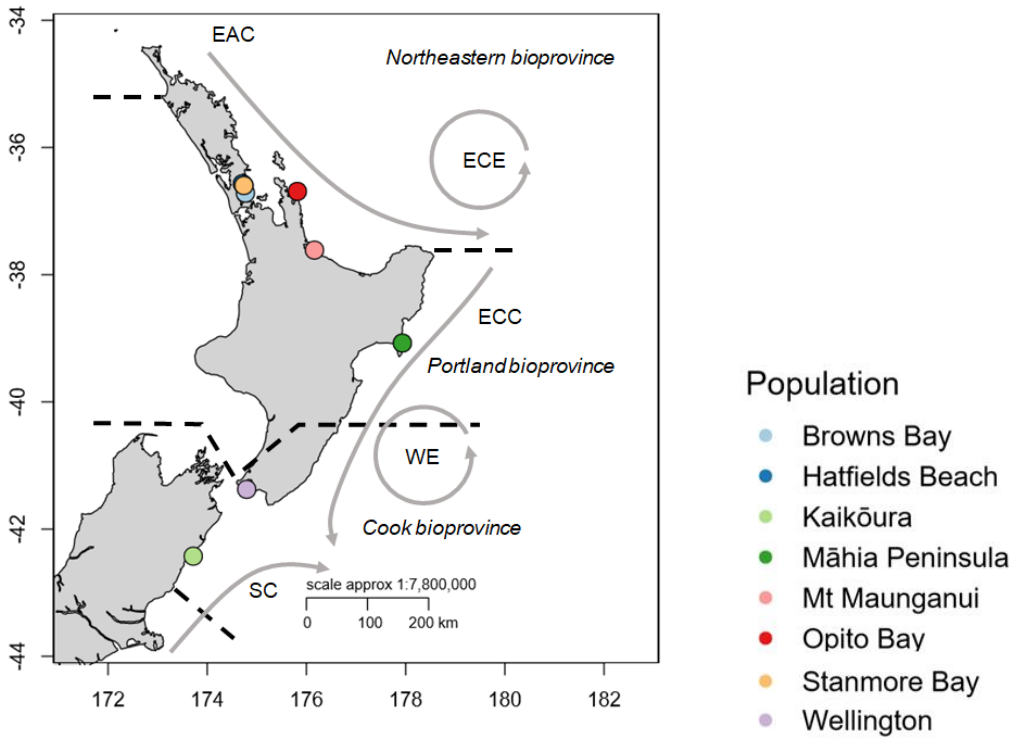


Figure 3.9 - Map of sampling localities (coloured dots), alongside prevailing ocean currents and biogeographic breaks proposed by Shears et al (2008). Black dashed lines and unabbreviated labels indicate bioprovinces. Grey arrows indicate currents, which are abbreviated as follows - EAC is the East Auckland Current, ECE is the East Cape Eddy, ECC is the East Cape Current, WE is the Wairarapa Eddy, SC is the Southland Current.

Within the North group, I was able to use the strong Isolation-By-Distance relationship to estimate the average adult-offspring dispersal distance (section 3.3.8). This suggested that the average parent-offspring dispersal distance is approximately 116m. This result is consistent with previous research that described *I. armatus* swimming in the incoming tide (Jansen, 1971). Therefore, tidal displacement of individuals (particularly small juveniles) may explain this dispersal distance. The above analysis assumes a linear and continuous distribution of *I. armatus* across the range of the calculation and is therefore only indicative of the average dispersal distance in an idealized system. As such, it likely underestimates dispersal distances because many populations of *I. armatus* will be intersected by gaps of > 116m. In order to gain a better estimate of average parent-offspring dispersal distance, it will be necessary to work within a site that better suits the assumptions that this model makes, or develop a more complex model that accounts for non-continuous distributions, or to use a less restrictive model. Additionally, this model assumes the effective population size, mutation rates, and population densities are all reasonably accurate - therefore these results should be interpreted with caution. These results, unsurprisingly, indicate that long distance

dispersal of *I. armatus* is relatively rare across a linear distribution. These results are consistent with expectations for direct developing species (Puritz et al., 2017).

Within the South group I found directional migration from Kaikōura into the Wellington population. However, this gene flow was relatively low and on a scale similar to that observed between between Opito Bay and Mt. Maunganui. These results were interesting as BayesAss-3 estimates contemporary migration rates - indicating somewhat frequent migration between populations. Concerns have been raised over the ability to accurately estimate migration, and to distinguish between contemporary and historical migration because marine populations are rarely at equilibrium (Marko & Hart, 2011; Slatkin, 1985). This should be less problematic for BayesAss estimates because BayesAss explicitly avoids a reliance on populations existing in equilibrium (Wilson & Rannala, 2003).

The Kaikōura samples were collected shortly prior to a large earthquake (7.8 Mw, 14/11/2016 ("GeoNet M 7.8 Kaikōura Mon, Nov 14 2016," n.d.)) while the Wellington samples were collected a few years after this earthquake. Although the resulting uplift may have had population level consequences, I suggest that the impact is likely minimal. This is because the effects would be more pronounced if the samples had been taken after the earthquake or if the uplift had occurred in Wellington because recolonisation in Wellington would have occurred, leading to greater genetic variation within this population, which has been reported at Kaikoura after the earthquake (Peters et al., 2020). Regardless, these results suggest that the Cook Strait may not play a significant role in isolation of these populations as I observed similar migration rates between Wellington and Kaikōura as I did for other pairs of populations with similar or smaller distances between them, this is also consistent with other research within New Zealand (Veale & Lavery, 2012; Arranz Martinez, 2018).

North-South Division

My results suggest there may be a North-South division between Māhia and Wellington that could be the result of factors beyond Isolation-By-Distance.The AMOVA suggested that the second biggest source of variation (after individuals) is this North-South break and it accounts for 28.9% of the total variation in my data (Table 3.7).

Barriers to gene flow across the East Coast of New Zealand have been relatively well studied, and research on larval developers suggests a genetic break near East Cape (Fig. 3.9) e.g. *Haliotis iris* (pāua) (Will, Hale, Schiel, & Gemmell, 2011) and *Buccinulum vittatum*

(Gemmell et al., 2018). This barrier has also been shown to affect direct developers - such as *Actinia tenebrosa* and two species of amphipods (M. I. Stevens & Hogg, 2004; Veale & Lavery, 2012). However, in contrast to these other direct developers, this break does not appear to affect *Isocladus armatus*, as the migration estimates and Fst between Mt. Maunganui and Māhia were no greater than migration estimates and Fst between the north group populations.

Surprisingly high estimates of migration between the North populations were observed, which was supported by the observation of three highly admixed individuals. Two of these samples were from Māhia but were admixed with Mt Maunganui according to both the STRUCTURE and PCA analyses, the other individual exhibit the same pattern except with the populations switched. This result is particularly surprising for two reasons 1) the presence of a strong biogeographic barrier near East Cape that was anticipated to play a role in reducing gene flow between these populations, and 2) while the direction of migration was stronger in the direction of the ocean currents, these results suggest unusually high bidirectional migration between these two populations. The direction of the East Cape Current may enable distribution of *I. armatus* into Māhia from northern populations, while distributing individuals further south, away from Wellington. This is consistent with my observations as I found greater migration from Mt. Maunganui into Māhia than vice versa. In addition, gene flow against prevailing ocean currents has been well modelled recently and suggests that's while ocean currents may play a role, near surface movements (such as stokes drift) can contribute to migration patterns more than ocean currents (Fraser et al., 2018).

Because the distance between Māhia and Wellington is greater than the distance between any other pair of adjacent populations I expected to see a large degree of genetic distance between these two populations, even in the absence of a geographic barrier. Therefore, it is difficult to disentangle the effects of a barrier to gene flow from the effects of distance between these populations. However, while aspects of the Isolation-By-Distance indicated the possibility of a divide, some aspects also appeared to be a continuation of the linear relationship observed within the Northern group. If a split is present (as indicated by the PCA, BayesAss, and STRUCTURE analyses) then it may be permeable rather than a sharp or hard boundary. This split could be attributable to several different processes, such as 1) the presence of a permeable barrier to gene flow resulting in rare dispersal between the groups (Cowman & Bellwood, 2013), or 2) the presence of a cryptic species (Wells & Dale, 2018). It is highly unlikely that this relationship is due to the lack of suitable habitat over any appreciable distance between Māhia and Wellington because there are many known

populations between these two regions (Hurley & Jansen, 1977; “iNaturalist.org,” n.d.; O. Wade, Pers. Comm).

I found that the pair of populations with the greatest dissimilarity was Welington and Māhia (Fst between these populations was 0.2, while Fst between Māhia and Stanmore Bay was 0.079 despite these pairs being similar distances apart). This result is consistent with some other work which suggests that a break may occur near Māhia which predominantly affects species with low dispersal capacity (such as species that have a very short, or complete lack of a larval stage) (Arranz Martinez, 2017).

This result is also consistent with the biogeographic regions proposed by Shears et al. (2008) (Fig. 3.9) though within these regions my results were inconsistent with the above study due to the proposed placement of a biogeographic region border at East Cape. The clustering of Wellington and Kaikōura populations is also consistent with the observations made by Veale and Lavery (2012). Possible seascape features between Māhia and Wellington that could act as a biogeographic break are the strong currents found in the Wairarapa Eddy, or the East Cape Current. While these results are consistent with the previous observations of biogeographic barriers within this region, the relative strength or effect of these barriers are difficult to determine in this study.

Alternatively, the presence of a cryptic species could also account for the observed patterns. Cryptic species have previously been observed within intertidal organisms, as well as within isopods (Leese, Kop, Wägele, & Held, 2008). Within New Zealand, strong North-South divergence between populations has previously been suggested to be the result of cryptic speciation in the New Zealand Brooding Brittlestar (Sponer & Roy, 2002). The degree of genetic divergence between the North and South group are also similar in scale to the divergence found between other cases of cryptic species within *Isopoda* (Leese et al., 2008). Therefore, this hypothesis represents a strong possibility. The potential for cryptic species is further supported by the observation of a single individual from Browns Bay (BBFA3, which was excluded from all analyses) which, despite appearing morphologically similar to *I. armatus*, lacked 93% of SNPs that were present in other samples. While missing data is a common feature of reduced representation datasets, an excessively high missingness can be problematic. Missingness refers to loci that lack sequence data for some individuals, either the result of failed amplification due to technical issues, a lack of cutting sites in the individual, or lack of the polymorphism. Some research has shown that very high missingness in RADseq data may be associated with divisions between species rather than populations (Tripp, Tsai, Zhuang, & Dexter, 2017)). These results therefore suggest that our

sample with high missingness may represent a cryptic species. This possibility should be included for consideration in future studies, from both a morphological and genetic perspective.

3.4.2 Conclusion

Isocladus armatus exhibits a surprisingly high amount of gene flow between populations at spatial scales <20 Km. However, at distances greater than 20 Km I find that the level of gene flow is consistent with the expectations regarding direct developing species. My results suggest that direct developers may rely on rafting for migration and may exhibit Isolation-By-Distance patterns of population connectivity (Teske et al., 2007).

My results suggest that population structure may arise at a spatial scale of between 5 and 15 km, because sites 15 km apart showed low but significant genetic differentiation based on Fst analyses. On a large spatial scale, I found that migration directions appeared to be stronger in the direction of prevailing ocean currents - suggesting a role of a rafting in dispersal for *I. armatus*. Additionally, a strong pattern of Isolation-by-Distance became evident, suggesting that on larger spatial scales dispersal of *I. armatus* is consistent with expectations regarding dispersal capacity for direct developers.

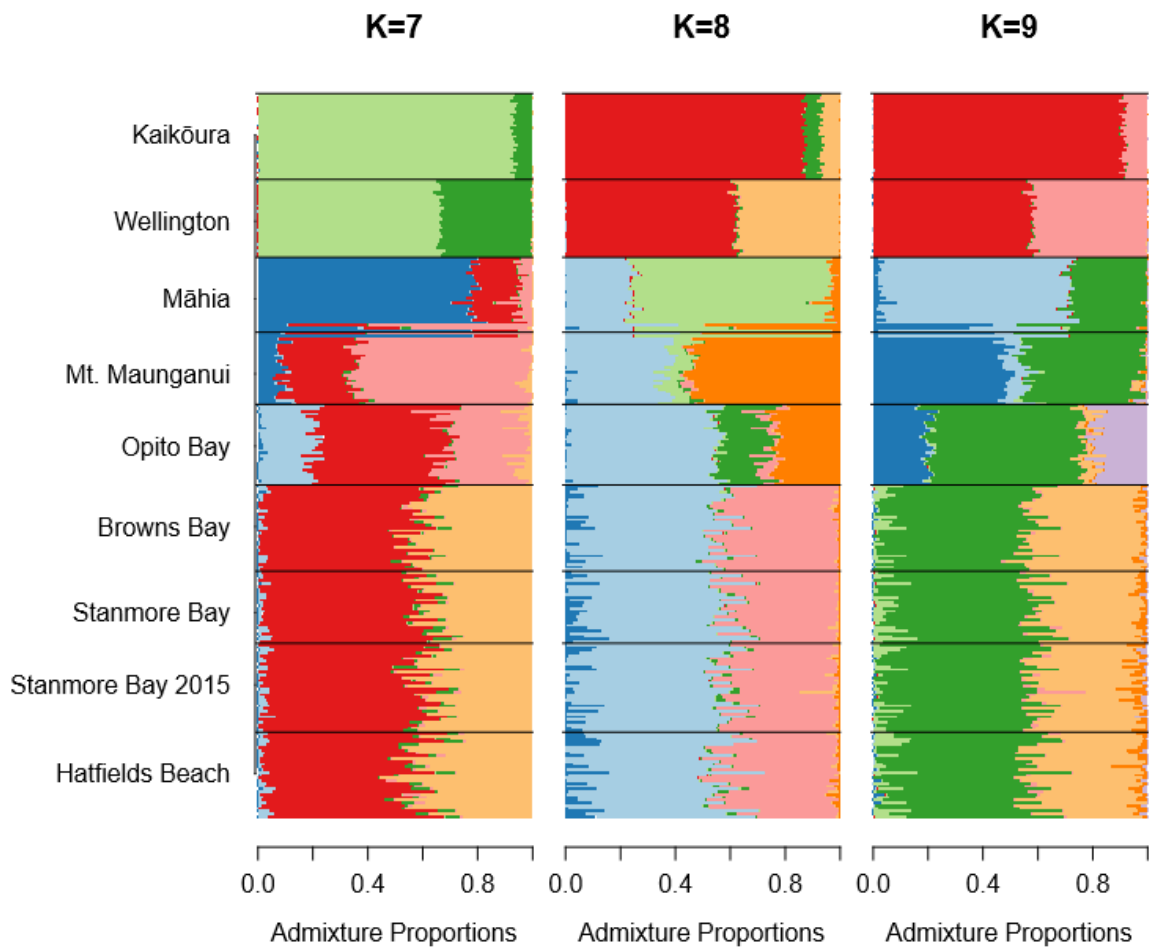
I observed evidence for relatively high migration between populations separated by a well-described biogeographic break. The lack of effect of this break was surprising and indicated higher migration over this break than between many other pairs of populations. Some evidence was found that suggested a break may affect this species, however this evidence suggested the break was likely located between Māhia peninsula and Wellington. Alternatively, a cryptic species hypothesis may explain the genetic dissimilarity between these populations, based on the possibility of a single individual of a cryptic species being sampled in Auckland, and the scale of dissimilarity between Māhia and Wellington was similar to other cryptic species in *Isopoda* (Leese et al., 2008).

3.4.4 Future Work

Future work is necessary to further elucidate the nature of the relationships between the North and South groups. This work should specifically focus on identifying the position of the North-South break by sampling between Māhia and Wellington across regular small intervals. This should incorporate known populations at Castlepoint and Cape Kidnappers, and ideally be performed using the same methods described in this study. This approach

would enable inclusion of this dataset into future works to provide a much more comprehensive understanding of population structure within *Isocladus armatus*.

Supplementary



Supplementary Figure 1 - Admixture barplots generated through STRUCTURE and CLUMPP for values of K above 6. There is a relatively large amount of noise in these values, as largely evidenced by the apparent admixture between Kaikōura and Opito Bay at K=8.

