A Nation-wide Phylogenetic and Phylogeographic Investigation

of the Endemic New Zealand Oyster Borers, *Haustrum scobina* and *Haustrum*

albomarginatum

By

Seamus O'Mahony

A thesis

submitted to the Victoria University of Wellington

in partial fulfilment of the requirements for the degree of

Master of Science in

Ecology & Biodiversity

Victoria University of Wellington

2020

Acknowledgements

I am lucky to have an encouraging network of people around me, who have made the past two years a fulfilling and enjoyable experience. First and foremost, my thanks must go to Nina Eastgate for her support over the past two years, and particularly for taking care of all the chores and cooking I neglected in the final weeks of this work.

My supervisors Peter Ritchie and Nicole Phillips have been invaluable in guiding this research. I am especially grateful for the freedom I had to investigate what interested me the most, which made this research a genuinely enjoyable time.

Kerry Walton has been an extremely valuable source of information and guidance. I never thought I would meet a person who knew so much about snails, and without his expertise, this work would have been a frustrating ordeal to say the least.

Thanks must go to my lab mates for their support over the past year. Leah Kemp was a great mentor in the lab. I would have been overwhelmed by the data wrangling, interpretation and confusing computer jargon without the guidance of Yvan Papa, Tom Oosting and Guillermo Rodriguez.

My gratitude goes to Bruce Marshall for supplying samples from Te Papa, and to Claire Elliot and Meredith Walton who collected samples for me. Thanks is due to the authors of Barco *et al.* (2015) and Barco *et al*. (2010) for the sequences of theirs used in this study. Sebastian Logan deserves huge recognition for contributing many sequences to this study too. An honorary mention must go to George Rzoska and his family for having me stay during a sampling trip and allowing me to store strange marine creatures in their freezer.

I would like to thank Victoria University of Wellington for funding this work through the Victoria Masters by Thesis Scholarship. The attention I gave this research would not have been possible without it.

Finally, a huge thank you must go to my parents, John and Antonia O'Mahony. Without their encouragement and financial support, this accomplishment would have been a mere dream.

Abstract

The endemic New Zealand sea snails *Haustrum scobina* and *Haustrum albomarginatum* are rocky shore intertidal dogwhelks of the Muricidae family. They have direct developing young and are carnivores. Their radula is used to drill into the shells of their prey, and they are commonly referred to as oyster borers. The taxonomic status of these species is still unresolved and therefore the name *Haustrum scobina sensu lato* is used.

The overall goal of this thesis research was to investigate the phylogeny and phylogeography of *Haustrum scobina sensu lato* using mitochondrial DNA sequences. Comparisons made to phylogeographic studies of ecologically similar species such as *Cominella* spp. provide an opportunity to identify the common environmental determinates of population migration route, genetic differentiation and speciation whenever similar patterns are found.

A nation-wide collection of samples was used to generate 277 new sequences from a 610 bp portion of the cytochrome c oxidase subunit I (COI) gene. This enabled the formation of a dataset of 654 DNA sequences, which was comprised of the 277 new sequences, 16 retrieved from a published study that deposited them in GenBank, and 361 from a previous unpublished thesis study. An unexpectedly diverse phylogeny of 58 COI haplotypes from 31 sample sites was recovered. These formed three clusters using K-means clustering by pairwise mutational distance. The in-group species did not form reciprocal monophyly groups, and the expected closest outgroup species (*Haustrum haustorium*) appeared to be as similar to the in-group clusters as they were to each other. A dataset of 27 DNA sequences from an 827 bp portion of the large sub-unit 28S nuclear rRNA gene was produced with the intention of corroborating the findings from the analyses of the COI dataset. This consisted of 26 new sequences and one sequence from a published study that deposited the sequence on GenBank. The expected taxonomic arrangement of *Haustrum scobina sensu lato* could not be matched by COI sequences due to incongruence with the 28S phylogeny and shell morphology.

The 28S dataset and shell morphology indicated there are two species in *Haustrum scobina sensu lato*. These are most likely *Haustrum scobina* and *Haustrum albomarginatum*, but they could not clearly be identified in the COI data. As a result, the phylogeographic certainty was limited when using the COI dataset because of the lack of clarity between the haplotypes of the two putative species. Possible reasons for the complicated COI dataset are discussed. Phylogenetic analysis of both the 28S and COI datasets did not support the expected conclusion that members of *Haustrum scobina sensu lato* are each other's closest relatives. *Haustrum haustorium* was the expected immediate outgroup species but formed a polytomy with the in-group.

A decrease in COI haplotype diversity was observed in southern samples when they were compared to the samples collected at northern locations. Taranaki sites shared a haplotype with multiple South Island sites that had no haplotype diversity. This suggested post-glacial re-colonisation of southern sites after displacement by ice-age conditions from these locations, a hypothesis consistent with results from studies of the *Cominella* genus. Association between Purau Bay in Lyttleton Harbour, Titahi Bay, Port Ahuriri and Kawau Island with no associated haplotypes between these locations suggested human-mediated translocation events. A genetic disjunction was also apparent between the south Wellington/Wairarapa coast and the eastern Wairarapa coast. This pattern was consistent with one study of *Cominella maculosa* and other studies have attributed similar patterns of other species in the region to recent uplift events affecting coastal community composition. The phylogeny of *Haustrum scobina sensu lato* will require further investigation before it can be used to more confidently resolve the phylogeographic history of the species.

Research Limitations

The research reported in this thesis comprises of a large and informative dataset. However, a phylogenetic issue was not resolved in the primary COI dataset. This limited the extent to which phylogeographic inferences could be made and limited the confidence in conclusions.

Normally, increasing the sample size or using a large dataset from a different marker would be the next step to resolve such issues. New Zealand enforced significant public restrictions as a result of the COVID-19 pandemic between the 25th of March 2020 and the 13th of May 2020. This meant that further lab work could not be conducted before the final analysis of the dataset presented in this thesis.

Contents

List of Tables

List of Figures

Chapter 1 - General Introduction

New Zealand is arguably one of the most suitable locations in the world to study the effects of environmental change on a species' distribution. New Zealand is characterised by some of the most volatile geological and environmental forces on the planet today (Newnham *et al.* 1999). Genetic structuring in contemporary populations can be used to understand how past geological, geographic, and environmental processes and change have structured populations. Where genetic breaks or connectivity is found, patterns can be associated with environmental influence. New Zealand's dynamic land and seascape means recent effects of environmental forces are still evident in the contemporary structuring of populations.

Direct developing species are those that have no larval stage, and direct developing benthos typically only have a planktonic phase if their egg-masses float (Kyle & Boulding 2000). Direct developing gastropods, for instance, are generally limited in their dispersal ability and rate. Studying the population genetics of these types of species can eliminate the somewhat random nature of dispersal patterns that may occur during a planktonic phase and give an opportunity to identify some of the consistent patterns of gene and genetic isolation. Contemporary population structure may also reveal how past processes have shaped a population. In species with highly realised dispersal ability, signals of historic environmental influence can be quickly lost with genetic homogenisation. Phylogenetic studies of direct developing taxa can reveal influences of historic environmental change, which may not be evident in species that are more mobile.

New Zealand Setting, Geography and Geology

As a unique archipelago consisting of over 700 islands, New Zealand hosts a vast array of differing ecological systems and species along its coastline, stretching from the Kermadecs at 29° S to Campbell Island at 52° S (Arranz Martinez 2017; Ross *et al.* 2009). These islands span from sub-Antarctic to sub-tropical latitudes. Oceanic

inflows interact with complex coastal currents and eddies, making predictions of where a planktonic vessel might travel near impossible (Heath 1985; Chiswell & Roemmich 1998; Chiswell & Booth 1999; Ross *et al.* 2009). Whilst such a varied seascape presents many caveats for marine researchers to overcome, it can also provide an ideal environment for studying the consequences varying life histories have for different species (Arranz Martinez 2017).

New Zealand boasts an impressive physical landscape, aptly described by Fleming (1975) as a narrow-gutted but lofty archipelago. Zealandia, the continent which barely breaches the ocean surface to make this archipelago, split from Australia 80 million years ago (Laird & Bradshaw 2004; Wallis & Trewick 2009). Today only 6% of Zealandia sits above sea level, these are the islands that form New Zealand (Mortimer *et al.* 2017). Geological processes of uplift and subsidence in combination with eustatic sea level change have continuously modified the terrestrial and coastal landscape making significant changes within timescales of millions- to thousands of years. New Zealand sits on the convergence zone between the Pacific and Australian plates, which produce an array of large fault lines such as the Alpine fault, one with a dextral strike-slip motion of up to 29 mm/yr, dip-slip motion of 7.8 mm/yr and large surface rupturing earthquakes at least every 380 years (Howarth *et al.* 2016). The recent 2016 Kaikoura earthquake resulted in highly varied vertical displacement ranging from -2.5 m to +6.5 m along the coast South Island's east coast (Clark *et al.* 2017). Evidence for historical uplift events and subsidence can be seen around New Zealand, such as the Whanganui marine terraces and terraces along the Wairarapa south coast (Grapes 1999; McSaveney *et al.* 2006; Pillans 2017). Sudden events like these can have profound effects on coastal communities and might even lead to significant loss of genetic diversity in these areas (McSaveney *et al.* 2006; Parvizi *et al.* 2019; Hay 2020).

Parts of New Zealand have become isolated and reconnected over millennia. This has split populations and been proposed as an explanation for speciation and genetic isolation that we now see in contemporary New Zealand (Winkworth *et al.* 1999; Craw *et al.* 2016). Regular glacial cycles caused by Milankovitch cycles of the earth's rotational axis and orbit around the sun bring global changes in ice-cover and eustatic sea level. The Last Glacial Maximum (LGM) onset in the Southern Hemisphere about

27 kya with the Last Isotope Maximum (LIM) occurring between 21.5 – 18 kya, before the climate began to warm again (Suggate & Almond 2005). The LGM reduced sea level by ~120 m relative to today (Lewis *et al.* 1994). The shallowest possible connection between the North and South Islands today is along Farewell Rise peaking at 94 m (Lewis *et al.* 1994). Sea level was reduced enough to close the Cook Strait between Taranaki and the South Island during the last glacial maximum (LGM) 20 kya (Proctor & Carter 1989; Lewis *et al.* 1994). This may have facilitated migration allowing North and South Island populations of terrestrial and coastal benthic species to share more genetic diversity. It may also have allowed South Island populations to move north during the LGM.

The fossil record in sedimentary basins like the Whanganui Basin often records a bias in species. With consideration to rocky reef species, they are often not fossilised since their habitat lacks the sediment to quickly preserve their remains (Beu & Maxwell 1990). On the other hand, soft substrate marine species are over-represented in the New Zealand fossil record. The Whanganui basin is fed by 12 million tonnes of sediment from the north-west South Island and Whanganui, 6.7 million tonnes is carried northward around Farewell Spit and 8 million tonnes is fed into the south-eastern entrance to the Cook Strait each year (Lewis *et al.* 1994). This has resulted in a 2 km thick sedimentary basin (Pillans 2017). An extensive fossil record can be found in the Whanganui marine terraces (Stevens & Vella 1981; Beu & Maxwell 1990), but this mostly consists of sand and mud dwelling species (Beu & Maxwell 1990). While the basin and land bridge probably consisted mostly of sand, mud and gravel (Lewis *et al.* 1994; Pillans 2017), it is probable that some rocky outcrops were present in places allowing rocky reef species to survive as they do on the vast sandy beaches of South Taranaki today (pers. obs. 2019).

Beu *et al.* (2004) give evidence for a historical sea connection across the present-day Auckland Isthmus, between the Hauraki Gulf and Manukau Harbour. The Auckland Isthmus consists of relatively young sediment. Part of the present Auckland Isthmus might have been below sea level between 20 – 50 ka before volcanoes in the area erupted (Beu *et al.* 2004). Whilst the Auckland Isthmus was likely breached in the Pliocene (Beu *et al.* 2004; Stevens & Hogg 2004), dates are poorly constrained and it is still unknown whether it was breached during the Pleistocene (Beu *et al.* 2004;

Walton 2017). Marine organisms were also likely transported between the east and west sides of the Auckland Isthmus through human activity. Māori, for example, would drag waka between the east and west coasts across the isthmus (Fox 1977).

What we observe today is only a brief snapshot of current species' historical distribution and geographical barriers to dispersal. The influence of past events on ecological communities might be revealed through the genetic structure of populations. Areas of low genetic diversity, for example, may indicate recent and on-going re-colonisation after a disruptive event, such as sudden uplift or a period of cold climate. By examining the current range of a species and inferring a historical distribution, we can anticipate how it might react to future changes. Effective management of commercially exploited species and ecological conservation efforts have never been of greater concern than they are today. No society before us has knowingly faced the climactic challenges and loss of biodiversity that loom over us. Genetic assessment of populations can give unparalleled insight of how they are shaped by their environment.

Direct developing species are likely to be more sensitive to such influences than those with a planktonic stage which can, under the right circumstances, quickly re-colonise sites and recover biodiversity. Studies of direct developing intertidal species in a situation like New Zealand can provide useful information about general patterns of dispersal and isolation, which can be used to guide how we manage our marine biota in the future.
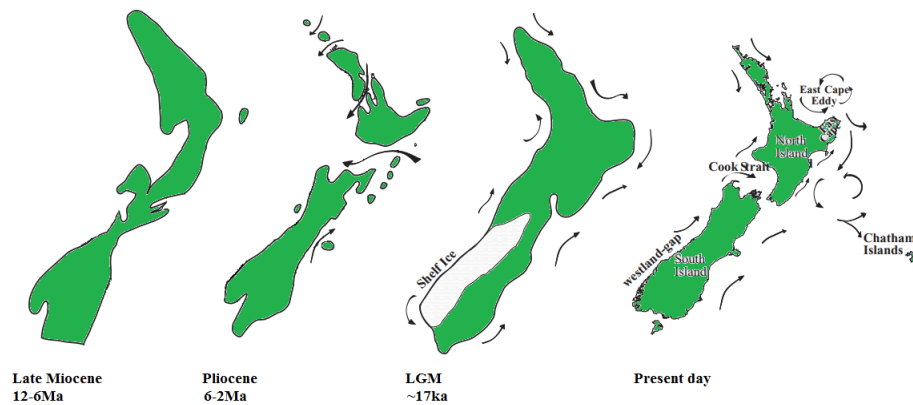
Figure 1.1 - New Zealand coastline from the late Miocene to present day. Green is land above sea level, off-white is shelf ice and black arrows show prevailing currents. Adapted from (Stevens & Hogg 2004).