Chapter 4 - Colour Polymorphism in *Isocladus armatus*

Abstract

Colour Polymorphism is widespread across *Animalia*, has been well studied across *Crustacea*, *Aves*, and *Insecta*. The adaptive function of this polymorphism varies across *Animalia* but is often thought to play a role in predator avoidance. In *Isopoda* it is commonly attributed to crypsis, however the specific mechanisms for expression and selection for this trait are relatively poorly understood. This is further complicated by cases of both diet-based colouration influencing crypsis and genetic based colouration. In this study I use genome wide association approaches to understand the genomic basis of colour polymorphism, and to infer the selective mechanism for this trait. I found that SNPs associated with Colour Polymorphism exhibited signatures of disruptive selection. I propose that substrate heterogeneity in *Isocladus armatus'* habitat results in microhabitats, each of which imposes a selective pressure benefiting a specific morph type. The size of these microhabitats is so small that high levels of interbreeding between these microhabitats, and thus between morphs, results in the maintenance of polymorphism across the population.

4.1. Introduction

Colour polymorphism (CP) is the co-existence of at least two discrete colour morphs within a population, for which the presence cannot be attributed to recurrent genetic mutations, and where each morph has a genetic basis (Huxley, 1955). By this definition, reversible colour change or polyphenic responses are separate and distinguishable from CP. Although CP does not necessitate the presence of continuous variation in colour within a population, continuous variation may exist alongside discrete distinguishable morphs. However continuous variation can be maintained separately of CP and through different mechanisms - such as in birds where sexual selection and diet can play a role in the maintenance of colouration (Galeotti et al., 2003; Hill, Inouye, & Montgomerie, 2002). CP has been found and studied widely across both plants and animals. Within animals it has particularly well studied in vertebrates such as fish (i.e cichlids (Kusche, Elmer, & Meyer, 2015)) and birds (i.e Gouldian finch (Kim et al., 2019)) (Galeotti et al., 2003; Wellenreuther, Svensson, &

Hansson, 2014). CP has also been researched, albeit to a lesser extent, in insects such as *Timema cristinae* (a stick insect), and ladybugs (Gautier et al., 2018; Nosil, 2007).

While CP is defined as having a genetic component, there may also be interactions between genes and the environment. Although the physiological mechanisms responsible for colouration are relatively well understood in many organisms and model systems, these mechanisms vary with taxa. This research focuses specifically on colouration in *Crustacea* and *Isopoda*. Additionally, while CP is defined by having a genetic basis, variation in colouration itself is not necessarily genetic and can be based on environmental factors such as diet and interactions between genetics and the environment. For example, variation in carotenoid based colouration has a basis in diet, social status, and individual fitness (Sefc, Brown, & Clotfelter, 2014). In some of these cases it has been shown that carotenoid acquisition and metabolism are associated with ornamentation and colour polymorphism (Andrade et al., 2019). This is because carotenoid-based signalling is hypothesized to be an honest signal of individual quality as it requires diversion of carotenoids from their role as an antioxidant into a role in pigmentation (Sefc et al., 2014). This diversion into pigmentation is costly and thus requires high levels of carotenoid intake; as a result, individuals with greater colouration may be higher quality because they can better afford this cost.

Colouration can be affected by both pigmentation and expansion/contraction of chromatophores. This is the case in *Idotea granulosa*, where while colouration is largely contingent on epidermis pigmentation, rapid colour change can be affected by the chromatophores (Lee, 1966b, 1966a). Chromatophore are cells containing pigments responsible for colouration, and function by expansion or contraction resulting in translocation of the pigment within the cell, where expansion disperses the pigment to change colour. In a species of rock lobster, *Panulirus cygnus*, it is thought that both ecology (such as diet) and genetics may contribute to colouration (Wade, Melville-Smith, Degnan, & Hall, 2008). Chromatophore distribution thus may be genetically determined, while pigmentation may be environmentally determined, as is the case in *Asellus aquaticus* (Hargeby, Johansson, & Ahnesjö, 2004).

Selective Mechanisms for Colour Polymorphism

Colour Polymorphism is often hypothesised to be maintained by a form of balancing selection, which enables the persistence of multiple phenotypes within a population (T. E. White & Kemp, 2016). However, the specific mechanism behind this selection can vary. The

four primary mechanisms proposed to maintain CP within populations are: 1) balancing selection, such as negative frequency dependent selection (NFDS), 2) sexual selection, 3) disruptive selection mediated by interbreeding between phenotypes, and 4) divergent selection mediated by gene flow between populations. Specific terms for of selection used in this chapter and described in table 4.1.

Table 4.1 Definitions of forms of selection discussed in this chapter, alongside the form that BayeScan (as Fst-outlier approach) would classify it as.

Selective Term	**Definition**	**BayeScan Classification**
Divergent selection	Selection that acts to diverge populations based on contrasting directional selection between them.	Disruptive
Disruptive selection	Selection that occurs within a population, in contrasting directions that can lead to increased divergence between populations.	Disruptive
Negative Frequency Dependent Selection	A form of selection that results in maintenance of multiple phenotypes due to rarer phenotypes possessing higher fitness that more common ones.	Putatively Balancing
Directional	Selection acting a direction to increase the abundance of a phenotype due to increased fitness of this phenotype.	Disruptive
Balancing	Selection acting to maintain polymorphism within a population.	Balancing
Stabilizing	Selection against extreme phenotype, resulting in stabilizing frequency of non-extreme phenotypes.	Balancing
Sexual selection	Selection for traits that result in greater reproductive success.	Depends on specific mechanism

Negative frequency dependent selection (NFDS) is a form of balancing selection, where rarer forms (i.e shape or color) have a fitness advantage. If NFDS maintains CP, then temporal variation in morph frequency may be observed due to variation in morph fitness (Takahashi & Kawata, 2013). This is because NFDS results in rare morphs producing more offspring - reducing that morphs rarity and increasing rarity of another morph, this change

would result in maintenance of each morph type through oscillations in frequency (Fig. 4.1). There are many potential adaptive benefits for this, i.e in apostatic selection (a form of NFDS mediated by predation) rarer morphs may have greater fitness because predators are unable to form cognitive search images for them.

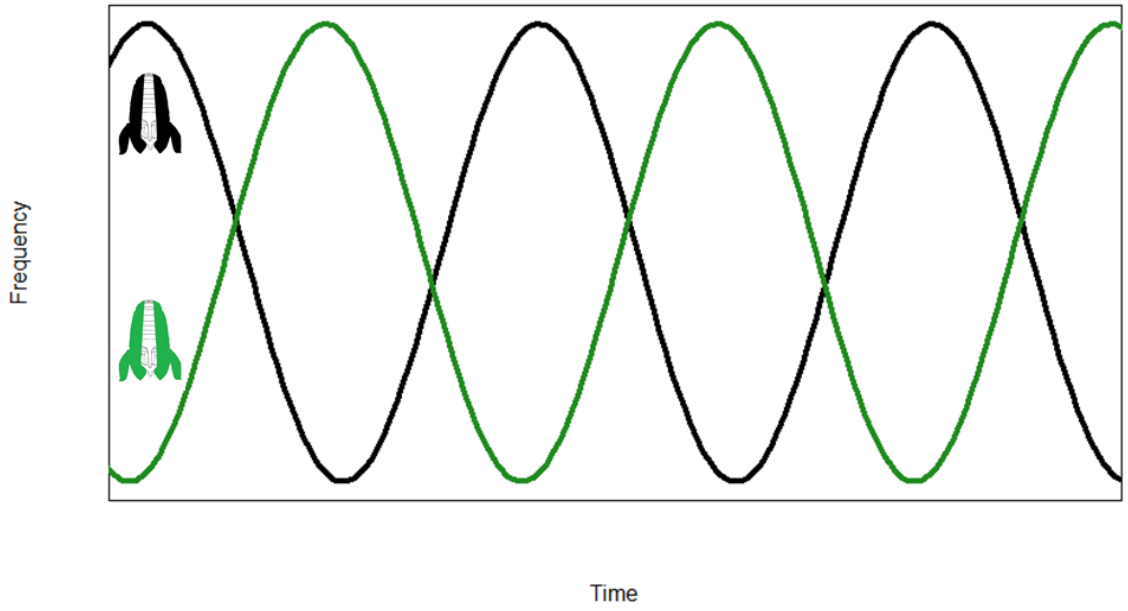


Figure 4.1 A simple example of the maintenance of multiple phenotypes through Negative Frequency Dependent Selection, time is on the X-axis and simulated frequency of the two morphs is on the Y-axis.

Alternatively, sexual selection can follow a NFDS pattern. For example, in the common bluetail, *Ischnura senegalensis*, NFDS has been postulated to be the mechanism behind the maintenance of female colour polymorphism (Takahashi, Yoshimura, Morita, & Watanabe, 2010). This was evidenced by the presence of oscillations in frequency of females morphs that were consistent with a mathematical model of morph frequencies under negative frequency-dependent selection. The biological mechanism behind this is thought to be increased costly male harassment of the more common morph. This harassment has a significant effect on the number of eggs laid by the female, and as such, morph frequency is expected to oscillate under selective pressures of male harassment. This case is a simple one, as the complexity of selection is increased by the number of morphs present. This is because genetic interactions (such as intermediate morphs) may be under selection, as well as fitness interactions between morph types such as situations where morph fitness can be treated as a zero-sum game (Kokko, Griffith, & Pryke, 2014).

Disruptive selection is a form of selection where, within a population, extreme phenotypes experience greater fitness than intermediates experience (Rueffler et al., 2006). This leads to selection for extremes and can lead to sympatric speciation events. However, interbreeding between phenotypes alongside disruptive selection can result in the maintenance of polymorphism without speciation. In the example below disruptive selection leads to the separation of the two phenotypes due to selection for the extremes within the phenotype range (Fig. 4.2). In the absence of interbreeding between phenotypes, and in a closed population, this could lead to sympatric speciation across phenotype boundaries (Rueffler et al., 2006). Although if this interbreeding between fitness peaks is sufficient to overcome the effects of selection, we would see the retention of both morphs within the population. It is difficult to distinguish between disruptive selection and divergent selection with gene flow because they represent essentially the same process, but on different scales: while disruptive selection operates within a population, divergent selection operates between populations. They do however, share in common a requirement for gene flow (at different scales). Disruptive selection requires gene flow between morphs within a population, while divergent selection requires gene flow between populations (Rueffler et al., 2006).

If divergent selection is responsible for CP, isolated populations should exhibit different morph frequencies, with some morphs entirely absent or entirely dominant in some populations. However, if NFDS selection were responsible, colour morphs should be maintained in similar frequencies on average (though temporal variation may occur) amongst isolated but similar populations (similar on the basis of habitat type, dominant species, selective pressures, etc).

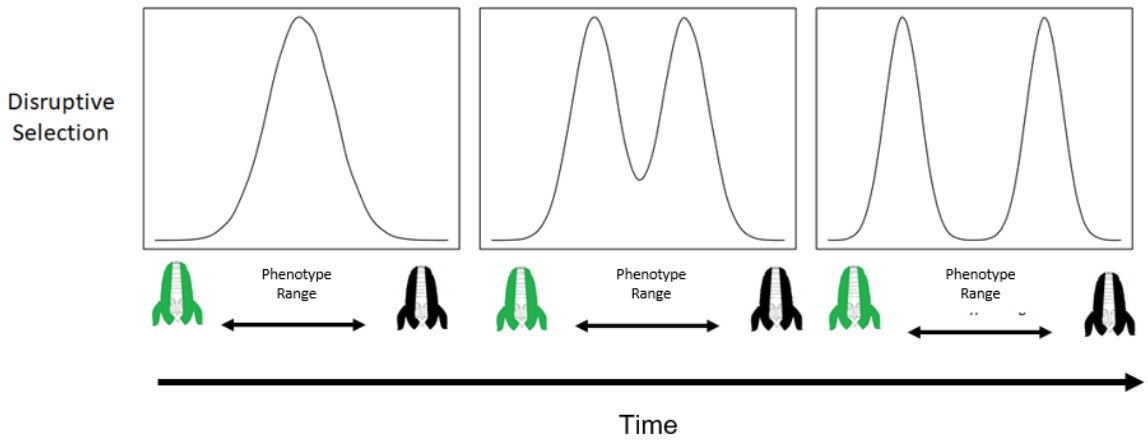


Figure 4.2 How disruptive selection works in the absence of interbreeding, each plot shows simulated frequencies of morphs at different time points.

An example of colour polymorphism maintained by divergent selection mediated by gene flow, is that of *Timema cristinae* (Nosil, 2007), a walking stick native to North America, that lives on a variety of host plants. This species exhibits two morphs – one with dorsal stripe, and one without. The frequency of these morphs changes with host plant, as some morphs exhibit greater fitness benefits on specific host species. It would be expected then, that a population of this species would diverge along host boundaries, and lead to the fixation of one morph for each host. However, this is not the case because continued gene flow between populations on different hosts results in the maintenance of each morph in each host population (Nosil, Crespi, Sandoval, & Kirkpatrick, 2006; Nosil, 2007).

It has also been proposed that sexual selection could maintain CP as a means of status signalling (discussed above with relation to carotenoids), but could also play a role in mate recognition (Wellenreuther et al., 2014). Additionally, disassortative mating has been recognized to play a role in the maintenance of colour polymorphism (Huynh et al., 2011). Disassortative mating is when individuals of different phenotypes mate more frequently than expected by chance. These hypotheses are most relevant for species which exhibit complex social and mating behaviours, such as birds. For example in *Zonotrichia albicollis*, a strong disassortative mating pattern is observed where more than 96% of breeding pairs consist of different morphs (Huynh et al., 2011).

Colour Polymorphism within Crustacea

Among *Crustacea*, CP is widespread and potentially maintained through a range of mechanisms - however the most common hypothesis of adaptive significant is crypsis (see Table 4.2). In the shrimp *Hippolyte obliquimanus*, it has been proposed that CP is maintained through apostatic selection (Duarte, Stevens, & Flores, 2016), while in the fiddler crab it has been proposed to be maintained through mate recognition (Detto et al., 2008). Within *Isopoda* there exists a wide range of CP among many species (see Figures 4.3-4.5), however the mechanism behind their maintenance across species is relatively poorly understood. There is evidence that divergent selection maintains CP within at least one species of isopod (*Idotea baltica* (Merilaita, 2001)), but limited evidence is available for the effect of NFDS or sexual selection.

Table 4.2 Species of marine invertebrates that exhibit colour polymorphism, and their associated adaptive values and selective mechanisms. Most studies suggest that crypsis provides the greater adaptive advantage of colour polymorphism.

Selective Mechanism	**Species**	**Adaptive Significance**	**Study**
Unknown	*Isocladus armatus*	Unknown	(Jansen, 1971; Wells & Dale, 2018)
Divergent selection; Disruptive selection	*Idotea baltica*	Predator avoidance	(S. Merilaita, 1998; Sami Merilaita, 2001)
Unknown	*Hippolyte obliquimanus*	Crypsis	(Duarte et al., 2016)
Unknown	*Cancer irroratus*	Crypsis	(Palma & Steneck, 2001)
Putatively NFDS	*Carcinus maenas*	Crypsis	(M. Stevens et al., 2014)
Unknown	*Birgus latro*	Crypsis	(Nokelainen et al., 2018)
Sexual selection	*Uca capricornis*	Mate recognition	(Detto et al., 2008)
Putative apostatic selection	*Paracerceis sculpta*	Unknown	(Shuster et al., 2014)
Unknown	*Dynamene sp.*	Predator avoidance	(Arrontes, 1991)
Putative sexual selection or balancing selection	*Sphaeroma rugicauda*	Unknown	(Khazaeli & Heath, 1979)

Isopods, particularly marine shallow water species, are relatively well known for exhibiting CP (Arrontes, 1991). Within the Sphaeromatidae colour polymorphism is reasonably common – with species in the following genera all exhibiting some degree of CP: Dynamene (Arrontes, 1991; Vieira, Queiroga, Costa, & Holdich, 2016), Sphaeroma (Heath, 1979), Cymodoce (Arrontes, 1991), and Isocladus (K. Jansen, 1968). However, to date, no studies have clearly identified the adaptive significance or selective mechanisms for CP in a *Sphaeromatid* isopod. CP is also common outside of *Sphaeromatidae*, one example of this is *Idotea baltica* (family: Idoteidae), where colour morph has been shown to be related to

predator avoidance in heterogeneous environments (Merilaita, 2001). However, background colouration alone was not sufficient to explain the colour polymorphism, and molecular work suggested that this 'cryptic' colouration was counteracted by gene flow, resulting in non-cryptic morphs being maintained within the population (Merilaita, 2001).