In a more contemporary context, New Zealand boasts an impressively dynamic seascape. Hypothetically, a floating body could be carried anywhere around New Zealand in a relatively short period by coastal currents alone. Coastal currents are convoluted around New Zealand so it is difficult and arguably impossible to reliably predict where a planktonic body may be taken. Whilst some currents connect the entire archipelago, others circle local areas for years. The Wairarapa Eddy has been shown to retain *Jasus edwardsii* (spiny rock lobster) larvae for up to 12 months (Chiswell & Booth 1999). Converging oceanic inflows such as the Subtropical Convergence Zone can act as significant barriers to migration (Verry *et al.* 2020). Migrating individuals face large climactic changes across these areas, for which they are not always well adapted. Floating egg capsules and larvae may also be carried offshore by currents and never have an opportunity to settle in a suitable habitat.

Oceans currents are variable through time (Chiswell & Rickard 2011), this can make summarising their influence on dispersal difficult. The relative percentile of transfer between areas can also be difficult to determine due to confounding environmental and ecological influences such as temperature, salinity and predator-prey interactions (Chiswell & Rickard 2011). Sensible conclusions must be generated depending on the species being studied. For a species such as *Haustrum scobina*, where young are direct developing and eggs are cemented to rocks, currents are less likely to have a significant influence than other environmental and ecological factors.
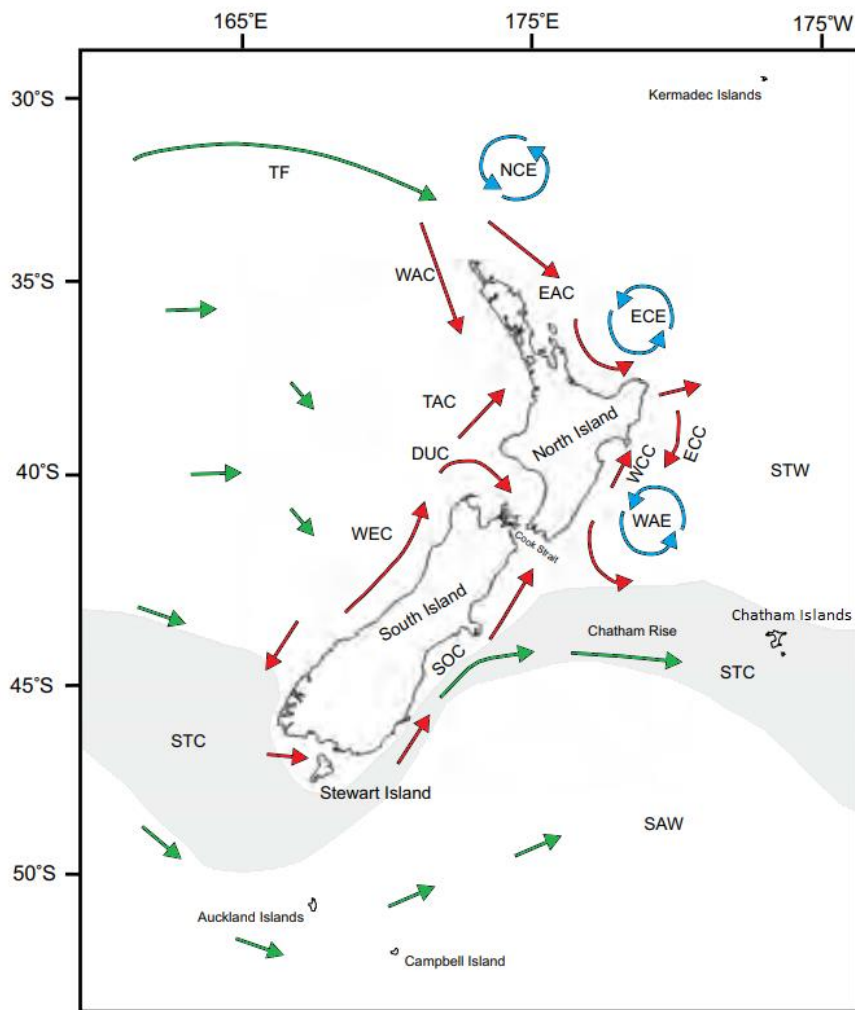
Figure 1.2 – Oceanic currents (green), coastal currents (red), eddies (blue) and convergence zones around New Zealand. DUC = D'Urville Current; EAC = East Auckland Current; ECC = East Cape Current; ECE = East Cape Eddy; NCE = North Cape Eddy; SOC = Southland Current; SAW = Sub-Antarctic water; STC = Subtropical Convergence; STW = Subtropical water; TAC = Tasman Current; TF = Tasman Front; WAC = West Auckland Current; WAE = Wairarapa Eddy; WCC = Wairarapa Coastal Current; WEC = Westland Current. Adapted from Ross *et al.* (2009).

Phylogenetics and Phylogeography

Phylogenetics is the study of the evolutionary relationship between a group of organisms. DNA sequence characters are widely used to construct phylogenetic trees. These methods of analysis can be applied to members of the same species or span

across a range of species and distantly related taxonomic groups. Gene sequencing provides a high-resolution method of comparing relationships between organisms and allows us to construct phylogenies on an intra-specific level. Phylogenetic relationships can be presented in several ways such as haplotype networks, but bifurcating trees are undoubtedly the most familiar representation to many. Presentation of the most likely connections in a dataset is an important goal of constructing phylogenies, and failure to consider how DNA sequences change or using incorrect assumptions can result in conflicting or misleading phylogenies. A phylogenetic tree for example, is a more simplified version of a haplotype network. A haplotype network will show the relation of all haplotypes to each other without organising them into hierarchical clades. A phylogenetic tree can make a group of haplotypes or species appear more closely related to each other than they really are. Different tree building methods (such as Bayesian or maximum likelihood) or using an inappropriate model of nucleotide selection might also produce slightly different phylogenies to each other (Hills *et al.* 2011). Often a variety of methods are used when constructing phylogenies so the different features of the dataset can be better understood, and the topology corroborated using different methodological approaches.

Phylogeography is the application of phylogenetic information to the study of a species' dispersal patterns and population structure. Population level studies typically compare diversity and shared haplotypes between regions. Indices of diversity and regional associations can be used to infer movement through time and space. These methods can provide detailed information of how the natural environment has influenced a species' evolution and distribution.

Marine Life-history Strategies and Dispersal

Early life-history strategies of marine organisms can generally be divided into three groups. Typically, direct developing species do not undergo metamorphosis after hatching. An example of such a species is *Haustrum scobina* (oyster borer) (Jones & Ross 2018). Lecithotrophic species have a short stage as planktonic larvae. The pelagic larval duration (PLD) of lecithotrophic species is typically limited by yolk

supply. *Haliotis iris* (pāua) is an example of a lecithotrophic species, where larvae tend to settle within five to nine days (Stephens *et al.* 2006). Finally, there are planktotrophic species, which spend an extended amount of time as planktonic larvae before settling and taking their adult form. *Jasus edwardsii* (spiny rock lobster) provides a prime example of this life-history strategy, spending 12-24 months as planktonic larvae before settling (Chiswell & Booth 1999). These are not necessarily discrete categories. Varied environmental influence such as temperature can determine the length of time for which larvae or eggs float (Stephens *et al.* 2006).

A phylogeographic analysis may expect to detect one of several types of population structure when analysing the distribution of genetic diversity. Panmixia describes a situation where there is no evidence of differentiation among sampled locations in a population. Haplotype frequencies are homogenized among all sites in this scenario. Panmixia is typically used as the null model in phylogeographic studies. Panmixia may be expected in populations with a high potential for dispersal (planktotrophic species) when there are no environmental factors constraining dispersal and gene flow. A pattern of panmixia will often be reported when the genetic marker used in a study does not provide sufficient intraspecific resolution.

An isolation by distance (IBD) pattern describes a situation where there is a correlation between haplotype frequencies and distance between sampled locations. When suitable methods have been applied, patterns like this can reveal influential environmental processes. An IBD pattern typically indicates that restriction to gene flow increases over longer distances and the further apart subpopulations are from each other geographically, the further they are from each other genetically. It is important to test whether the marker used in the analysis is neutral to selection, and analysis methods are impartial before making a conclusion about this pattern.

A similar pattern to IBD may be seen where there are geographic clines. These are areas where geographic and environmental conditions change over a spatial gradient and genetic distance between subpopulations increases over the geographic cline. Genetic patterns resulting from geographic clines may only be evident if the locus analysed is under selection for the varying environmental factor. Geographic clines

may be mistaken for IBD if the genetic marker and analysis methods are not appropriate for the intended purpose.

A distinct phylogeographic break is the third pattern that could be observed. This could also be comprised of several breaks. In this scenario, haplotype frequencies between populations rapidly change across a narrow geographic area. Phylogeographic breaks tend to indicate there is a barrier to gene flow between sampled sites. A genetic barrier could result from several factors such as ocean currents, physical barriers, dispersal power, and life history characteristics.

Finally, a chaotic pattern may be observed. A chaotic pattern is revealed where some haplotypes are specific to some locations but not found in others, with no apparent geographic correlation. Fine scale processes may cause a chaotic pattern, including phenomena like local eddies, wave action retaining larvae or river mouths transporting larvae out past the surf zone into oceanic currents.
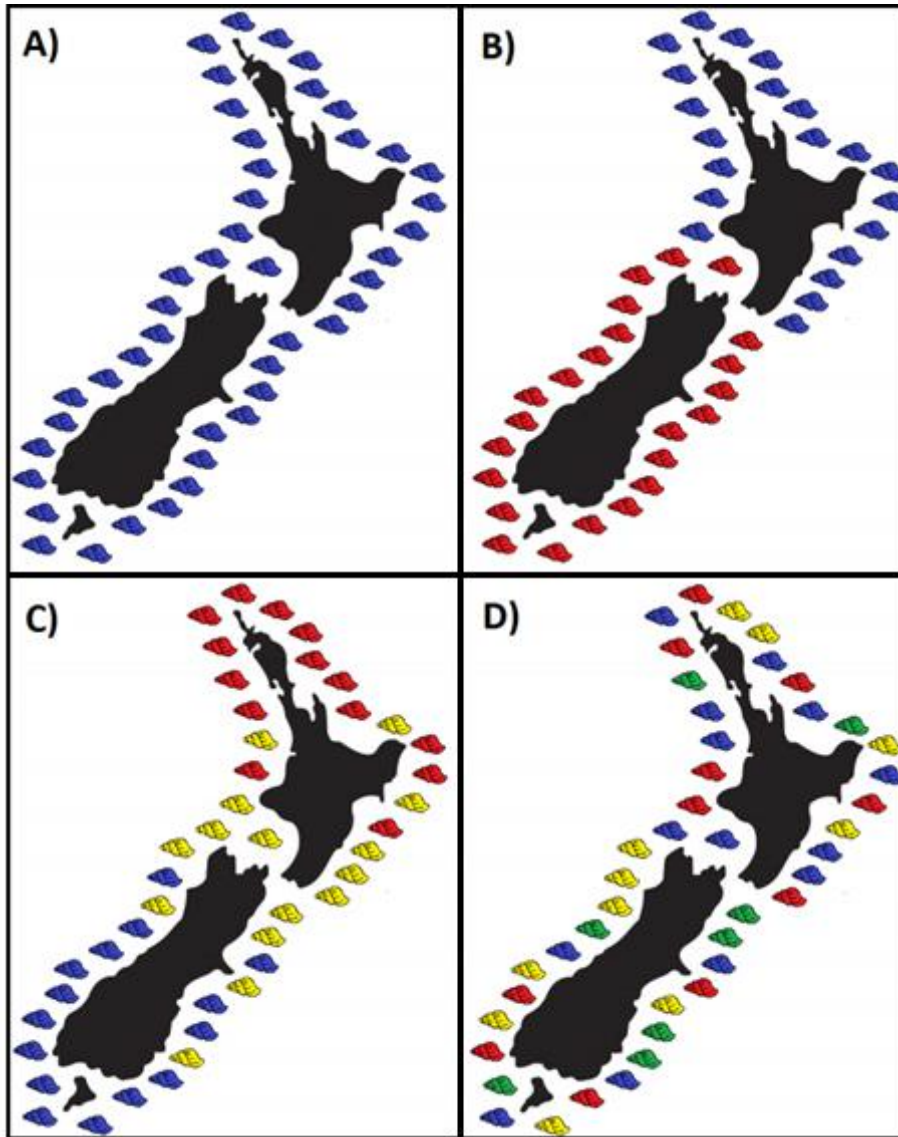
Figure 1.3 – Patterns of population structure where shell colour represents distribution of haplotypes at a site. A) Panmictic, high gene flow maintains genetic homogeneity. B) A distinct break, where gene flow is limited across a specific geographical point. C) Isolation by distance, gene flow is limited by distance and dispersal ability. D) Chaotic pattern, patchy distribution of haplotypes. This also represents a panmictic pattern with high diversity. Adapted from Logan (2019).