Figure 4.3 Examples of colour polymorphisms in *Idotea baltica* (family: *Idoteidae*), images were retrieved from iNaturalist. Top left is the "Uniformis" morph, top right is "Nigrum-lineata", bottom left is "Immaculatum", and the bottom right is "Immaculatum-lineata".

Interestingly some CP phenotypes appear to be present across many isopod species - specifically the striped morph (defined by a distinct dorsal stripe extending the length of the organism from the cephalothorax to the pleon) which is present among many *Sphaeromatid* species as well as within *Idotea* - suggesting either persistence through numerous speciation

events, or regular convergence for this phenotype (Figures 4.3-4.5). The presence of this phenotype across *Isopoda* suggests that this colouration pattern likely has a strong adaptive function.

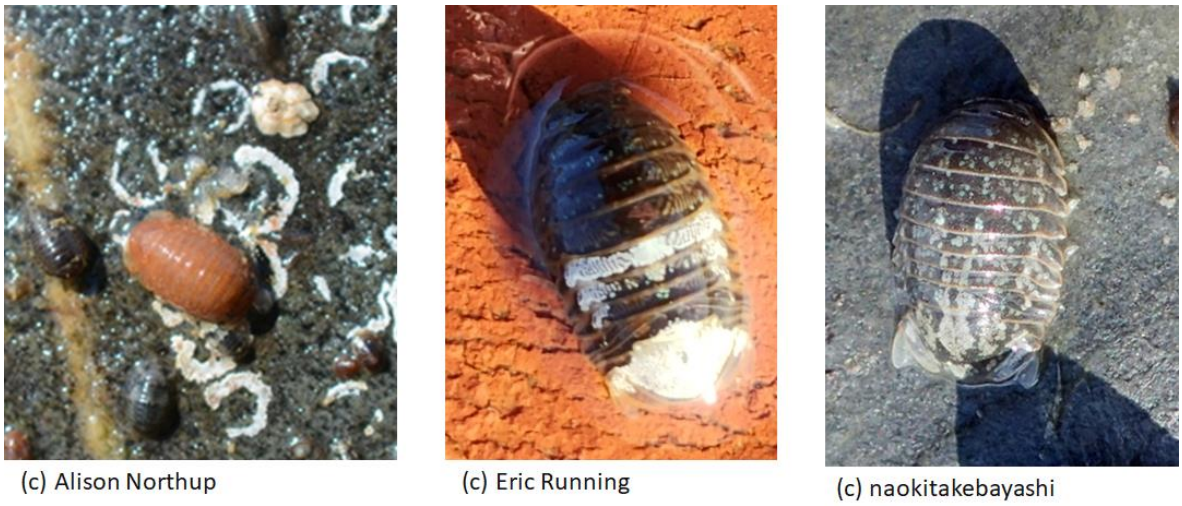


Figure 4.4 Examples of colour polymorphism in *Gnorisphaeroma oregonense* (family: *Sphaeromatidae*), images retrieved from iNaturalist.

Isocladus armatus exhibits an exceptionally large range of colour polymorphism that varies across a continuum, however some distinct morphs appear common across this continuum. (Jansen, 1971) recognized 2 main morphs – variegated and striped, however recent work has indicated that there are other distinctive morphs – such as green, red, white, un-patterned, and dotted (see Figure 4.5). Variation exists between population frequencies of morph, but to the best of our knowledge these morphs are present across many populations. Colouration in these species may provide an adaptive advantage in terms of predator avoidance in a highly variable environment.

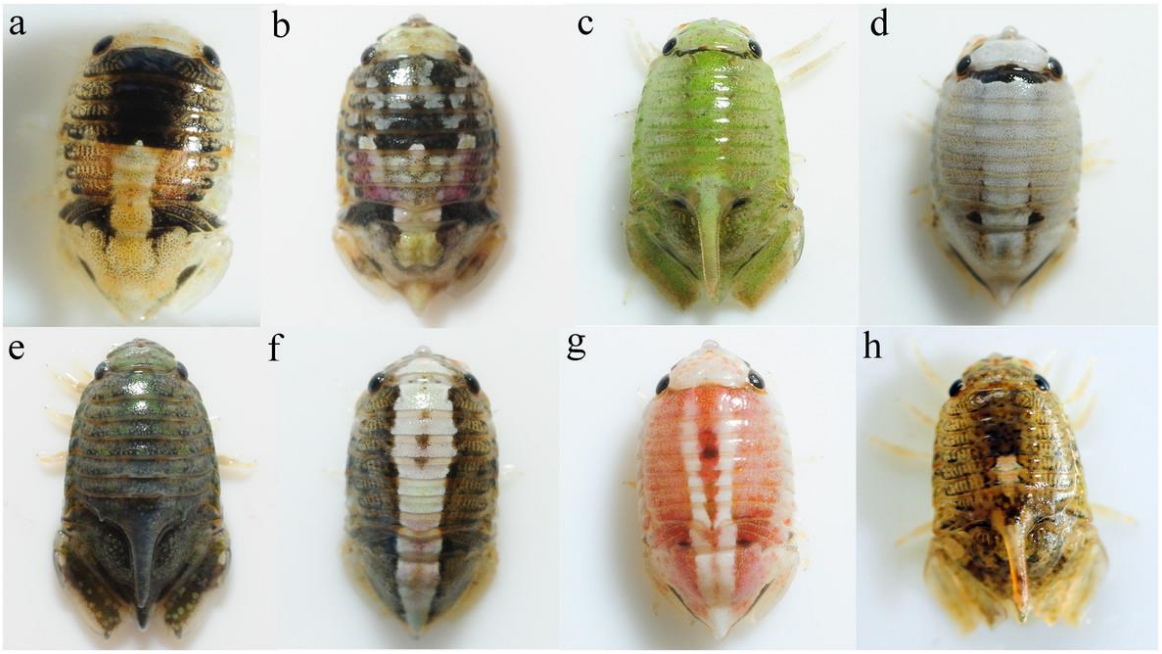


Figure 4.5 the range of colour polymorphism in *Isocladus armatus*, image reproduced from Wells and Dale (2018). A and B represent examples of the variegated morph, C represents the green morph, D the white morph, E the unpatterned morph, F and G the striped morph, and H the spotted morph.

Maintenance of CP in *Isocladus armatus* is likely the result of either balancing selection (such as NFDS) or divergent/disruptive selection. Sexual selection, while possible, is unlikely to play a significant role due to the fact that isopods do not exhibit complex social behaviours, which sexual selection predominantly acts upon. However, some species of *Sphaeromatid* isopods exhibit a range of selective mechanisms, including potential sexual selection, based on the presence of both autosomal and sex linked colouration morphs (Shuster et al., 2014; Shuster & Levy, 1999).

Both divergent and disruptive selection have the potential to maintain CP in *Isocladus armatus*, on the basis that certain morphs have greater fitness than others in different environments. For example, the white morph may have greater fitness in a sandy environment than the green morph, while the green morph may exhibit greater fitness in an algal environment (possibly due to crypsis). *I. armatus* is found in highly heterogeneous substrates that may impose contrasting selective pressures. On this basis, where environmental 'borders' are virtually non-existent and morph types are not reproductively isolated, gene flow between morphs should maintain CP. Alternatively, balancing selection through NFDS could maintain CP on the basis of morph frequency affecting fitness. This is what was observed in some studies *Idotea baltica* (Jormalainen et al., 1995; Salemaa, 2009). This is similar to early research on *I. armatus* by (Jansen, 1971) which suggested an

effect of temperature on morph frequency that could result in CP being maintained by NFDS, and is supported by early observations of another *Sphaeromatid* isopod, which exhibits seasonal variation in morphotype in response to low salinity and temperature (Khazaeli & Heath, 1979).

Genetic Approaches to Understanding Colour Polymorphism

The genetic basis for colouration can be determined through techniques such as cross-breeding, this can be used to determine patterns of Mendelian inheritance associated with different morphs (Shuster et al., 2014; Shuster & Levy, 1999). However, techniques for determining the genomic basis for CP have become widely accessible in recent years. It is important to recognize that some studies treat genetic and genomic as synonymous, however for this research, genetic basis will simply mean determining that a trait is genetically determined, while the genomic basis refers to the specific elements of the genome that dictate the trait. Understanding the genomic basis of traits is now a routine approach for model species. Most commonly, the first step is to assay a genome-wide panel of SNPs (Single Nucleotide Polymorphisms) across many individuals of different phenotypes. Tests to check for associations between certain alleles and the presence of a phenotype are then conducted (Iles, 2008). This does not, of course, test for causation, and can only help deduce whether a trait has a genetic component. Further tests can also be conducted to identify loci that are under selection for these phenotypes, using methods such as Fst-outlier analysis (Narum & Hess, 2011). This approach aims to identify loci that exhibit allele frequencies that are atypical compared to other loci in the dataset, and this is used to suggest that they may be acted upon by selection.

Most widely used methods to generate genome-wide SNP data will generate data for many neutral SNPs. This may have reduced utility in identifying SNPs under selection or testing for associations with phenotypes. However, recent developments in these approaches have started to be able to target active regions of the genome (Kilian et al., 2012). These new approaches may have greater power to detect loci associated with/or under selection for traits that may be actively expressed, such as colouration.

Specific Objectives

This research aims to 1) identify whether or not CP has a genetic basis in *Isocladus armatus*, and 2) provide initial evidence to indicate whether disruptive/divergent or balancing selection is maintaining CP. Because our methods are not able to distinguish between

disruptive and divergent selection (as both are forms of directional selection, but on different scales), I will use the term ‘disruptive’ when in reference to a single population, while ‘divergent’ will be used when referencing multiple populations. I hypothesize that NFDS will be the selective mechanism for different morphs within *I. armatus*. Whichever form of selection maintains CP, it must also function across species boundaries - as some of the colour morphs observed within *I. armatus* are present across species, both within the genus, but also at a higher level within the order *Isopoda*, such as in *Idotea baltica* (Figure 4.3).

I use a reduced representation genotyping approach, DArTseq (Kilian et al., 2012), to assay SNPs from across the genome of *Isocladus armatus*, for many individuals (total of 120, see Table 4.3) amongst three sites. This approach generates a large number (~80,000) of loci, enabling characterization of the genomic basis of CP. This modified genotyping-by-sequencing approach, uses PstI as one of the primary restriction enzymes for DNA digests. Because this enzyme is highly sensitive to methylation at CXG sites, it is able to be used to preferentially sequence low copy number informative regions of the genome compared to highly repetitive genomic elements which exhibit greater CXG methylation (Kilian et al., 2003). This means that this approach can be particularly useful in understanding traits being actively expressed. I employ this technique to identify loci under putative selection for color polymorphism and the underlying selective mechanism behind their maintenance.

4.2. Methods

4.2.1 Sample Collection and Processing

Sample Collection

Specimens of *Isocladus armatus* were collected from Browns Bay and Stanmore Bay in June 2018. Samples were chosen based first on morphotype (preferentially selecting unambiguous morphs from each of the four morphotypes) (see Table 4.3 for collection details). I collected 32 samples from both locations and stored these at -80°C in 100% ethanol until DNA extraction. For details regarding samples from 2015 see Wells and Dale (2018). Samples from 2015 are included as I find no evidence for significant temporal variation (see chapter 3), and due to similar sampling strategies the incorporation of these sites should increase power to detect colour associated loci.

Table 4.3 Number of each sex, and morph type per sampling event. Sampling events without year specified were collected in 2018. Sampling events from 2015 are from Wells & Dale 2018.

Site	Sex	Striped	Unpatterned	Variegated	White	Total
Hatfields Beach 2015	Female	3	6	3	2	14
Hatfields 2015	Male	0	4	4	3	11
Stanmore Bay 2015	Female	4	4	4	4	16
Stanmore Bay 2015	Male	4	4	4	3	15
Browns Bay	Female	4	4	4	4	16
Browns Bay	Male	4	4	4	4	16
Stanmore Bay	Female	4	4	4	4	16
Stanmore Bay	Male	4	4	4	4	16
Total	-	27	34	31	28	120

DNA Extraction

DNA was extracted using a modified protocol for the Qiagen DNEasy Blood and Tissue kit. Samples were first rinsed with ddH2O to remove residual ethanol and the cephala were removed using an ethanol and flame sterilized blade. They were then homogenized in a 65°C solution of 22 µl 20% SDS and 178 µl of 0.5M EDTA (pH 8). 10 µl of Proteinase K was then added and the solution was incubated at 65°C at 800 rcf for 24 hours. After 24 hours 400 µl of Qiagen Buffer ATL was added, alongside another 15 µl aliquot of Proteinase K and incubated again for 24 hours.

After this, 400 µl of Buffer AL and 400 µl of 100% ethanol was added to each sample and vortexed for 20 seconds. The solutions were passed through a spin column, in two steps due to the large volume. The column was then washed according to the manufacturer's protocol, before a three step elution took place. Each elution consisted of 50 µl of nuclease-free water

incubating on the column for 15 minutes before centrifugation at 7,000 rcf, this resulted in a total of 150 µl of eluate.

DNA samples were concentrated in an Eppendorf vacuum centrifuge at 30°C, to reduce total volume to approximately 50 µl. At this point the concentration was determined by using 2 µl of sample in a Qubit fluorometer assay.

The concentrations were then adjusted through the addition of nuclease-free water to be between 10 and 50 ng/µl before being plated and submitted for sequencing.

DNA Sequencing and GBS Method

DNA samples were submitted to Diversity Arrays Technology Ltd for a genome complexity reduction approach, similar to Genotyping-By-Sequencing. In this case, my samples were processed with the same approach as employed by Wells and Dale (2018), so as to ensure datasets were comparable, this is important as the DArTseq approach is optimized for each species.

The DArTseq approach involves digesting high molecular weight DNA using restriction enzymes (in this case PstI and SphI), after which adapters corresponding to restriction enzyme overhangs are then bound to the digested DNA (whereby overhangs associated with cutting sites have adapters bound to them). The samples are then amplified, using primers associated with the adapters, to amplify fragments containing the adapters. The amplified products were then sequenced on an Illumina HiSeq 2500. This approach is explained in more detail in (Elshire et al., 2011; Kilian et al., 2012; Wells & Dale, 2018).

The resulting sequences were then demultiplexed based on individual barcodes, then filtered to remove reads with Q-scores < 25. DArT proprietary bioinformatics pipelines were then implemented to process the data, this approach first removes low quality sequences, and stringently assesses barcode accuracy to ensure samples are correctly assigned to the appropriate individual. The reads are then assessed to both GenBank viral and bacterial databases, as well as an internal Diversity Arrays database to identify potential contaminants.

The filtered data is then processed into a secondary pipeline, which performs SNP calling. In this pipeline monomorphic sequences are removed, and SNPs are only called if both homozygous and heterozygous genotypes can be identified. This approach incorporates

technical replicates from the library preparation step, to ensure consistency in SNP calls across replicates.

4.2.2 Morph Classification

To understand the genomic basis for colouration in *Isocladus armatus* I took a high resolution photo of each individual immediately upon collection, and classified individuals into 5 groups (Fig. 4.6). Each photo was taken using a Nikon D5300 camera, with an ISO of 200, aperture F-stop of 14, and exposure time of 1/200 seconds. Photos were taken against a grey card with a ruler in each photo. Isopods were blot-dried with a tissue prior to photographing, to reduce glare.

Morph Name	Depiction
White	
Striped	
Unpatterned	
Variegated	
Other	N/A

Figure 4.6 Images of each morph of *Isocladus armatus* used in this study. “Other” represents any non-standard morph (such as green or spotted, see figure 4.5 for examples). The spines present in the first two photos are indicative of males, however all morphs are found in both sexes.

4.2.3 Data Filtering and Processing

I used the same data generated in the population genomics aspect of this thesis (see Chapter 3), however I did not impose the stringent filtering on this data as I did for the population analyses. Instead filtering was conducted as follows: first I used the R package dartR to remove all populations exception for both Stanmore Bay populations, Hatfields Beach, and Browns Bay, and then recalculated the statistics that I filtered on. I then removed the individual BBFA3 due to high missing data levels (93%), and then also removed any individuals for which a morph was considered ambiguous or fell outside of the four main morphs for which I was interested.

I then selected SNPs based on a call rate of at least 0.8, read depths of at least 3 and less than 500, replicability of at least 0.8. Finally, I removed monomorphic and secondary loci. An in-depth explanation of what these parameters are and why I filter on them is found in Chapter 3, Methods - SNP Filtering and Data Processing.

4.2.4 Testing for Selection and Association

I performed two analyses to identify the genomic basis of colouration. The first of these used BayeScan 2.1 (Foll & Gaggiotti, 2008) to understand how colouration is maintained within populations. BayeScan is a software designed to identify loci under natural selection among groups, based on differences in allele frequencies. It first identifies, for each loci, if there is a significant difference in frequency between each group and the pooled group of each sample. This enables the calculation of both a locus level, and group level value of Fst. This value is then decomposed into two components - essentially representing differences in allele frequency due to a group effect, and differences due to a locus effects (alpha). If alpha is significantly different to zero (based on a q-value, which is analogous to the p-value - in this case q-value represents the proportion of outlier loci that are expected to be false positives), then this suggests that a locus is under selection resulting in a significant contribution to the locus-specific Fst. In contrast, if it is not different to zero, then it suggests that the group effect is responsible. BayeScan can distinguish between loci under disruptive/divergent selection vs purifying or balancing selection, under the hypothesis that loci subject to balancing or purifying selection should exhibit lower genetic differentiation that expected to the pooled samples - while the opposite would be the case for disruptive or divergent selection. This does however mean that we cannot distinguish between disruptive and divergent selection.

I used this tool to identify SNPs that were potentially under selection for colouration. For this analysis, I set the prior odds to 100 (from the default of 10), and grouped individuals by morphotype. The prior odds value was chosen to increase stringency of my analysis without massively reducing power. In order to test the possibility of false positives, I used the same dataset but randomly assigned individuals to different morphotypes. I then performed the BayeScan analysis described above on this dataset as well, expecting that if this analysis was robust then we should observe no signal of selection on any loci.

The second analysis aimed to identify loci that were associated with colouration, this was done using Plink 1.9. This analysis used the same dataset as above, however I implemented a filter to remove oci with missingness greater than 0.05. I then tested for associations between SNPs and morph type, using chi-square tests, as implemented in Plink. Each colour was analysed separately. One morphotype was then used as a case for phenotype in Plink, while the remaining morphotype were combined into a 'control' group. This was repeated for each morphotype.
These tests were then adjusted using Bonferroni correction in order to account for multiple testing, using 10,000 permutations with the --mperm function.

Using the R package SNPrelate (Zheng et al., 2012), I conducted Principal Component Analyses for SNPs with a Q-Value of < 0.1 for BayeScan, as well as for significant SNPs from the Plink analysis. This approach allows visualization of genetic structure in the dataset, and by limiting this analysis to SNPs associated with colouration we are able to identify genetic structure due to colouration. Clustering significance was determined through a permanova and k-means analysis in the R package "vegan".

4.3. Results

4.3.1 SNP Filtering

I removed 9 individuals from my analyses due to morphotype ambiguity, leaving a total of 112 samples to conduct the rest of the analyses. I filtered the data down to a total of 20,725 loci from an initial count of 56,015 (Table 4.4)

Table 4.4 Results of the SNP quality filtering, and the number of SNPs remaining after each filter was implemented.

Condition	Number of SNPs remaining
Initial number of SNPs	56,015
Call rate > 0.8	35,311
Depth > 3 and < 500	35,307
Replicability > 0.95	34,554
Monomorphic loci removed	34,554
Secondary loci removed	20,725

4.3.2 Selection Analyses

Using BayeScan, I identified six loci putatively under selection for colouration or traits associated with colouration (q-value < 0.1), two of which had a q-value reported as 0 (Table 4.5). These loci all exhibited α greater than 0, which suggests that disruptive or divergent selection is acting upon these loci, rather than purifying or balancing selection.

Table 4.5 BayeScan results for loci with a Q-Value < 0.1. Alpha indicates the type of selection, positive values suggest disruptive selection, while negative values indicate purifying or balancing selection.

Locus Name	Posterior Probability	Alpha	Q-Value	Fst	Significant in Plink?
13986227-21-T/G	1	2.67	0	0.13	No
14004074-28-T/A	1	2.24	0	0.01	Yes
13987455-33-A/C	0.91	1.67	0.03	0.06	Yes
13996433-20-C/T	0.91	1.57	0.05	0.05	No
44796519-19-C/A	0.83	1.41	0.07	0.05	No
13996259-12-A/G	0.77	1.40	0.1	0.05	No

A PCA of these loci also indicated that colouration had a genetic basis (Fig. 4.7), the strongest effect for which was on the white and striped morphs. PC1 explained 34.6% of the variation, PC2 explained 25.9%, and PC3 explained 24.1%. This clustering was relatively strong for the striped and white morphs, but less evident for the variegated and unpatterned morphs. This was evidenced by a permanova analysis in the R package “vegan”, which indicated a significant ($P < 0.001$, $R2 = 0.45$) effect of morph type on clusters within the principal component data, k-means analysis also indicated strong clustering for the striped morph, most significantly for PC2 vs PC3.

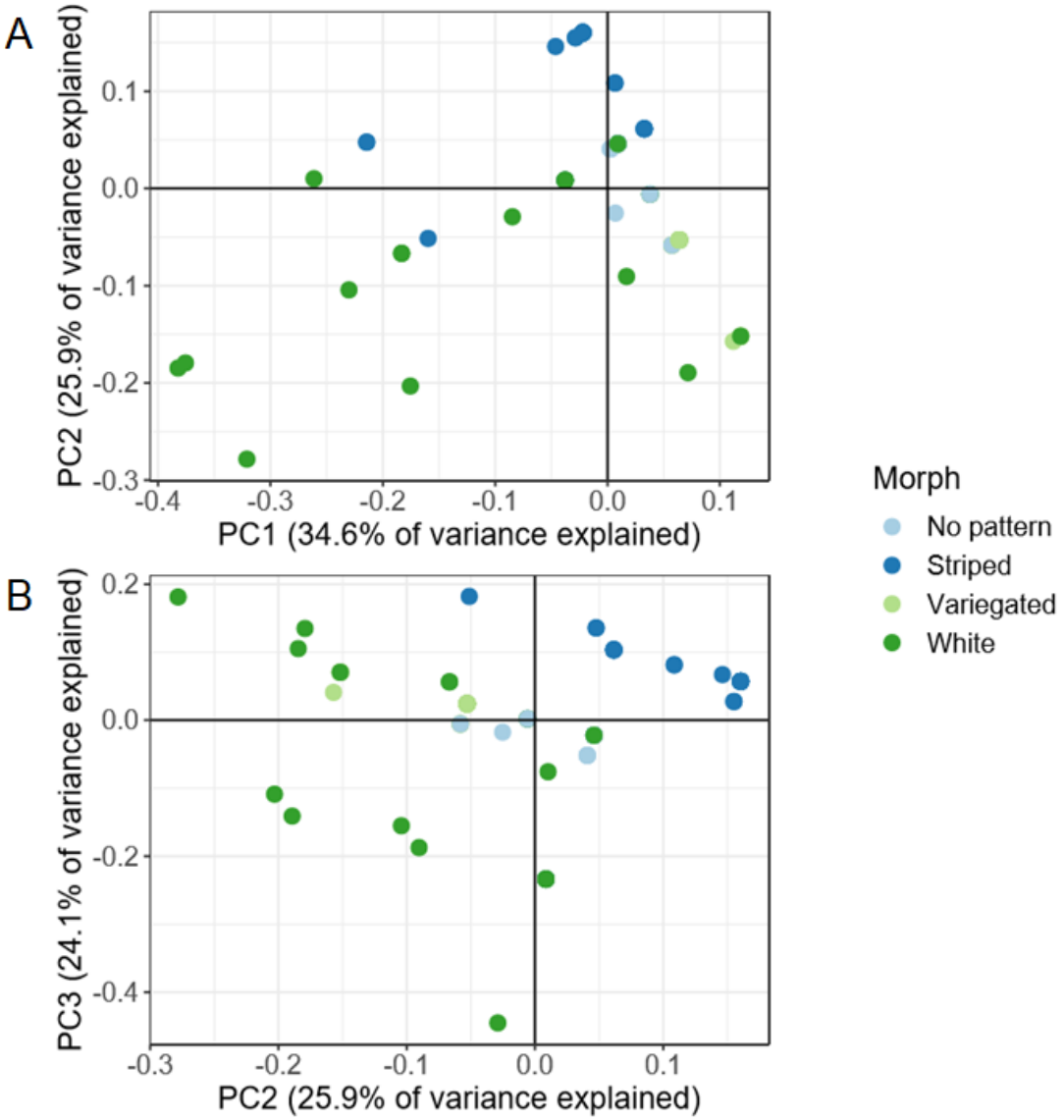


Figure 4.7 Principal Component Analyses for the six significant SNPS using BayeScan with a Q-Value < 0.1, coloured by morph type. A is PC1 vs PC2, while B is PC2 vs PC3.

Conversely for the randomized individual analysis I found only a single locus that appeared to be under selection. This suggests that BayeScan may not be especially sensitive to false positives and should be used alongside other analyses.

4.3.3 Genetic Association Analyses

The genetic significance of colouration was then further tested using PLINK 1.9 (Chang et al., 2015), this analysis found 1 locus strongly associated with the striped morph, 1 locus associated with the variegated morph, two loci associated with the white morph, and finally

no loci associated with the unpatterned morph (Table 4.6). Two of these loci also appeared to be strongly under selection in the BayeScan analysis.

Table 4.6 Plink results for significant loci (as determined by Plink with a corrected P-value < 0.05). Frequencies indicate the proportion of individuals having the associated allele at that locus.

Morph Association	**Locus Name**	**Bonferroni Corrected P-Value**	**Significant in BayeScan?**	**Frequency among morph**	**Frequency among control**
Striped	13991447_6_G_A_5	0.015	No	0.19	0.02
Variegated	14004074_28_T_A_27	< 0.0001	Yes	0.63	0.13
White	14003583_7_G_A_6	0.009	No	0.14	0
White	13987455_33_A_C_32	0.027	Yes	0.2	0.02

These SNPs showed some clustering for the white morph, but I observed relatively little clustering overall for this set of SNPs (Fig. 4.8). However, this may be, in part, due to the relatively few SNPs that were deemed significant. Despite this, PC1 still explained a large amount of the variance (31.7%) and PC2 explained a similar amount (28.4%). Permanova analysis within the R package “vegan”, indicated a strong effect of morph type on clustering ($P < 0.001$, $R2 = 0.2$), however k-means analysis indicated that only mild clustering was found for the white morph and the variance was too high to draw strong conclusions.

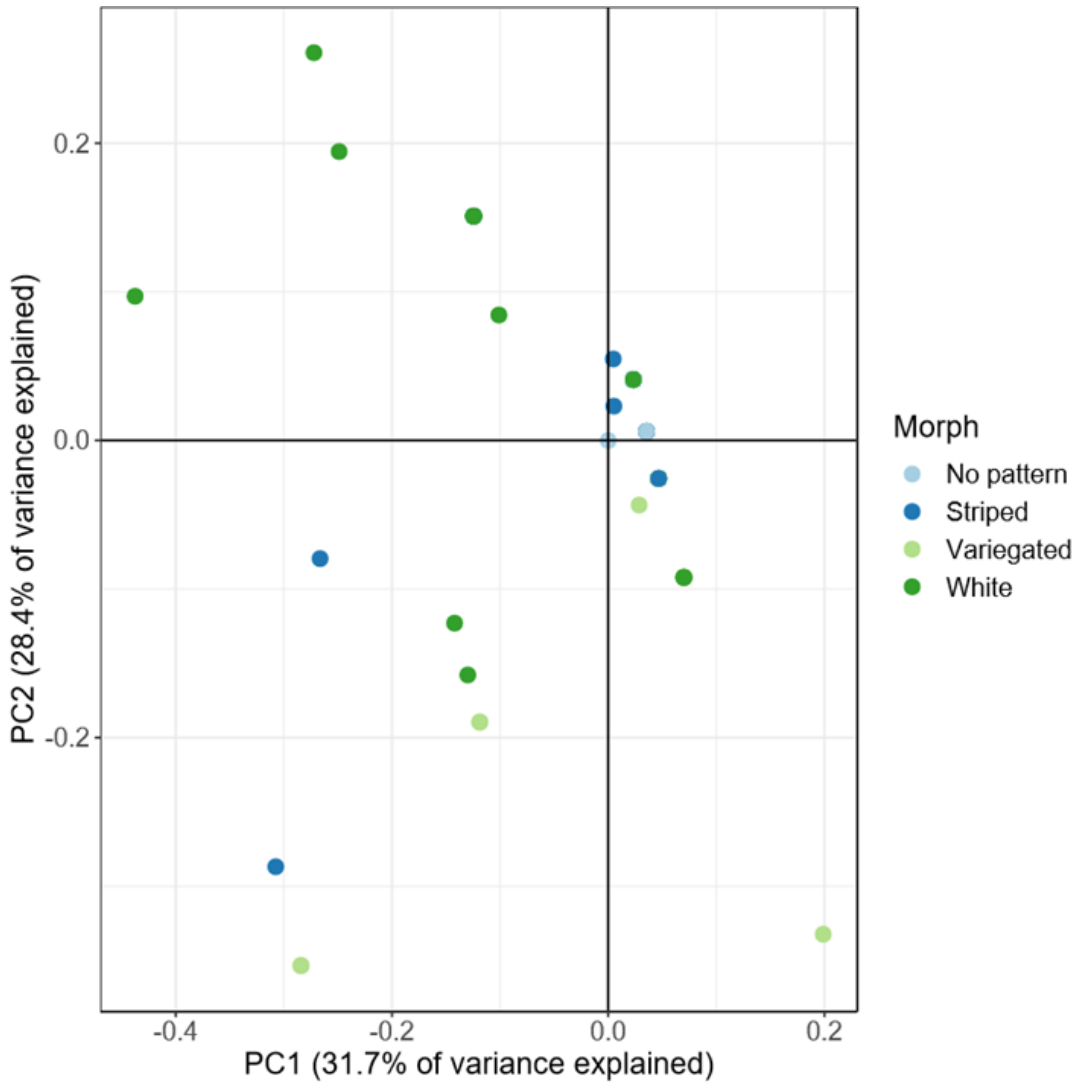


Figure 4.8 PCA of the four significant SNPs from Plink, using corrected p-value of < 0.05.

4. 4. Discussion

I found significant evidence for the genetic basis of colouration within *Isocladus armatus*, a result which reinforces previous research (Wells & Dale, 2018), but also indicates the genetic basis is not exclusive to the striped morph. This result is interesting as this previous work indicated a strong genetic basis for the striped morph, but not for any other morph types.

These results indicate that selection for colouration is driving a large amount of variation observed within loci associated with colouration. A genome wide association analysis indicated that the variegated, striped, and white morphs all have a strong genetic basis,

however I found no evidence to suggest that the unpatterned morph was under genetic control. This is not necessarily surprising, as while I treat this group as a separate morph, it may be the result of loci interactions or could be an intermediate morph, which may not be detectable within the current study design.

The eight loci I found significantly associated with CP (from both Plink and BayeScan) suggests that colouration within *I. armatus* is a complex trait controlled by multiple loci, and interactions between these loci. While we found some differences in the SNPs between our two analyses, this may be due to the fact the Plink used a much more conservative missingness filter than BayeScan. Additionally, Plink tests for association rather than selection, and as such is likely to identify SNPs that show no evidence for selection, but do exhibit evidence of association (i.e through linkage). The additive effect of CP loci on phenotype proposed by Wells and Dale (2018) is also partially supported by these results, as not all individuals possessing SNPs associated with certain morphs, expressed that specific morph (Table 4.6).