The marine environment enables a vast array of life-history strategies. Given its superficially open nature, traditional paradigms in marine sciences have maintained that high levels of gene flow are common across large distances (Cowen & Sponaugle 2009; Puritz *et al.* 2017). Closely related to this line of thought is the belief that population structure can be predicted by the PLD of a species, where a shorter PLD correlates to less gene flow between sub-populations, and hence a more distinctively structured metapopulation. Whilst this is a reasonable assertion, the pattern rarely holds and PLD is not always a reliable predictor of gene flow (Weersing & Toonen

2009). When one fully considers the complex nature and variability of the marine environment, it becomes apparent why it is difficult to pinpoint the primary explanatory factor for a pattern of genetic diversity.

The more obvious variables which might negate the influence of PLD on population structure include ocean currents and local eddies either transporting or retaining larvae and adult species (Chiswell & Roemmich 1998; Chiswell & Booth 1999; Galarza *et al.* 2009). Other factors such as founder effects and competition (Waters *et al.* 2013), local selection pressure (Gardner *et al.* 2010) and habitat patchiness (Pinsky *et al.* 2012) can all have a significant influence. Of course, that is not to say PLD does not have any influence on population structure, but that three-dimensional systems should not be approached with such two-dimensional assumptions. Furthermore, Ross *et al.* (2009) found that the relationship between genetic differentiation and PLD breaks down in species with a particularly short, or no PLD. This suggests that environmental and ecological factors play a more significant role in shaping migration than geographical distance, especially in species with a short PLD.

Across New Zealand literature there is evidence that species specific traits and local environmental effects are likely to have a more pronounced effect on the dispersal ability of species with shorter larval durations (Ross *et al.* 2009). Such variation means there is often a consistent pattern to be found in studies of direct developers. To date in New Zealand, direct developers have shown evidence of leapfrog type dispersal, where they are most closely related to subpopulations not directly adjacent to them (Dohner *et al.* 2018), long distance translocation (Fleming *et al.* 2017) and post-glacial maxima re-colonisation of southern regions (Fleming *et al.* 2017; Dohner *et al.* 2018; Walton *et al.* 2018). Many of these studies only provide a preliminary indication of these processes, with nearly all of them stating further research is required. For example, if a leapfrog pattern is observed with a slow evolving marker, the same populations should be studied again with a faster evolving marker. If a more linear pattern is found with the faster evolving marker, the leapfrog pattern can be disregarded and explained by differences in haplotype variation due to a higher mutation rate. It is also necessary to replicate, closely mirror or expand on such studies using a variety of different species. To find an interesting phenomenon in one species gives little indication of how environmental processes may have influenced

the phenomenon. Factors including differing life-history strategies, competition and niche specification must be controlled for by comparing results between co-distributed but evolutionarily independent species (Teske *et al.* 2011; Keeney *et al.* 2013).

Demographic and Genetic Connectivity

Genetic measures of connectivity between subpopulations have an advantage over direct observations or measurements of migration because it is easier to subsample individuals from many populations over large spatial scales. Genetic studies also offer a significant advantage over methods such as mark-recapture and tracking as they can reveal whether migrating individuals have actually bred and transferred genes between populations. Results of genetic studies can often produce misleading results though. Genetic connectivity does not necessarily mean there is a significant or important demographic link (the dependence of populations upon each other for diversity and growth). Demographic connectivity is dependent on the proportion of reproductively successful migrants per generation by population size (Lowe & Allendorf 2010; Ovenden 2013). The absolute number of migrants alone is responsible for genetic connectivity, and relatively few migrants per generation can homogenise genes between subpopulations (Ovenden 2013).

Another relevant issue is how we define a population. This is a difficult term to define, as populations can be defined subjectively depending on the context. Broadly, the issue of how a population is defined can be divided into the ecological paradigm, where demographic cohesion is the most important factor, and the evolutionary paradigm, where reproductive cohesion is the most important factor (Waples & Gaggiotti 2006). Waples & Gaggiotti (2006) argue that neither line of thought alone is sufficient. The ecological paradigm simply relies on the migration rate, and the evolutionary paradigm relies on the number of reproductively successful migrants per generation. More complex algorithms are necessary to denote populations, and methods of testing for population structure without *a priori* allocations should be employed if possible. Even still, genetic methods are limited in analysis under the

ecological paradigm as the boundary between demographic dependence and independence occurs where there are high rates of migration, and genetic markers may not be able to distinguish significant differences between subpopulations under such conditions (Waples & Gaggiotti 2006).

Ovenden (2013) proposes the concept of 'crinkled populations'. This describes a scenario in which subpopulations are genetically linked, but migration rates are not high enough to classify them as a single population or demographically dependant on each other. Ovenden (2013) suggests combining genetic data with other data such as mark/recapture studies in order to directly quantify demographic links. Whilst not always possible to conduct such studies in conjunction with genetic studies, they should be considered in follow-up research. Comparisons of studies from similar species may also be useful in interpreting genetic data and estimating the significance of any findings in a genetic study.

The scope of this thesis research is to investigate genetic connectivity between samples of a direct developer collected from around the New Zealand coastline. For this reason, statistical analysis methods must be carefully chosen. For example, using analysis of molecular variance (AMOVA) methods would require *a priori* assignment of subpopulations, which can impart bias in the results. Arguably it is best to use a cluster analysis algorithm which can group haplotypes without any prior assignments or assumptions. From there we can then begin to investigate whether there is a significant correlation between clusters of haplotypes and geographical location by means such as AMOVA. If this is not the case, then we cannot confidently state that a meta-population is split into defined subpopulations. In such a case, mark-recapture investigation into how individuals are moving may reveal actual rates of migration. Genetic connectivity might be a misleading artefact of small numbers of migrants homogenising genes between subpopulations.

In this thesis, reference to populations is made where there is genetic association between a particular set of sample sites within a given geographical range. For example, Taranaki is referred to as a population as there are three sample sites with high frequencies of shared haplotypes. The South Island samples are also referred to as a population, even though they occur over a much larger area. This is because all

South Island sites had a very low degree of diversity and all shared a common haplotype. The entire distribution of the in-group is referred to as a metapopulation.

Choice of Genetic Marker

Different genetic markers can give different types of information about a species' phylogeny and phylogeography. The different sizes of the dataset and combinations of DNA sequence loci can improve resolution power, though a balance must always be struck between cost, practicality, and the efficacy of the type of data that is used. Over time, more advanced DNA sequencing techniques have become cheaper, superseding older methods and datasets. For example, many low-budget studies will still choose to use mitochondrial DNA (mtDNA) sequencing over large-scale genome sequencing because of costs and the availability of comparable data from other populations and species. The application of universal polymerase chain reaction (PCR) primers has allowed genetic studies to combine datasets among separate studies and improve on previous findings. One such example being where significant genetic structuring in *Perna canaliculus* (New Zealand greenshell mussels) was found using Random Amplification of Polymorphic DNA (RAPDs) and mtDNA sequencing (NADH IV region) where no structure had previously been observed using allozyme electrophoresis (Apte & Gardner 2001; Apte & Gardner 2002; Star *et al.* 2003). Generally, a highly variable or 'fast evolving' mitochondrial DNA marker will work for analysing intraspecific phylogenies or phylogeography. Using a range of different markers is always best though in order to fully resolve patterns which may not be shown by a single marker (Gardner *et al.* 2010). It can be impractical to employ a large range of markers due to cost and time, so a choice and compromise is generally made. The following is an overview of the genetic markers that have been used for population genetic studies.