Interestingly, I only observed a few loci associated with colouration, while other studies investigating similar phenomena have observed many more significant loci (Gautier et al., 2018; M.-H. Li, Tiirikka, & Kantanen, 2014; Nazari-Ghadikolaei et al., 2018). This could be due to a variety of reasons, one of which could be that other species have a more complex genomic architecture involved with colouration, and as such more SNPs are associated with CP. This would result in a greater number of SNPs appearing associated with CP based on greater number of sites for potential linkage. In some species of goat and sheep, coat colour has had many significant associations across multiple chromosomes - suggesting a complex genetic basis (M.-H. Li et al., 2014; Nazari-Ghadikolaei et al., 2018). However, in both a species of ladybird, and songbird, CP has been shown to be mostly strongly associated with only one or two regions of the genome, and significance of SNPs may primarily arise due to linkage rather than complexity of this trait (Gautier et al., 2018; Kim et al., 2019). Therefore, the complexity of CP in *I. armatus* remains unclear.

Of the four significant loci detected by Plink, two of these were significant (Tables 4.5 and 4.6) in the BayeScan analysis. These loci were associated with the striped and white morphs. Both of these loci exhibited high alpha values in the BayeScan analysis, which suggests that these loci are maintained by a form of directional selection (either disruptive, or divergent). Because these analyses focus on populations that show very little genetic differentiation between them, I propose that the mechanism for this selection is disruptive selection as opposed to divergent. These analyses support my hypothesis that colouration is

largely driven by genetics and suggests that colour variation likely confers a selective advantage that results in it being maintained by selection.

This result was unanticipated because I predicted that balancing selection would maintain the colour polymorphism. Specifically, I hypothesized that rarity in morph type may confer a selective advantage resulting in NFDS for CP. Nevertheless, disruptive/divergent selection as a mechanism for CP maintenance has been found in other organisms. For example, divergent selection appears to maintain CP in *Idotea baltica* - where gene flow between populations maintains morphotypes across populations (Merilaita, 2001). In addition, research on the asp viper (*Vipera aspis*) also indicates that disruptive selection can act to maintain CP (Dubey et al., 2015). However, other examples of disruptive selection acting to maintain CP are relatively rare and evidence for it is generally indirect, as in this study (S. M. Gray & McKinnon, 2007).

Although my results are supportive of some form of directional selection maintaining CP, it is worth noting that it is not entirely clear as to what signal NFDS would exhibit in BayeScan. Previous research has indicated the NFDS exhibits a strong signature of balancing selection in BayeScan (Cavedon et al., 2019), and the authors of BayeScan have indeed used it in the context of frequency dependent selection (Fischer, Foll, Heckel, & Excoffier, 2014). Despite this, in complex cases of CP, where there are many morph types (as is the case in *I. armatus*) with a putative rare-morph advantage, some morphotypes may undergo NFDS on a smaller scale. Additionally, multiple forms of selection can act upon the same trait, especially if the associated genomic elements are found on both sex and autosomal chromosomes, as postulated by Shuster et al. (2014). Finally, NFDS could also contribute to the maintenance of CP, for example a trait in a population could hypothetically fluctuate between being under NFDS and disruptive selection (Rueffler et al., 2006). Take a scenario where there is an ‘optimum’ cryptic morph type, the population will eventually stabilize on this phenotype. However, if a predator eventually adapts to this, then the phenotype is not optimal anymore, and phenotypes that deviate from this ‘optimum’ now experience higher fitness - a case of disruptive selection. This could fluctuate over time, resulting in both NFDS and disruptive selection maintaining polymorphism (an in-depth discussion of this can be found in (Rueffler et al., 2006)). When analysing selection in genetic data, if the effects of disruptive selection are greater than that of NFDS, then this signal could overwhelm that of the NFDS in BayeScan. Therefore, I suggest caution in interpretation of these results, and suggest that further research is dedicated to understanding the contribution of each mechanism to CP maintenance. This should incorporate substantial modelling to better

understand how NFDS in massively polymorphic systems and directional selection can be distinguished.

Galeotti et al. (2003) proposed that if disruptive selection maintains CP in birds, then two functions may contribute to this selection. The first is crypsis, where crypsis against a background may result in differential predation, while the second is environmental adaptation linked to CP (such as temperature tolerance or exposure tolerance as a linked trait to CP). The latter hypothesis may explain CP maintenance in *Isocladus armatus*, as both Jansen (1971) and Khazaeli and Heath (1979) proposed a role of salinity and temperature in colour morph frequency. However, because morphotypes are found in relatively equal abundances in our study sites, and were sampled evenly, this may only play a small role and there is limited power to test this within our experimental design. The first hypothesis may be important in intertidal species because rocky shores are highly dynamic environments, with major fluctuations in exposure, sediment load, plant growth, and presence of other species. This dynamic environment also consists of variable substrate types (such as sand, stones, and shells), and background colours, the result of other species in the environment such as many different colors of algae, as well as snails and various shellfish. This likely results in many 'microhabitats' each of which could result in differential selection for different morphs. An example of potential fitness landscapes is displayed below in Fig. 4.9. In this example, the landscape represents the relative average fitness for a population, where each peak represents a dominant 'microhabitat' which imposes a selective pressure, while the range of the landscape represents the 'macrohabitat' or the local range of the population. Because continual interbreeding can occur between microhabitats, each peak is not isolated, and we observe intermediate morphs that could occur between fitness peaks.

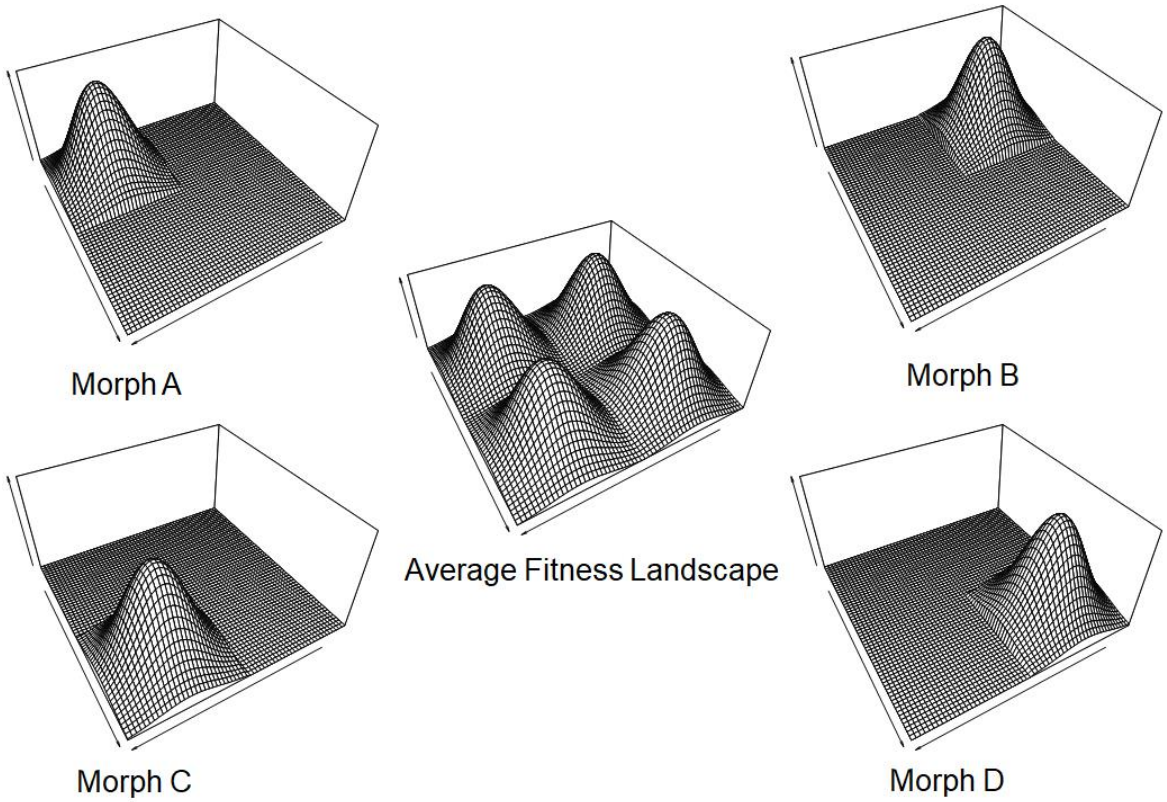


Figure 4.9 Potential fitness landscape for each colour polymorphism maintained by disruptive selection. Each morph would be a peak on the landscape. X and Z axis arrows indicate position within an environment, while the Y axis arrow represents the relative fitness at each point within the landscape. Each individual plot indicates the fitness landscape for that specific morph, while the average fitness landscape represents the fitness landscape for a population containing all four morphs.

In the absence of regular and continual interbreeding between morph types, it may be expected that divergence of species based on morph type may occur. However, if the microhabitats imposing selective pressures are much smaller than population range or dispersal distance of an individual, then continual interbreeding between morph types may result in the maintenance of CP by disruptive selection within a single population. Additionally, this would provide for the maintenance of intermediate morphs within a population.

While this provides a possible mechanism for CP maintenance through a form of directional selection, this mechanism is purely hypothetical. Other research has indicated that CP is maintained by divergent selection in *Idotea baltica*, through a similar mechanism on a larger spatial scale - where selection for morph type occurs at a population level rather than within a population (Merilaita, 2001). It is likely then, that this trait is maintained as a result of a

range of selective pressures, such as environmental factors (i.e predation, salinity, or temperature), and mechanisms, such as sexual selection or disruptive selection. This is a new and novel result that warrants further investigation in greater detail.

Future Works

Understanding of the genomics of colouration in *Isocladus armatus* would be greatly advanced by having a well resolved genome, alongside low coverage genomes for each morph. This method would enable greater understanding of how genetics contribute to the presence of different colour morphs. The addition of a transcriptomic element to this research would also be beneficial as it would enable a better breakdown of how colouration is controlled on a genetic level. In addition, these data would also enable the understanding of how the specific loci of interest interact to result in a specific phenotype. Ideally this work should be done using a line of captively bred individuals for each morph as well, to enable an understanding of the heritability of these traits.

Chapter 5 - Conclusion and Summary

In Chapter 2 of my thesis, I examined the effectiveness of two different mitochondrial DNA enrichment approaches on a marine isopod. As both approaches did not yield significant enrichment of mitochondrial DNA, I used whole genome shotgun sequencing to sequence and assemble the mitochondrial genome for the marine isopod *Isocladus armatus*. I did this to gain a better understanding of the evolution of atypical mitochondrial structure in *Isopoda*. In Chapter 3, I used population genomics to investigate population structure and connectivity across two spatial scales for this species, to inform the understanding of dispersal and population structure in a direct developing species. Finally, in Chapter 4 I use a subset of the data from Chapter 3 to identify loci that are significantly associated with colour polymorphism, to inform understanding of the genomics of this trait, as well as to further understand the adaptive significance of this trait and to understand how it is maintained within populations.

Mitochondrial Enrichment and Sequencing

The initial stages of my work aimed to assess the effectiveness of two approaches for mitochondrial DNA enrichment genomic DNA using *Isocladus armatus* as a study species. I attempted a multiple displacement amplification approach (MDA) (Wolff et al., 2012) and a differential centrifugation approach (Macher et al., 2018) to enrich for mitochondrial DNA. Initial results suggested that the enrichments may have been successful, as both qPCR and gel electrophoresis indicated enrichment in the MDA, while the differential centrifugation showed two distinct bands on an agarose gel, indicating a high abundance of DNA at two sizes which were expected to be due to the presence of a nuclear and mitochondrial DNA band. Despite these results, when I sequenced the DNA from these approaches, neither indicated any significant enrichment relative to abundance of mitochondrial reads from shotgun sequencing.

I therefore used nanopore shotgun sequencing, to generate and assemble a complete mitochondrial genome from whole genome sequencing data. The mitochondrial assemblies strongly suggested an atypical structure. This is significant because it was previously thought that atypical mitochondrial genomes were absent within this lineage.

These results suggested that mitochondrial genome was multipartite - possessing a dimeric circular chromosome, and potentially a single monomeric linear chromosome. The circular chromosome, which is termed the ‘dimer’, consisted of two slightly different copies of a

‘typical’ mitochondrial genome fused together as inverted repeats. I was unable to completely resolve the structure of the linear chromosome, because, as in other species, it is likely identical to 50% of the circular chromosome. This high level of similarity means that it can be difficult to resolve the structure with high confidence due to the fact that assembly algorithms may be unable to identify the correct placement of the read. However, the likely presence of this linear chromosome was suggested through analysis of the sequencing signal and sequencing speed for mitochondrial reads. Previous research has indicated a hairpin structure at the end of the linear chromosome (Peccoud et al., 2017), and I hypothesized that this hairpin should leave an identifiable signature within the sequencing signal (R. White et al., 2017). Additionally, a significant difference between the sequence speeding during translocation across the nanopore was observed between mitochondrial reads containing the putative hairpin and non-mitochondrial reads, where mitochondrial reads tended to travel slower through the pore.

The circular mitochondrial chromosome possessed key differences to the dimers observed in other species of *Isopoda* with this atypical structure. The first difference was the absence of a single tRNA heteroplasmy. In other species of isopods there are single nucleotide changes at three otherwise identical loci, and this results in changes to the site coding for three tRNA’s (see the figure below for details) (Chandler et al., 2015; Peccoud et al., 2017). For example, on one unit of the circular chromosome, the sequence at the heteroplasmic site is ‘tc**g**’, encoding for the tRNA-Arginine, while on the other unit, a single base change results in the coding site tc**c**, coding for the tRNA-Glycine. In *I. armatus* I observed two of these heteroplasmies, but not a third that has been described elsewhere (Peccoud et al., 2017). The second difference, I observed from other species of *Isopoda* are larger junctions between copies of the mitochondrial unit. This larger junction also contained a putative tRNA.

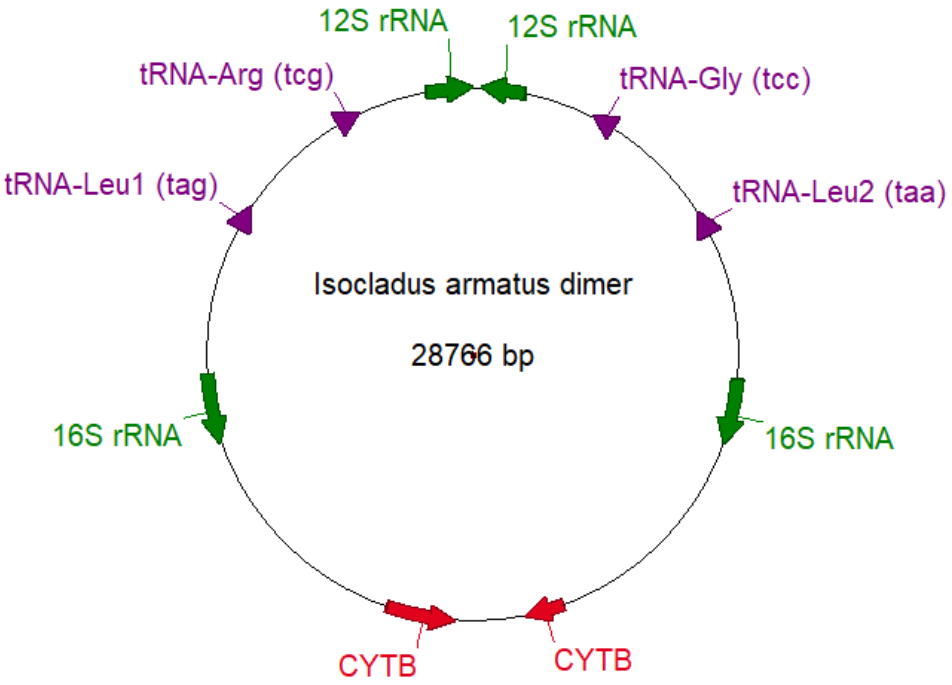


Figure 5.1 Annotation of rRNAs, junctions, and heteroplasmic tRNAs in *Isocladus armatus*. Directionality is indicated by direction of arrows, these annotations suggest duplication of the units as inverted repeats.

These results suggest that atypical structure is likely ancestral within *Isopoda*, or at least ancestral to the most recent common ancestor of the lineages described below, and that this atypical structure has been lost repeatedly within some lineages. This is evidenced by the same heteroplasmic sites and atypical structure being observed in three highly diverged lineages of isopods (*Asellota*,*Oniscidea*, and *Sphaeromatidea* (Doublet et al., 2012)). It is possible that this could be the result of convergent evolution, but I consider this explanation to be less parsimonious. While one other species of isopod within the order *Sphaeromatidae* has been reported to not possess this atypical structure, this could be the result of a reversion to a 'typical' metazoan mitochondrial structure, or alternatively could be due to a lack of sensitivity in those experiments due to the reliance on pooling samples (Doublet et al., 2013, 2012).

If this atypical structure is truly ancestral, then that suggests that the heteroplasmies and atypical size have been maintained for an order of magnitude longer than previous estimates. Doublet et al (2008) suggested that this structure was at least 40 million years, given the presence across *Oniscidea*, however Lins et al. (2012) suggest that the most recent ancestor for Sphaeromatidae and *Oniscidea* existed approximately 400 million years ago. Given the presence of this atypical structure and the presence of the heteroplasmies in both lineages, this suggests a strong adaptive significance of this structure. This likely

results from the essential role of mitochondria in energy metabolism. Loss of this structure without first re-acquiring the heteroplasmic tRNAs (as hypothesized to have occurred within *Sphaeroma serratum*) would likely be either highly deleterious or even lethal.

Population genomics reveals substantial structure across populations of *I. armatus*

In Chapter 3, I used a dataset of 8,020 SNPs to infer population structure and assess gene flow in *I. armatus* across multiple spatial scales. At larger spatial scales (above 20 kms) I found evidence of relatively low levels of gene flow between various populations. However, at smaller spatial scales of 5-20 kms, gene flow was much higher between populations. For example, among the Hauraki Gulf populations, gene flow was approximately 7 times greater than between populations 110 kms apart. This suggests that populations within the Hauraki gulf may exist as a metapopulation (Goldstien et al., 2010; Nagel et al., 2015). Additionally, I found a strong relationship between geographic and genetic distance, suggesting Isolation-By-Distance as a strong factor in population differentiation. However, this relationship only explained 5% of the total amount of genetic variation, which suggests many other factors may contribute to population differentiation.

This geographical relationship is also supported by a Principal Component Analysis, where Procrustes transformation of the principal components saw a strong association between geographic arrangement and principal component arrangement.

I found evidence for a genetic break between Māhia Peninsula and Wellington, New Zealand, in contrast to initial expectations and the results observed by many other studies (Ross, Hogg, Pilditch, Lundquist, & Wilkins, 2012; M. I. Stevens & Hogg, 2004; Veale & Lavery, 2012). Interestingly, a meta-analysis conducted by Arranz Martinez (2017) suggests that a genetic break occurs near Māhia peninsula, predominantly affecting species with low dispersal capability (such as direct developers). While this research was not able to identify any features of the area that could account for this break, seascape features between Māhia and Wellington that could contribute to the division between populations include currents such as the Wairarapa Eddy and the East Cape Current, both of which have the potential to distribute individuals away from Wellington, however it is unclear why a break in this region would predominantly affect species with low dispersal capacity (Arranz Martinez, 2017). This result was surprising as many other studies have found evidence for a genetic break near East Cape for both direct developers (M. I. Stevens & Hogg, 2004; Veale & Lavery, 2012) and biphasic species (Ross et al., 2012). In contrast to my initial hypothesis, I found high

gene flow across this break relative to other pairs of populations, with the presence of 3 highly admixed individuals from Māhia and Mt Maunganui that indicated recent migration between these two populations.

These results suggest a need to further study how biogeographic breaks can affect species with different life histories. Additionally, given the lack of effect of the break at East Cape on *I. armatus*, it suggests further study is required to understand dispersal in direct developers. This research supports the hypothesis that the strong swimming ability of *I. armatus*, enables greater dispersal than has generally been hypothesised for species lacking a pelagic larval stage. This is supported by the high estimates of bidirectional migration over East Cape, which suggests migration against ocean currents. This can indicate that either 1) seasonal variation in ocean currents, or 2) the presence of local in-shore currents or sporadic storm events, enables rafting, or 3) that this species' strong swimming ability increases dispersal potential. These results are in contrast to specific studies of direct developers, but complementary to a recent meta-analysis, thus it is unclear whether dispersal is *I. armatus* is unusual or representative or other direct developers.

The break I observed between Māhia and Wellington clearly differentiates the North group of populations (all populations excluding Kaikōura and Wellington), from the South group (Kaikōura and Wellington). However, it is possible that this South group consists of a cryptic species and the genetic break between these groups is due to speciation rather than seascape features, as this has been observed in another intertidal species in New Zealand (Sponer & Roy, 2002). It is however unlikely to be the result of geographic isolation of these groups, as there are many populations of *I. armatus* between Māhia and Wellington (Hurley & Jansen, 1977; "iNaturalist.org," n.d.); O. Wade, Pers. Comm), and the genetic distance I observed is greater than expected given an Isolation-By-Distance relationship.

Colour polymorphism has a strong genetic basis

I identified six SNPs strongly associated with colour polymorphism in *Isocladus armatus*. These SNPs suggested that a genetic basis may underlie the white, striped, and variegated morphs, while no SNPs were associated with the unpatterned morph. Four of these SNPs appear to be maintained through disruptive selection. This result indicates that some of the colour morphs have a very strong genetic basis and supports the conclusions of Wells and Dale (2018). My analysis indicates that colour polymorphism is not maintained by balancing selection, as has been proposed for other species of invertebrates (M. Stevens et al., 2014; Takahashi et al., 2010).

In theory, disruptive selection could lead to speciation across morph boundaries. However, the population genomics supports the idea that populations of *I. armatus* span a relatively large (at least 5km) range. This is substantially larger than any potential microhabitats, thus potentially explaining why speciation between morphs has not been observed. Additionally, colour polymorphism in *I. armatus* is maintained across populations around New Zealand, supporting this hypothesis.

I propose that selection drives diversity in morphotype due a potential role in crypsis, as proposed in other species of isopod (Hultgren & Mittelstaedt, 2015; S. Merilaita, 1998; M. Stevens et al., 2014). While this could lead to fixation of morphs, rocky shores (the habitat for *I. armatus*) are highly variable environments with a range of potential backgrounds that would each impose a selective pressure for crypsis against this background (i.e. white morphs would be selected for against a white background such as shells). Because these backgrounds are small compared to the population range, the selective pressures imposed by these microhabitats may not be strong enough to overcome the effect of gene flow between morphs within a population. In effect, disruptive selection could maintain CP due to high levels of gene flow between microhabitats.

Future Work

Draft genome sequencing of *I. armatus*

Future work should prioritize the sequencing of a draft (or low coverage) genome for *I. armatus*. This approach should make use of both long and short read sequencing, to generate a contiguous chromosome level assembly.

A draft genome would enable a greater number of SNPs to be discovered and would allow direct estimation of linkage (Mousavi et al., 2016) and Identity-By-Descent (Browning, 2008; Y. Wang et al., 2017). This would enable greater understanding of historical population demographics, such as population size, migration, and relatedness (Gauvin et al., 2014; Palamara, 2014). Additionally, having the entire DNA sequence of the genome would help to further understand the genomic basis of colouration. For example, genomic elements responsible for colouration may be linked with SNPs that we find are associated with colouration and thus should be investigated further (Gautier et al., 2018; Kim et al., 2019; Stinchcombe & Hoekstra, 2008). This work could include transcriptomics, which would also aid in understanding how certain genes are able to interact and result in colour

polymorphism. The methods outlined by Gautier et al. (2018) and Kim et al. (2019) and those proposed by Stinchcombe & Hoekstra (2008) would provide a good approach for understanding the genomics of colouration in *I. armatus*. These approaches would need to generate transcriptomes, or at the least cDNA sequences for the loci of interest, from as early on in the isopod life cycle as possible. This is because colouration loci may not be continually expressed across the life of an individual, and may only be expressed during development and moulting. A series of experiments should thus be conducted using a captive breeding system to generate transcriptomes and cDNA sequences for each morph, over their life cycle.

Phylogenetic analysis for the *Exosphaeroma* clade and *Sphaeromatidae* family

Additionally, a phylogenetic analysis should be conducted for many species within the *Exosphaeroma* clade and *Sphaeromatidae* family in order to resolve the relationships between species in these groups and within New Zealand isopods. This analysis should incorporate a range of nuclear loci, but ideally also make use of whole mitochondrial genomes for the phylogenetic analysis.

Recent DNA based analyses have failed to fully resolve relationships between species of isopod within the *Exosphaeroma* clade (Wetzer et al., 2018). The incongruence observed was between the phylogenetic placement of these species using both a nuclear and a mitochondrial loci (18S and 16S rDNA). Increasing the amount of data used in these analyses may contribute to increased resolution and reduced incongruence. Additional work should be performed to understand the benefits of whole mitochondrial phylogenetics, with a comparison of the results to a range of nuclear loci.

Sequencing the mitochondrial genomes of these species would shed light on how widespread atypical mitochondrial structure is among *Isopoda*. Including species within the *Sphaeroma* genus in such an analysis would be particularly useful as there is no evidence that suggests that these species have atypical mitochondrial structure (Doublet et al., 2012; M. Yang et al., 2019). This absence leaves many questions about how mitochondrial structure has evolved across many lineages of isopod. This would provide insights into evolutionary responses to loss of essential genes (such has been proposed to be a primary factor in the evolutionary maintenance of this atypical mitochondrial structure (Doublet et al.,

2012; Peccoud et al., 2017)). The approach described above would go a long way to help answer these questions.

Bibliography

Al-Nakeeb, K., Petersen, T. N., & Sicheritz-Pontén, T. (2017). Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics*, *18*(1), 510. https://doi.org/10.1186/s12859-017-1927-y

Amos, W. (2010). Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? *Proceedings. Biological Sciences / The Royal Society*, *277*(1686), 1443–1449. https://doi.org/10.1098/rspb.2009.1757

Andrade, P., Pinho, C., Pérez I de Lanuza, G., Afonso, S., Brejcha, J., Rubin, C.-J., … Carneiro, M. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5633–5642. https://doi.org/10.1073/pnas.1820320116

Arakawa, K., Suzuki, H., & Tomita, M. (2009). Quantitative analysis of replication-related mutation and selection pressures in bacterial chromosomes and plasmids using generalised GC skew index. *BMC Genomics*, *10*, 640. https://doi.org/10.1186/1471-2164-10-640

Arranz Martinez, V. (2017). *Connectivity among marine communities: a multi-species approach to determining the major drivers of larval connection between populations of coastal species in New Zealand* (ResearchSpace@Auckland). Retrieved from http://hdl.handle.net/2292/37344

Arrontes, J. (1991). Colour polymorphism in relation to spatial distribution in some intertidal isopods in northern Spain. *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom*, *71*(4), 749–758. https://doi.org/10.1017/S002531540005342X

Ayre, D. J., Minchinton, T. E., & Perrin, C. (2009). Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology*, *18*(9), 1887–1903. https://doi.org/10.1111/j.1365-294X.2009.04127.x

Baratti, M., Filippelli, M., & Messana, G. (2011). Complex genetic patterns in the mangrove wood-borer Sphaeroma terebrans Bate, 1866 (Isopoda, Crustacea, Sphaeromatidae) generated by shoreline topography and rafting dispersal. *Journal of Experimental Marine Biology and Ecology*, *398*(1), 73–82. https://doi.org/10.1016/j.jembe.2010.12.008

Baratti, M., Goti, E., & Messana, G. (2005). High level of genetic differentiation in the marine isopod Sphaeroma terebrans (Crustacea Isopoda Sphaeromatidae) as inferred by

mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology*, *315*(2), 225–234. https://doi.org/10.1016/j.jembe.2004.09.020

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(Database issue), D36–D42. https://doi.org/10.1093/nar/gks1195

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bertram, L., & Tanzi, R. E. (2009). Genome-wide association studies in Alzheimer's disease. *Human Molecular Genetics*, *18*(R2), R137–R145. https://doi.org/10.1093/hmg/ddp406

Browning, S. R. (2008). Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics*, *178*(4), 2123–2132. https://doi.org/10.1534/genetics.107.084624

Brusca, R. C., & Gilligan, M. R. (1983). Tongue replacement in a marine fish (Lutjanus guttatus) by a parasitic isopod (Crustacea: Isopoda). *Copeia*, *1983*(3), 813–816.

Cavedon, M., Gubili, C., Heppenheimer, E., vonHoldt, B., Mariani, S., Hebblewhite, M., … Musiani, M. (2019). Genomics, environment and balancing selection in behaviourally bimodal populations: The caribou case. *Molecular Ecology*, *28*(8), 1946–1963. https://doi.org/10.1111/mec.15039

Chandler, C. H., Badawi, M., Moumen, B., Grève, P., & Cordaux, R. (2015). Multiple Conserved Heteroplasmic Sites in tRNA Genes in the Mitochondrial Genomes of Terrestrial Isopods (Oniscidea). *G3* , *5*(7), 1317–1322. https://doi.org/10.1534/g3.115.018283

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Cole, L. W. (2016). The Evolution of Per-cell Organelle Number. *Frontiers in Cell and Developmental Biology*, *4*, 85. https://doi.org/10.3389/fcell.2016.00085

Cotton, J. (2001). Homologous recombination in animal mitochondria. *Genome Biology*, *2*(10), reports0034. https://doi.org/10.1186/gb-2001-2-10-reports0034

Cowman, P. F., & Bellwood, D. R. (2013). Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. *Proceedings. Biological Sciences / The Royal Society*, *280*(1768), 20131541. https://doi.org/10.1098/rspb.2013.1541

Detto, T., Hemmi, J. M., & Backwell, P. R. Y. (2008). Colouration and colour changes of the fiddler crab, Uca capricornis: a descriptive study. *PloS One*, *3*(2), e1629. https://doi.org/10.1371/journal.pone.0001629

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

Doublet, V., Helleu, Q., Raimond, R., Souty-Grosset, C., & Marcadé, I. (2013). Inverted repeats and genome architecture conversions of terrestrial isopods mitochondrial DNA. *Journal of Molecular Evolution*, *77*(3), 107–118. https://doi.org/10.1007/s00239-013-9587-7

Doublet, V., Raimond, R., Grandjean, F., Lafitte, A., Souty-Grosset, C., & Marcadé, I. (2012). Widespread atypical mitochondrial DNA structure in isopods (Crustacea, Peracarida) related to a constitutive heteroplasmy in terrestrial species. *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*, *55*(3), 234–244. https://doi.org/10.1139/g2012-008

Doublet, V., Souty-Grosset, C., Bouchon, D., Cordaux, R., & Marcadé, I. (2008). A thirty million year-old inherited heteroplasmy. *PloS One*, *3*(8), e2938. https://doi.org/10.1371/journal.pone.0002938

Dray, S., Dufour, A.-B., & Others. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. Retrieved from http://pbil.univ-lyon1.fr/JTHome/Stage2009/articles/SD839.pdf

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., … Others. (2011). Geneious. *Biomatters Ltd*.