Allozymes – Allozymes are allelic variants of enzymes/proteins. These are proteins that have differing electrical charges, and as a result will move through an electrically charged matrix at different rates. Running protein homogenates through a medium such as a gel in an electrical field separates them. Then by using catalytic activity of

the enzymes visible bands can be produced (Hellberg *et al.* 2002). Typically, allozymes are co-dominantly inherited, fitting with Mendelian inheritance. They are cheap to analyse but have significant drawbacks. Allozymes can only reveal a small portion of an organism's DNA substitutions, and only reveal mutations that have affected proteins. This means they may only show variation of genes which are under selection, leaving them prone to over-representing phenomena such as geographic clines or giving little phylogenetic information where non-synonymous mutations have not become fixed in a population (Hellberg *et al.* 2002).

Mitochondrial DNA (mtDNA) – Mitochondrial DNA is usually maternally inherited haploid DNA found in the mitochondria of all eukaryotes. Cases of heteroplasmy and double uniparental inheritance (DUI) have been observed in some molluscs though (Ghiselli *et al.* 2019). Sequencing mtDNA is a popular and informative method of phylogenetic and phylogeographic analysis. Amplifying short regions of the mitochondrial genome is done by PCR. This involves developing primers (short synthesised DNA strands) which bind to complementary regions of mtDNA. A primer binds to each end of the region chosen to amplify and the gene is copied. This occurs multiple times in a chain reaction producing a final product that has a very high concentration of the chosen marker. This can then be sequenced by methods such as Sanger sequencing. Primers are often interchangeable between species, depending how closely they are related. Once developed, primers can be used across a range of studies and species. This enables studies to produce comparable datasets, and reduces the development costs. Whole mitochondrial genomes can be sequenced as well. Methods including next-generation sequencing can yield resolution an order of magnitude higher than PCR and Sanger sequencing. While modern studies tend to be moving towards whole genome and mitochondrial sequencing, single markers are still relevant and informative. Maternal inheritance means mtDNA is haploid. It does not undergo recombination, making analysis simple, though rare cases have shown certain species such as some bivalves to have biparental inheritance of mtDNA (Hoeh *et al.* 1991). Maternal inheritance also means there is a smaller effective population size compared to loci characterised by Mendelian inheritance. Generally, mtDNA will have a high degree of variation making it well suited to intraspecific studies. Lack of recombination also means mtDNA can retain genetic signals for long periods of time through a lineage. Sequencing mtDNA is an effective, economical, and easily

analysed method of studying genetics. However, mtDNA can produce misleading phylogenies in some cases. Phenomena such a nuclear-mitochondrial DNA transfer may produce confusing signals, or the haploid nature may not show the full picture of how a population and the different sexes are structured. Mitochondrial DNA will often produce a different signal to nuclear DNA (nuDNA), leaving researchers with uncertainty. Many researchers have emphasised the need to combine mtDNA with other markers such as nuclear genes (Shaw 2002; Ballard & Whitlock 2004). DNA 'barcode' systems often rely upon a single mtDNA marker, commonly COI for animals (Hebert *et al.* 2003a). The idea behind these is that a single gene can be used for a rigid and comparable database to taxonomically classify a range of species. Whilst mtDNA has proven useful in such applications, it should be noted that no single marker can be used to confidently assign species delimitation in such a way.

Nuclear DNA (nuDNA) – Nuclear DNA markers can be amplified in the same way that mtDNA markers can, through PCR. A major difference though is that nuDNA is bi-parentally inherited and undergoes recombination. This means that it has a much larger effective population size and can fit with classic evolutionary models such as Mendelian inheritance. Recombination can remove genetic signals from a population quickly by homogenising differentiation between populations. Analysis of nuclear markers is often more complex than mitochondrial markers, as mutations coming together from more than one lineage can confound genetic signals. Typically, nuDNA evolves slower than mtDNA, so it can be useful for looking at deeper lineages if the genetic signal has been retained. Recombination can break up sequences, and patterns of population history can be lost more easily compared to mtDNA. Combining nuDNA markers with mtDNA markers can provide a robust method of analysis. Various genes evolve at different rates, so looking at faster and slower evolving markers together can reveal patterns which may not be shown with a single marker. PCR products can be unreadable if the loci are diploid and heterozygous, adding to the complexity of choosing a nuDNA marker.

Microsatellites – Microsatellites are short tandem repeats (di-, tri- or tetra- nucleotide repeats arranged in tandem) (Wright & Bentzen 1995). Repeats can be any combination of nucleotides. Microsatellite markers are typically short (<100bp) and are flanked either side by unique DNA sequences. When the sequence of the flanking

regions is known, primers can be developed to isolate the microsatellite by PCR. Various alleles of the microsatellite can then be analysed. Development of primers for microsatellite techniques can be a technically challenging and expensive process, however once they are developed, they can be re-used and further study can be cheaply and easily reproduced. Microsatellite markers are typically co-dominantly inherited and are not under any selective pressure. These markers are excellent for revealing recent genetic signals as they tend to have a high degree of polymorphism (Gardner *et al.* 2010).

Whole genome sequencing – This is the process of sequencing the entire genome of an organism. This includes all chromosomal locations and the entire mitochondrial genome. Whole genome sequencing provides the highest level of genetic resolution possible but is significantly more expensive and complicated than other methods. For many phylogeographic studies the whole genome level of resolution is not necessary. Genome sequencing can be useful for testing hypotheses that are often proposed by earlier studies that used a lower-resolution marker. When the entire genome is sequenced in a population level study, mechanisms of evolution and selection can be more confidently explained.

The COI Barcode

The mitochondrial cytochrome oxidase subunit-I (COI) gene was chosen as the DNA sequence to analyse in this thesis research. The mitochondrial COI gene is widely applied across zoological phylogenetic and phylogeographic studies. Worldwide it has become a standardly applied marker for delimiting species and for species-level analysis (Hebert *et al.* 2003b). The COI region, with its relatively high rate of mutation has been consistently used to effectively delimit closely related species across all animal phyla except Cnidaria (Hebert *et al.* 2003b). Distinguishing taxonomy across millions of species is obviously an immense amount of work, so it is not surprising that scientists have searched for an effective and accurate but widely useable method of distinguishing species. Morphological classifications are essential

but given the possibility of cryptic species or species with high levels of morphological diversity, we cannot rely on morphological classifications alone.

The COI region has been adopted as the most widely used genetic marker for animal species delimitation for almost 20 years. The initial concept of a COI 'barcode' was proposed in 2003 (Hebert *et al.* 2003a). Hebert *et al.* (2003a) found the COI region correctly classified 100% of the 200 species surveyed in their initial study. They argued that widespread use of the marker would improve the database used to identify species correctly. Jumping forward nearly twenty years, we see that the COI barcode system has been widely adopted and used across a range of species. Today PCR amplification and sequencing of COI is economical and simple. Its use across a wide range of species has proven useful for taxonomically difficult groups. With easily searchable databases such as GenBank, researchers have been able to collate a database of millions of species which can be quickly discovered and compared with functions like Basic Local Alignment Search Tool (BLAST). In addition, COI has proven effective across a range of intra-specific studies. The use of a common marker maintains consistency and a way to build collective knowledge among studies across the globe whilst avoiding fragmentation of information by using different systems (Mallet & Willmott 2003).

The COI region is not without shortcomings. Conflicts between gene trees and species trees due to incomplete lineage sorting and introgression can obscure the true evolutionary history of species. There is no shortage of such examples where has COI produced unexpected phylogenetic relationships and produced results that are incongruent with other markers (Cong *et al.* 2017). It is often suggested that studies should use a second genetic marker to collaborate with COI findings, however in some cases, it is unclear how to resolve conflict between markers. Faster evolving markers and/or more conserved markers (such as nuclear rRNA regions) may reveal a more accurate result that fits with other lines of evidence and allow for a better understanding of the issues in a COI dataset.

Species Concepts

The definition of a species has been one of the most widely discussed matters of conjecture in natural sciences for centuries, and yet there is still no one definitive conclusion on the matter. Taxonomic issues are likely to present themselves when studying taxa that are under-represented in literature. This is especially true in a place such as New Zealand, where so many endemic and unique groups are found, but proper taxonomic documentation has not been conducted on scales comparable to the likes of Europe. Population level genetic studies of intertidal invertebrates are likely to reveal new cryptic species which were previously unnoticed. The debate then arises as to whether a divergent clade of specimens can be classified as a separate species. Genetic data can produce confusing signals. For example, it may be impossible to identify reciprocal monophyly where a clade from the in-group might appear just as genetically distant from the rest of the samples as the immediate outgroup species. Various confounding effects can cause such issues, notably recent and rapid divergence of lineages can lead to incomplete lineage sorting, an issue that is likely to arise in New Zealand as speciation is often thought to occur after recent glacial cycles and regular geomorphological change (Winkworth *et al.* 1999; Craw *et al.* 2016). Closely related species may also interbreed and hybridise, a core issue which relates back to how we define a species or sub-species. Species boundaries are poorly defined in many taxa (Hare 2001), and endemic New Zealand groups are certainly no exception to this. Species boundaries might be difficult to define during the process of speciation, and boundaries can be 'semipermeable' to gene flow even after speciation (Hare 2001). Discrepancy between different genetic markers can reveal just how influential such phenomena can be.

Considering what defines a species is therefore of primary importance when undertaking a phylogeographic study, as the underlying principles rely on phylogenies. This topic is often over-looked or covered with insufficient detail (Zachos 2016). There are many species concepts, such as the biological/isolation species concept or the ecological species concept, but many of these have criteria which are incompatible with each other (De Queiroz 2007).

The biological/isolation species concept is defined as "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (Mayr 1963). Essentially this defines a species as a population that is reproductively isolated from other populations. Immediate violations of this are well known. *Equus ferus* (horse) and *Equus asinus* (donkey) may be bred to create a mule but we do not define these two animals as the same species. Such issues can become even more confusing when the offspring is a viable individual and occurs naturally, such as the 'grolar bear' produced when *Ursus arctos* (brown bear) and *Ursus maritimus* (polar bear) breed (Yadav *et al.* 2019). Hybrids of distinct species can effectively make a species or population go extinct if the hybrids are successful and show little hybrid-breakdown. *Anas superciliosa* (grey duck) in New Zealand are nearly all now hybrids with *Anas platyrhynchos* (mallard duck) (Gillespie 1985; Sheppard 2017).

The ecological species concept defines a species as a lineage which "occupies an adaptive zone minimally different to that of any other lineage in its range and which evolves separately from all lineages outside its range" (Van Valen 1976). An "adaptive zone" (Van Valen 1971) is similar to an ecological niche; it is the environment a species inhabits inclusive of physical and biological factors but exclusive of the life-histories and behaviour of species that inhabit it. This allows populations that are geographically isolated but evolutionarily similar and inhabit the same adaptive zones to be classified as the same species. However, this definition places little importance on mechanisms like allopatric speciation. Whilst it allows us to think more sensibly about the definition of geographically isolated populations, it dismisses essential modes of speciation. Niche partitioning within species is also a reasonably common phenomenon. The ecological species concept avoids this contradiction by using "adaptive zones" in its definition, but this only stands to weaken its integrity. Cleary neither of these classic species definitions stand up to much criticism.

Further definition such as 'regularly interbreeding to produce viable offspring' may stand to make a more rigorous definition of a species, but then the definition becomes confounded by the likes of geographical barriers which can isolate populations of the same species (Sokal & Crovello 1970). Alternative approaches include for example

the cohesion species concept (Templeton 1989), which places emphasis on demographic and genetic exchangeability. The cohesion concept encourages us to think of species as populations that maintain genetic and demographic connectivity, rather than simply reproducing on infrequent occasions. The underlying principles of the cohesion concept involve identifying traits which cause a population to maintain cohesion (Templeton 1989). This ties the definition of a species closely to how a population and subpopulations are defined. It also encourages thinking about how a species should be defined on a case-by-case basis.

The natural world and evolution of life is wonderfully messy. Attempting to classify such dynamic systems across billions of life forms can feel chaotic. We should not expect rigorous definitions to fit every scenario, and we should not attempt to fit everything into perfect sets. When taxonomic problems are encountered, a fluid approach may be the best way of solving it. It is unrealistic to expect that phylogenies can always be lineage sorted with 100% confidence, but a well-informed argument should be developed to support the classification of species. This should involve considering many approaches and species concepts. Researchers should not be tempted then to choose the one that best fits their narrative. Nor should they approach a dataset or project with a favourite species definition, as this will impart heavy bias in their classification of species. A combination of theories should be considered and even combined to give the most likely explanation for a given scenario. Consideration for likely speciation mechanisms must be considered. The constant break-up of landmasses in New Zealand is likely to have caused allopatric speciation in many lineages. Factors like this must be given heavy consideration on a case-by-case basis. From there if a new species is suspected, official classification can begin by following guidelines from a relevant organisation. For animals, this is the International Code of Zoological Nomenclature (ICZN).

Study Aims

The research findings of this thesis are reported in Chapter 2. Direct developing benthic gastropods (*Haustrum scobina sensu lato*) were sampled from 31 locations

around much of New Zealand. A total of 654 samples were included in the analysis. This study used the COI mitochondrial DNA gene as the primary marker for phylogeographic analysis. A smaller dataset of 27 nuclear 28S rRNA sequences is included to support phylogenetic interpretation of the dataset. The intention was to test for genetic connectivity between populations around New Zealand and use the findings to identify common causes of genetic patterns between *Haustrum scobina sensu lato* and ecologically similar species, which have previously been studied on a population level (such as *Cominella maculosa* and *Cominella virgata*). Studies of the *Cominella* genus have proposed hypotheses of gene flow mechanisms including sporadic long-distance dispersal events (Fleming *et al.* 2017; Dohner *et al.* 2018), human-mediated translocations (Fleming *et al.* 2017; Walton *et al.* 2018) and post-glacial maxima colonisation of southern sites (Fleming *et al.* 2017; Walton *et al.* 2018). Specific hypotheses and research questions are listed in Chapter 2.