Duarte, R. C., Stevens, M., & Flores, A. A. V. (2016). Shape, colour plasticity, and habitat use indicate morph-specific camouflage strategies in a marine shrimp. *BMC Evolutionary Biology*, *16*(1), 218. https://doi.org/10.1186/s12862-016-0796-8

Dubey, S., Zwahlen, V., Mebert, K., Monney, J.-C., Golay, P., Ott, T., … Ursenbacher, S. (2015). Diversifying selection and color-biased dispersal in the asp viper. *BMC Evolutionary Biology*, *15*, 99. https://doi.org/10.1186/s12862-015-0367-4

Duchêne, S., Archer, F. I., Vilstrup, J., Caballero, S., & Morin, P. A. (2011). Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PloS One*, *6*(11), e27138. https://doi.org/10.1371/journal.pone.0027138

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PloS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA

restriction data. *Genetics*, *131*(2), 479–491. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1644282

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12930761

Fischer, M. C., Foll, M., Heckel, G., & Excoffier, L. (2014). Continental-scale footprint of balancing and positive selection in a small rodent (Microtus arvalis). *PloS One*, *9*(11), e112332. https://doi.org/10.1371/journal.pone.0112332

Foll, M. (2012). BayeScan v2. 1 user manual. *Ecology*, *20*, 1450–1462. Retrieved from http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Fraser, C. I., Morrison, A. K., Hogg, A. M., Macaya, E. C., van Sebille, E., Ryan, P. G., Padovan, A., Jack, C., Valdivia, N., & Waters, J. M. (2018). Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nature Climate Change*, *8*(8), 704–708. https://doi.org/10.1038/s41558-018-0209-7

Galeotti, P., Rubolini, D., Dunn, P. O., & Fasola, M. (2003). Colour polymorphism in birds: causes and functions. *Journal of Evolutionary Biology*, *16*(4), 635–646. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14632227

Gautier, M., Yamaguchi, J., Foucaud, J., Loiseau, A., Ausset, A., Facon, B., … Prud'homme, B. (2018). The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird. *Current Biology: CB*, *28*(20), 3296–3302.e7. https://doi.org/10.1016/j.cub.2018.08.023

Gauvin, H., Moreau, C., Lefebvre, J.-F., Laprise, C., Vézina, H., Labuda, D., & Roy-Gagnon, M.-H. (2014). Genome-wide patterns of identity-by-descent sharing in the French Canadian founder population. *European Journal of Human Genetics: EJHG*, *22*(6), 814–821. https://doi.org/10.1038/ejhg.2013.227

Gee, H. (2003). [Review of *Evolution: ending incongruence*]. *Nature*, *425*(6960), 782. https://doi.org/10.1038/425782a

Gemmell, M. R., Trewick, S. A., Crampton, J. S., Vaux, F., Hills, S. F. K., Daly, E. E., … Morgan-Richards, M. (2018). Genetic structure and shell shape variation within a rocky shore whelk suggest both diverging and constraining selection with gene flow. *Biological Journal of the Linnean Society. Linnean Society of London*, *125*(4), 827–843. https://doi.org/10.1093/biolinnean/bly142

GenSkew - visualization of nucleotide skew in genome sequences. (n.d.). Retrieved August 2, 2019, from http://genskew.csb.univie.ac.at/

GeoNet M 7.8 Kaikōura Mon, Nov 14 2016. (n.d.). Retrieved October 10, 2019, from https://www.geonet.org.nz/earthquake/story/2016p858000

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., … Timp, W. (2019). Targeted Nanopore Sequencing with Cas9 for studies of methylation, structural variants and mutations (p. 604173). https://doi.org/10.1101/604173

Goldstien, S. J., Schiel, D. R., & Gemmell, N. J. (2010). Regional connectivity and coastal expansion: differentiating pre-border and post-border vectors for the invasive tunicate Styela clava. *Molecular Ecology*, *19*(5), 874–885. https://doi.org/10.1111/j.1365-294X.2010.04527.x

Gosselin, T. (2019). *thierrygosselin/radiator: Official release*. https://doi.org/10.5281/zenodo.2595083

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology*, *4*(9), a011403. https://doi.org/10.1101/cshperspect.a011403

Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, *22*(2), 71–79. https://doi.org/10.1016/j.tree.2006.10.005

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hahn, M. (2019). Population Structure. In Sinauer (Ed.), *Molecular Population Genetics* (pp. 105–109). 198 Madison avenue, New York, NY 10016, United States of America: Oxford University Press.

Hellberg, M. E. (2009). Gene Flow and Isolation among Populations of Marine Animals. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 291–310. https://doi.org/10.1146/annurev.ecolsys.110308.120223

Hill, G. E., Inouye, C. Y., & Montgomerie, R. (2002). Dietary carotenoids predict plumage coloration in wild house finches. *Proceedings. Biological Sciences / The Royal Society*, *269*(1496), 1119–1124. https://doi.org/10.1098/rspb.2002.1980

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *11*(s1), 117–122. https://doi.org/10.1111/j.1755-0998.2010.02967.x

Hultgren, K. M., & Mittelstaedt, H. (2015). Color change in a marine isopod is adaptive in reducing predation. *Current Zoology*, *61*(4), 739–748. https://doi.org/10.1093/czoolo/61.4.739

Hurley, D. E., & Jansen, K. P. (1977). The marine fauna of New Zealand: Family Sphaeromatidae (Crustacea Isopoda: Flabellifera). *New Zealand Oceanographic Institute Memoir*, *63*, 1–80. Retrieved from http://isopods.nhm.org/pdfs/2325/2325.pdf

Huynh, L. Y., Maney, D. L., & Thomas, J. W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (Zonotrichia albicollis). *Heredity*, *106*(4), 537–546. https://doi.org/10.1038/hdy.2010.85

Iles, M. M. (2008). What can genome-wide association studies tell us about the genetics of common disease? *PLoS Genetics*, *4*(2), e33. https://doi.org/10.1371/journal.pgen.0040033

iNaturalist.org. (n.d.). Retrieved May 16, 2018, from iNaturalist.org website: https://www.inaturalist.org/

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* , *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jansen, K. P. (1968). *a Comparative Study of Intertidal Species of Sphaeromidae (Isopoda Flabellifera)*.

Jansen, K. P. (1971). Ecological studies on intertidal New Zealand Sphaeromatidae (Isopoda: Flabellifera). *Marine Biology*, *11*(3), 262–285. https://doi.org/10.1007/BF00401274

Johri, S., Solanki, J., Cantu, V. A., Fellows, S. R., Edwards, R. A., Moreno, I., … Dinsdale, E. A. (2019). “Genome skimming” with the MinION hand-held sequencer identifies CITES-listed shark species in India’s exports market. *Scientific Reports*, *9*(1), 4476. https://doi.org/10.1038/s41598-019-40940-9

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jormalainen, V., Merilaita, S., & Tuomi, J. (1995). Differential predation on sexes affects colour polymorphism of the isopod Idotea baltica (Pallas). *Biological Journal of the Linnean Society. Linnean Society of London*, *55*(1), 45–68. https://doi.org/10.1111/j.1095-8312.1995.tb01049.x

Kahnt, B., Theodorou, P., Soro, A., Hollens-Kuhr, H., Kuhlmann, M., Pauw, A., & Paxton, R. J. (2018). Small and genetically highly structured populations in a long-legged bee, Rediviva longimanus, as inferred by pooled RAD-seq. *BMC Evolutionary Biology*, *18*(1), 196. https://doi.org/10.1186/s12862-018-1313-z

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kayal, E., Bentlage, B., Collins, A. G., Kayal, M., Pirro, S., & Lavrov, D. V. (2012). Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biology and Evolution*, *4*(1), 1–12. https://doi.org/10.1093/gbe/evr123

Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics*, *196*(1), 313–320. https://doi.org/10.1534/genetics.113.158758

Khazaeli, A. A., & Heath, D. J. (1979). Colour polymorphism, selection and the sex ratio in the isopod Sphaeroma rugicauda (Leach). *Heredity*, *42*(2), 187–199. https://doi.org/10.1038/hdy.1979.22Kierepka, E. M., & Latch, E. K. (2015). Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, *15*(3), 512–525. https://doi.org/10.1111/1755-0998.12332

Kilian, A., Huttner, E., Wenzl, P., Jaccoud, D., Carling, J., Caig, V., … Others. (2003). The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. *Proceedings of the International Congress in the Wake of the Double Helix: From the Green Revolution to the Gene Revolution*, 443–461. Retrieved from http://dista.unibo.it/doublehelix/proceedings/SECTION_IV/HELIX%20pp%20443-461.pdf

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., … Uszynski, G. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods in Molecular Biology* , *888*, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Kilpert, F., Held, C., & Podsiadlowski, L. (2012). Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. *Molecular Phylogenetics and Evolution*, *64*(1), 106–117. https://doi.org/10.1016/j.ympev.2012.03.013

Kim, K.-W., Jackson, B. C., Zhang, H., Toews, D. P. L., Taylor, S. A., Greig, E. I., … Burke, T. (2019). Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nature Communications*, *10*(1), 1852. https://doi.org/10.1038/s41467-019-09806-6

Kimura, M., & Weiss, G. H. (1964). The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, *49*(4), 561–576. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17248204

Kokko, H., Griffith, S. C., & Pryke, S. R. (2014). The hawk-dove game in a sexually reproducing species explains a colourful polymorphism of an endangered bird. *Proceedings. Biological Sciences / The Royal Society*, *281*(1793). https://doi.org/10.1098/rspb.2014.1794

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences

across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kusche, H., Elmer, K. R., & Meyer, A. (2015). Sympatric ecological divergence associated with a color polymorphism. *BMC Biology*, *13*, 82. https://doi.org/10.1186/s12915-015-0192-7

Leese, F., Agrawal, S., & Held, C. (2010). Long-distance island hopping without dispersal stages: Transportation across major zoogeographic barriers in a Southern Ocean isopod. *Die Naturwissenschaften*, *97*(6), 583–594. https://doi.org/10.1007/s00114-010-0674-y

Leese, F., Kop, A., Wägele, J.-W., & Held, C. (2008). Cryptic speciation in a benthic isopod from Patagonian and Falkland Island waters and the impact of glaciations on its population structure. *Frontiers in Zoology*, *5*, 19. https://doi.org/10.1186/1742-9994-5-19

Lessios, H. A., & Weinberg, J. R. (1993). Migration, gene flow and reproductive isolation between and within morphotypes of the isopod Excirolana in two oceans. *Heredity*, *71*, 561. https://doi.org/10.1038/hdy.1993.180

Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, *46*(3), 282–297. https://doi.org/10.1093/icb/icj024

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* , *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M.-H., Tiirikka, T., & Kantanen, J. (2014). A genome-wide scan study identifies a single nucleotide substitution in ASIP associated with white versus non-white coat-colour variation in sheep (Ovis aries). *Heredity*, *112*(2), 122–131. https://doi.org/10.1038/hdy.2013.83

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Lin, G.-M., Xiang, P., Sampurna, B. P., & Hsiao, C.-D. (2017). Genome skimming yields the complete mitogenome of Chromodoris annae (Mollusca: Chromodorididae). *Mitochondrial DNA Part B*, *2*(2), 609–610. https://doi.org/10.1080/23802359.2017.1372715

Lins, L. S. F., Ho, S. Y. W., Wilson, G. D. F., & Lo, N. (2012). Evidence for Permo-Triassic colonization of the deep sea by isopods. *Biology Letters*, *8*(6), 979–982. https://doi.org/10.1098/rsbl.2012.0774

Liu, H., Li, H., Song, F., Gu, W., Feng, J., Cai, W., & Shao, R. (2017). Novel insights into mitochondrial gene rearrangement in thrips (Insecta: Thysanoptera) from the grass thrips, Anaphothrips obscurus. *Scientific Reports*, *7*(1), 4284. https://doi.org/10.1038/s41598-017-04617-5

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, *4*(12), 981–994. https://doi.org/10.1038/nrg1226

Luo, S., Valencia, C. A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., … Huang, T. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(51), 13039–13044. https://doi.org/10.1073/pnas.1810946115

Lutcavage, M. E., Brill, R. W., Skomal, G. B., Chase, B. C., Goldstein, J. L., & Tutein, J. (2000). Tracking adult North Atlantic bluefin tuna (Thunnus thynnus) in the northwestern Atlantic using ultrasonic telemetry. *Marine Biology*, *137*(2), 347–358. https://doi.org/10.1007/s002270000302

Macher, J.-N., Zizka, V. M. A., Weigand, A. M., & Leese, F. (2018). A simple centrifugation protocol for metagenomic studies increases mitochondrial DNA yield by two orders of magnitude. *Methods in Ecology and Evolution / British Ecological Society*, *9*(4), 1070–1074. https://doi.org/10.1111/2041-210X.12937

Ma, J., & Amos, C. I. (2012). Principal components analysis of population admixture. *PloS One*, *7*(7), e40115. https://doi.org/10.1371/journal.pone.0040115

Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, *26*(9), 448–456. https://doi.org/10.1016/j.tree.2011.05.007

Martel, A., & Chia, F. S. (1991). Drifting and dispersal of small bivalves and gastropods with direct development. *Journal of Experimental Marine Biology and Ecology*, *150*(1), 131–147. https://doi.org/10.1016/0022-0981(91)90111-9

Martin, A., Quinn, K., & Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software, Articles*, *42*(9), 1–21. https://doi.org/10.18637/jss.v042.i09

McGaughran, A., Hogg, I. D., Stevens, M. I., Lindsay Chadderton, W., & Winterbourn, M. J. (2006). Genetic divergence of three freshwater isopod species from southern New Zealand. *Journal of Biogeography*, *33*(1), 23–30. https://doi.org/10.1111/j.1365-2699.2005.01338.x

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*(13), 3223–3231. https://doi.org/10.1111/mec.13243

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63. https://doi.org/10.1093/nar/gkz173

Merilaita, S. (1998). Crypsis through disruptive coloration in an isopod. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1401), 1059. https://doi.org/10.1098/rspb.1998.0399

Merilaita, S. (2001). Habitat heterogeneity, predation and gene flow: colour polymorphism in the isopod, Idotea baltica. *Evolutionary Ecology*, *15*(2), 103–116. https://doi.org/10.1023/A:1013814623311

Morin, P. A., Leduc, R. G., Archer, F. I., Martien, K. K., Huebinger, R., Bickham, J. W., & Taylor, B. L. (2009). Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources*, *9*(2), 498–504. https://doi.org/10.1111/j.1755-0998.2008.02502.x

Morin, P. A., Martien, K. K., & Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, *9*(1), 66–73. https://doi.org/10.1111/j.1755-0998.2008.02392.x

Morrison, C. L., Harvey, A. W., Lavery, S., Tieu, K., Huang, Y., & Cunningham, C. W. (2002). Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. *Proceedings. Biological Sciences / The Royal Society*, *269*(1489), 345–350. https://doi.org/10.1098/rspb.2001.1886Morton, J. E., & Miller, M. (1973). *The New Zealand Sea Shore*. Retrieved from https://researchspace.auckland.ac.nz/handle/2292/6658

Mousavi, M., Tong, C., Liu, F., Tao, S., Wu, J., Li, H., & Shi, J. (2016). De novo SNP discovery and genetic linkage mapping in poplar using restriction site associated DNA and whole-genome sequencing technologies. *BMC Genomics*, *17*, 656. https://doi.org/10.1186/s12864-016-3003-9

Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods in Ecology and Evolution / British Ecological Society*, *202*, 389. https://doi.org/10.1111/2041-210X.13252

Nagel, M. M., Sewell, M. A., & Lavery, S. D. (2015). Differences in population connectivity of a benthic marine invertebrate Evechinus chloroticus (Echinodermata: Echinoidea) across large and small spatial scales. *Conservation Genetics* , *16*(4), 965–978. https://doi.org/10.1007/s10592-015-0716-2

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*, 184–194. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-0998.2011.02987.x

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S. R., Staiger, E. A., Rashidi, A., & Huson, H. J. (2018). Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. *Frontiers in Genetics*, *9*, 105. https://doi.org/10.3389/fgene.2018.00105

Nikula, R., Fraser, C. I., Spencer, H. G., & Waters, J. M. (2010). Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Marine Ecology Progress Series*, *405*, 221–230. https://doi.org/10.3354/meps08523

Nokelainen, O., Stevens, M., & Caro, T. (2018). Colour polymorphism in the coconut crab (Birgus latro). *Evolutionary Ecology*, *32*(1), 75–88. https://doi.org/10.1007/s10682-017-9924-1

Nosil, P. (2007). Divergent host plant adaptation and reproductive isolation between ecotypes of Timema cristinae walking sticks. *The American Naturalist*, *169*(2), 151–162. https://doi.org/10.1086/510634

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*(5), 646–649. https://doi.org/10.1038/ng.139

Nunez, J. C. B., & Oleksiak, M. F. (2016). A Cost-Effective Approach to Sequence Hundreds of Complete Mitochondrial Genomes. *PloS One*, *11*(8), e0160958. https://doi.org/10.1371/journal.pone.0160958

Okie Jordan G., Smith Val H., & Martin-Cereceda Mercedes. (2016). Major evolutionary transitions of life, metabolic scaling and the number and size of mitochondria and chloroplasts. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1831), 20160611. https://doi.org/10.1098/rspb.2016.0611

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O'hara, R. B., Simpson, G. L., … Wagner, H. (2010). Vegan: community ecology package. R package version 1.17-4. *Http://cran. R-Project. Org>. Acesso Em*, *23*, 2010. Retrieved from https://kityna.ga/548893pdf.pdf

Palamara, P. F. (2014). Population genetics of identity by descent. Retrieved from http://arxiv.org/abs/1403.4987

Palma, A. T., & Steneck, R. S. (2001). Does Variable Coloration in Juvenile Marine Crabs Reduce Risk of Visual Predation? *Ecology*, *82*(10), 2961–2967. https://doi.org/10.2307/2679974

Pante, E., & Simon-Bouhet, B. (2013). marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PloS One*, *8*(9), e73051. https://doi.org/10.1371/journal.pone.0073051

Peccoud, J., Chebbi, M. A., Cormier, A., Moumen, B., Gilbert, C., Marcadé, I., … Cordaux, R. (2017). Untangling Heteroplasmy, Structure, and Evolution of an Atypical Mitochondrial Genome by PacBio Sequencing. *Genetics*, *207*(1), 269–280. https://doi.org/10.1534/genetics.117.203380

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. https://doi.org/10.1111/1755-0998.12129

Peters, J. C., Waters, J. M., Dutoit, L., & Fraser, C. I. (2020). SNP analyses reveal a diverse pool of potential colonists to earthquake-uplifted coastlines. *Molecular Ecology*, *29*(1), 149–159. https://doi.org/10.1111/mec.15303

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: an R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557–561. https://doi.org/10.1111/1755-0998.12323

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, *9*(3), e1000602. https://doi.org/10.1371/journal.pbio.1000602

Piertney, S. B., & Carvalho, G. R. (1994). Microgeographic genetic differentiation in the intertidal isopod Jaera albifrons Leach. I. Spatial distribution of allozyme variation. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society* . Retrieved from http://rspb.royalsocietypublishing.org/content/256/1346/195.short

Poore, G. C. B., & Bruce, N. L. (2012). Global Diversity of Marine Isopods (Except Asellota and Crustacean Symbionts). *PloS One*, *7*(8). https://doi.org/10.1371/journal.pone.0043529

Prezant, R. S., & Chalermwat, K. (1984). Flotation of the Bivalve Corbicula fluminea as a Means of Dispersal. *Science*, *225*(4669), 1491–1493. https://doi.org/10.1126/science.225.4669.1491

Pritchard, J. K., Wen, W., & Falush, D. (2010). *Documentation for STRUCTURE software: Version 2.3*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.323.9675&rep=rep1&type=pdf

Puebla, O., Bermingham, E., & Guichard, F. (2009). Estimating dispersal from genetic isolation by distance in a coral reef fish (Hypoplectrus puella). *Ecology*, *90*(11), 3087–3098. https://doi.org/10.1890/08-0859.1

Puritz, J. B., Keever, C. C., Addison, J. A., Barbosa, S. S., Byrne, M., Hart, M. W., … Toonen, R. J. (2017). Life-history predicts past and present population connectivity in two sympatric sea stars. *Ecology and Evolution*, Vol. 7, pp. 3916–3930. https://doi.org/10.1002/ece3.2938

Queiroz, N., Humphries, N. E., Mucientes, G., Hammerschlag, N., Lima, F. P., Scales, K. L., … Sims, D. W. (2016). Ocean-wide tracking of pelagic sharks reveals extent of overlap with longline fishing hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(6), 1582–1587. https://doi.org/10.1073/pnas.1510090113

Reich, D., Price, A. L., & Patterson, N. (2008). [Review of *Principal component analysis of genetic data*]. *Nature genetics*, *40*(5), 491–492. https://doi.org/10.1038/ng0508-491

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics: TIG*, *16*(6), 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2

Robin, E. D., & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of Cellular Physiology*, *136*(3), 507–513. https://doi.org/10.1002/jcp.1041360316

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*. https://doi.org/10.1111/mec.15253

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*, 94. https://doi.org/10.1186/1471-2148-12-94

Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current Biology: CB*, *27*(21), R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015

Ross, P. M., Hogg, I. D., Pilditch, C. A., Lundquist, C. J., & Wilkins, R. J. (2012). Population Genetic Structure of the New Zealand Estuarine Clam Austrovenus stutchburyi (Bivalvia: Veneridae) Reveals Population Subdivision and Partial Congruence with Biogeographic Boundaries. *Estuaries and Coasts*, *35*(1), 143–154. https://doi.org/10.1007/s12237-011-9429-z

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9093870

Rueffler, C., Van Dooren, T. J. M., Leimar, O., & Abrams, P. A. (2006). Disruptive selection and then what? *Trends in Ecology & Evolution*, *21*(5), 238–245. https://doi.org/10.1016/j.tree.2006.03.003

Sahyoun, A. H., Bernt, M., Stadler, P. F., & Tout, K. (2014). GC skew and mitochondrial origins of replication. *Mitochondrion*, *17*, 56–66. https://doi.org/10.1016/j.mito.2014.05.009

Salemaa, H. (2009). Geographical variability in the colour polymorphism of Idotea baltica (Isopoda) in the northern Baltic. *Hereditas*, Vol. 88, pp. 165–182. https://doi.org/10.1111/j.1601-5223.1978.tb01619.x

Schofield, G., Bishop, C. M., MacLean, G., Brown, P., Baker, M., Katselidis, K. A., … Hays, G. C. (2007). Novel GPS tracking of sea turtles as a tool for conservation management. *Journal of Experimental Marine Biology and Ecology*, *347*(1), 58–68. https://doi.org/10.1016/j.jembe.2007.03.009

Sefc, K. M., Brown, A. C., & Clotfelter, E. D. (2014). Carotenoid-based coloration in cichlid fishes. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *173C*, 42–51. https://doi.org/10.1016/j.cbpa.2014.03.006

Shears, N. T., Smith, F., Babcock, R. C., Duffy, C. A. J., & Villouta, E. (2008). Evaluation of biogeographic classification schemes for conservation planning: application to New

Zealand's coastal marine environment. *Conservation Biology: The Journal of the Society for Conservation Biology*, *22*(2), 467–481. https://doi.org/10.1111/j.1523-1739.2008.00882.x

Shuster, S. M., Embry, S. J., Hargis, C. R., & Nimer, A. (2014). The Inheritance of Autosomal and Sex-Linked Cuticular Pigmentation Patterns in the Marine Isopod, Paracerceis sculpta (Isopoda: Sphaeromatidae). *Journal of Crustacean Biology: A Quarterly of the Crustacean Society for the Publication of Research on Any Aspect of the Biology of Crustacea*, *34*(4), 460–466. https://doi.org/10.1163/1937240X-00002251

Shuster, S. M., & Levy, L. (1999). Sex-linked inheritance of a cuticular pigmentation marker in the marine isopod, Paracerceis sculpta Holmes (Crustacea: Isopoda: Sphaeromatidae). *The Journal of Heredity*, *90*(2), 304–307. https://nau.pure.elsevier.com/en/publications/sex-linked-inheritance-of-a-cuticular-pigmentation-marker-in-the-Slatkin, M. (1985). RARE ALLELES AS INDICATORS OF GENE FLOW. *Evolution; International Journal of Organic Evolution*, *39*(1), 53–65. https://doi.org/10.1111/j.1558-5646.1985.tb04079.x

Sponer, R., & Lessios, H. A. (2009). Mitochondrial Phylogeography of the Intertidal Isopod Excirolana braziliensis on the Two Sides of the Isthmus of Panama. *Proceedings of the Smithsonian Marine Science Symposium*, (August 2008).

Sponer, R., & Roy, M. S. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation. *Evolution; International Journal of Organic Evolution*, *56*(10), 1954–1967. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2002.tb00121.x/abstract

Stampar, S. N., Broe, M. B., Macrander, J., Reitzel, A. M., Brugler, M. R., & Daly, M. (2019). Linear Mitochondrial Genome in Anthozoa (Cnidaria): A Case Study in Ceriantharia. *Scientific Reports*, *9*(1), 6094. https://doi.org/10.1038/s41598-019-42621-z

Stevens, M. I., & Hogg, I. D. (2004). Population genetic structure of New Zealand's endemic corophiid amphipods: evidence for allopatric speciation. *Biological Journal of the Linnean Society. Linnean Society of London*, *81*(1), 119–133. https://doi.org/10.1111/j.1095-8312.2004.00270.x

Stevens, M., Lown, A. E., & Wood, L. E. (2014). Camouflage and individual variation in shore crabs (Carcinus maenas) from different habitats. *PloS One*, *9*(12), e115586. https://doi.org/10.1371/journal.pone.0115586

Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, *100*(2), 158–170. https://doi.org/10.1038/sj.hdy.6800937

Sullivan, K. A. M., Platt, R. N., Bradley, R. D., & Ray, D. A. (2017). Whole mitochondrial genomes provide increased resolution and indicate paraphyly in deer mice. *BMC Zoology*, *2*(1), 11. https://doi.org/10.1186/s40850-017-0020-3

Taanman, J. W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta*, *1410*(2), 103–123. https://doi.org/10.1016/s0005-2728(98)00161-3

Takahashi, Y., & Kawata, M. (2013). A comprehensive test for negative frequency-dependent selection. *Population Ecology*, *55*, 499–509. https://doi.org/10.1007/s10144-013-0372-7

Takahashi, Y., Yoshimura, J., Morita, S., & Watanabe, M. (2010). Negative frequency-dependent selection in female color polymorphism of a damselfly. *Evolution; International Journal of Organic Evolution*, *64*(12), 3620–3628. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2010.01083.x

Tamura, K., & Aotsuka, T. (1988). Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. *Biochemical Genetics*, *26*(11-12), 815–819. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3242492

Temunović, M., Franjić, J., Satovic, Z., Grgurev, M., Frascaria-Lacoste, N., & Fernández-Manjarrés, J. F. (2012). Environmental heterogeneity explains the genetic structure of Continental and Mediterranean populations of Fraxinus angustifolia Vahl. *PloS One*, *7*(8), e42764. https://doi.org/10.1371/journal.pone.0042764

Teske, P., Papadopoulos, I., Zardi, G., McQuaid, C., Edkins, M., Griffiths, C., & Barker, N. (2007). Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: planktonic, abbreviated and direct development. *Marine Biology*, *152*(3), 697–711. https://doi.org/10.1007/s00227-007-0724-y

Thorrold, S. R., Jones, G. P., Hellberg, M. E., Burton, R. S., Swearer, S. E., Neigel, J. E., … Warner, R. R. (2002). Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science*, *70*(1), 291–308. Retrieved from https://www.ingentaconnect.com/content/umrsmas/bullmar/2002/00000070/a00101s1/art00004

Tripp, E. A., Tsai, Y.-H. E., Zhuang, Y., & Dexter, K. G. (2017). RADseq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, *7*(19), 7920–7936. https://doi.org/10.1002/ece3.3274

Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., & Corbett, C. R. (2018). Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports*, *8*(1), 10931. https://doi.org/10.1038/s41598-018-29334-5

Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., … Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, Placopecten magellanicus, using RADseq-derived SNPs. *Evolutionary Applications*, *10*(1), 102–117. https://doi.org/10.1111/eva.12432

Vavrek, M. J. (2011). Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* , *14*(1), 16. Retrieved from https://www.uv.es/pe/2011_1/238/238.pdf

Veale, A. J., & Lavery, S. D. (2012). The population genetic structure of the waratah anemone (Actinia tenebrosa) around New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *46*(4), 523–536. https://doi.org/10.1080/00288330.2012.730053

Vendrami, D. L. J., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., … Hoffman, J. I. (2017). RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society Open Science*, *4*(2), 160548. https://doi.org/10.1098/rsos.160548

Verity, R., & Nichols, R. A. (2016). Estimating the Number of Subpopulations (K) in Structured Populations. *Genetics*, *203*(4), 1827–1839. https://doi.org/10.1534/genetics.115.180992

Wade, N. M., Melville-Smith, R., Degnan, B. M., & Hall, M. R. (2008). Control of shell colour changes in the lobster, Panulirus cygnus. *The Journal of Experimental Biology*, *211*(Pt 9), 1512–1519. https://doi.org/10.1242/jeb.012930

Wang, J. (2011). COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*(1), 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Wang, Y., Cao, X., Zhao, Y., Fei, J., Hu, X., & Li, N. (2017). Optimized double-digest genotyping by sequencing (ddGBS) method with high-density SNP markers and high genotyping accuracy for chickens. *PloS One*, *12*(6), e0179073. https://doi.org/10.1371/journal.pone.0179073

Waples, R. S. (2015). Testing for Hardy-Weinberg proportions: have we lost the plot? *The Journal of Heredity*, *106*(1), 1–19. https://doi.org/10.1093/jhered/esu062

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Wei, D.-D., Shao, R., Yuan, M.-L., Dou, W., Barker, S. C., & Wang, J.-J. (2012). The multipartite mitochondrial genome of Liposcelis bostrychophila: insights into the evolution of mitochondrial genomes in bilateral animals. *PloS One*, *7*(3), e33973. https://doi.org/10.1371/journal.pone.0033973

Weiser, E. L., & Powell, A. N. (2011). Evaluating gull diets: a comparison of conventional methods and stable isotope analysis. *Journal of Field Ornithology*, *82*(3), 297–310.

Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1557-9263.2011.00333.x?casa_token=JXMTuwCBSwgAAAAA:Vud_XbzDanREY8ijDJ45VITBZIldpehO61DkkhhlyEe6WVq0gjFj2bnp_U2XFgUpT25qZToag4doxr7f

Wellenreuther, M., Svensson, E. I., & Hansson, B. (2014). Sexual selection and genetic colour polymorphisms in animals. *Molecular Ecology*, *23*(22), 5398–5414. https://doi.org/10.1111/mec.12935

Wells, S. J., & Dale, J. (2018). Contrasting gene flow at different spatial scales revealed by genotyping-by-sequencing in Isocladus armatus, a massively colour polymorphic New Zealand marine isopod. *PeerJ*, Vol. 6, p. e5462. https://doi.org/10.7717/peerj.5462

Wetzer, R., Bruce, N. L., & Pérez-Losada, M. (2018). Relationships of the Sphaeromatidae genera (Peracarida: Isopoda) inferred from 18S rDNA and 16S rDNA genes. *Arthropod Systematics & Phylogeny*, *76*(1), 1–30.

White, R., Pellefigues, C., Ronchese, F., Lamiable, O., & Eccles, D. (2017). Investigation of chimeric reads using the MinION. *F1000Research*, *6*, 631. https://doi.org/10.12688/f1000research.11547.2

White, T. E., & Kemp, D. J. (2016). Colour polymorphism. *Current Biology: CB*, *26*(13), R517–R518. https://doi.org/10.1016/j.cub.2016.03.017

Whitlock, M. C. (2011). and D do not replace FST. *Molecular Ecology*, *20*(6), 1083–1091. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2010.04996.x

Will, M., Hale, M. L., Schiel, D. R., & Gemmell, N. J. (2011). Low to moderate levels of genetic differentiation detected across the distribution of the New Zealand abalone, Haliotis iris. *Marine Biology*, *158*(6), 1417–1429. https://doi.org/10.1007/s00227-011-1659-x

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, *163*(3), 1177–1191. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12663554

Wolff, J. N., Shearman, D. C. A., Brooks, R. C., & Ballard, J. W. O. (2012). Selective enrichment and sequencing of whole mitochondrial genomes in the presence of nuclear encoded mitochondrial pseudogenes (numts). *PloS One*, *7*(5), e37142. https://doi.org/10.1371/journal.pone.0037142

Woods, C. M. C. (1993). Natural diet of the crab Notomithrax ursus (Brachyura: Majidae) at Oaro, South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *27*(3), 309–315. https://doi.org/10.1080/00288330.1993.9516571

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*(2), 97–159. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17246615

Yang, E. C., Kim, K. M., Kim, S. Y., Lee, J., Boo, G. H., Lee, J.-H., … Yoon, H. S. (2015). Highly Conserved Mitochondrial Genomes among Multicellular Red Algae of the Florideophyceae. *Genome Biology and Evolution*, *7*(8), 2394–2406. https://doi.org/10.1093/gbe/evv147

Yang, M., Gao, T., Yan, B., Chen, X., & Liu, W. (2019). Complete mitochondrial genome and the phylogenetic position of a wood-boring Isopod Sphaeroma terebrans (Crustacea, Isopod, Sphaeromatidae). *Mitochondrial DNA Part B*, *4*(1), 1920–1921. https://doi.org/10.1080/23802359.2019.1613181

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* , *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zou, H., Jakovlić, I., Zhang, D., Chen, R., Mahboob, S., Al-Ghanim, K. A., … Wang, G.-T. (2018). The complete mitochondrial genome of Cymothoa indica has a highly rearranged gene order and clusters at the very base of the Isopoda clade. *PloS One*, *13*(9), e0203089. https://doi.org/10.1371/journal.pone.0203089