The implications of disputed taxonomic documentation are discussed. This thesis provides an example of how the findings of analysis from different genetic markers can conflict with each other. It also provides a description of the methods of working through taxonomic issues when using genetic data. The importance of proper phylogenetic interpretation before phylogeographic analysis is discussed. This research highlights issues commonly faced in phylogeographic studies, for species that have not been previously studied in significant detail.

Chapter 2 - The Phylogenetics and Phylogeography of *Haustrum scobina sensu lato*

Introduction

<u>Marine Phylogeography in New Zealand</u>

New Zealand's vast and complicated ocean environment makes a unique setting to study the movement and evolution of populations. New Zealand has the world's fifth largest Exclusive Economic Zone (EEZ) . Whilst this provides excellent opportunities to study marine processes, it also implies an inherent responsibility to understand, manage and conserve the ecosystem. Since Māori first arrived sometime in the thirteenth century A.D (Wilmshurstl & Higham 2004), people in New Zealand have had a strong relationship with the marine environment. This has directly influenced patterns of migration, breeding, and abundance of many marine species.

Geological processes have also had a profound influence on New Zealand species. Sitting on the convergence zone between the Pacific and Australasian plates means the physical landscape is constantly changing. Changes in eustatic sea level can be amplified or dampened by geological change. These processes can have a unique effect on New Zealand species relative to those in other parts of the world. This can provide opportunities to study how populations respond to fast rates of environmental change and infer how they might react to future changes such as human-induced climate change and sea level rise. It also means however, that New Zealand species should be studied with an extra degree of attention. With such a unique landscape, seascape, and vast array of endemic species, it is difficult to directly draw comparisons and infer results from studies in other parts of the world.

The most comprehensive review of New Zealand benthic phylogeography to date was conducted by Ross *et al.* (2009). They state in the 30 years prior to their review, 42 molecular studies had been conducted on the phylogeography of New Zealand's benthic coastal invertebrates and plants. Many biogeographic hypotheses have been

formed, particularly in New Zealand and ever increasing numbers of molecular studies have been testing them in recent years (Wallis & Trewick 2009). Since Ross *et al.* (2009) published their review of the field, New Zealand researchers have continued to expand the work. Notable recent publications that have focused on phylogeography of direct developing benthos in particular include Walton *et al.* (2018), Dohner *et al.* (2018), Fleming *et al.* (2017) and Keeney *et al.* (2013). These studies have focused on *Cominella maculosa*, *Cominella virgata*, *Zeacumantus subcarinatus* and *Zeacumantus lutulentus*, which are all direct developing gastropods.

The most common pattern observed in phylogeographic studies of New Zealand benthic invertebrates is a north/south divide around the top of the South Island (Ross *et al.* 2009). This has commonly been attributed to upwelling in the Cook Strait, a process that is likely to prevent pelagic larvae transportation. However, most claims are unsubstantiated, for example there has been no direct evidence to date that upwelling in the Cook Strait region has restricted gene flow. Only eight of the 42 papers reviewed by Ross *et al.* (2009) observed a panmictic pattern with no genetic subdivision, this comprised seven taxa which all had a pelagic larval stage. Ross *et al.* (2009) found a significant negative correlation between PLD and genetic differentiation across the studies they reviewed. Species with a PLD of less than ten days showed greater variability in genetic variation than those with a longer PLD. This suggests that when PLD is short, other environmental factors have greater influence on a species' distribution. Identifying common patterns across various species can help test hypotheses about what is determining migration and barriers (Teske *et al.* 2011).

Studying direct developers is therefore more likely to be informative about environmental processes effecting a species' distribution and evolution than studying species with a longer PLD. A common pattern seen in New Zealand direct developers is a lack diversity in southern populations (Fleming *et al.* 2017; Walton *et al.* 2018). This is often attributed to recent post-glacial re-colonisation of southern sites (Fleming *et al.* 2017; Dohner *et al.* 2018; Walton *et al.* 2018). Keeney *et al.* (2013) also observed a connection between South Island and south-western North Island populations with evidence for recent population expansion. They attribute this to connection via a land bridge between Farewell Spit and south Taranaki during the

LGM ~18 kya (Lewis *et al.* 1994; Trewick & Bland 2012). These studies each focus on a single genus; *Cominella* and *Zeacumantus*. The conclusions of such studies are therefore still hypotheses to be tested with ecologically similar species. This study aimed to test for genetic differentiation around New Zealand in two species of the *Haustrum* genus, which live on similar mid-low tide rocky reefs as *Cominella* spp.

Target Species – *Haustrum scobina sensu lato.* (Quoy & Gaimard 1833)

The subfamily Haustrinae is part of the Muricidae family (in the clade Neogastropoda), which is one of the most diverse marine gastropod families comprised of at least 1,600 known extant species (Barco *et al.* 2015). Within the Haustrinae subfamily there are two extant genera, *Bedeva* and *Haustrum* (Barco *et al.* 2015). Until recently, *Bedeva vinosum* was thought to belong to the *Haustrum* genus (Ayre *et al.* 2009), but phylogenetic analysis has since placed it in the *Bedeva* genus (Barco *et al.* 2010). The *Bedeva* genus is endemic to Australia, though recent translocation events have led to their establishment in South Africa, the Canary Islands and Madiera, whilst the *Haustrum* genus is endemic to New Zealand (Logan 2019). Muricidae are a family of predatory whelks and harbour adaptations allowing them to drill through shells (Carriker 1981), earning many of them the common name of 'oyster borers'. Muricids feed on a variety of benthic invertebrates (Kitching & Lockwood 1974) and fill niches as ectoparasites (Vaïtilingon *et al.* 2004), scavengers (Morton 2006) and predators (Blackmore 2000). Often found in high abundance, Muricid species significantly affect benthic community structure (Morton 1999). Members of the family are often considered significant pests of the aquaculture industry, having a detrimental impact on farmed species such as oysters (Buhle & Ruesink 2009). There have been significant uncertainties regarding the phylogeny of the Muricidae family, and though recent phylogenetic studies have attempted to clarify some of this, there are aspects left to be debated (Tan 2003; Barco *et al.* 2010; Barco *et al.* 2015).

The *Haustrum* genus is comprised of four extant species, *Haustrum scobina*, *Haustrum albomarginatum*, *Haustrum haustorium* and *Haustrum lacunosum*, all of which are endemic to New Zealand (Barco *et al.* 2015). The classification of two of

these, *Haustrum scobina* and *Haustrum albomarginatum,* has been subjected to debate for decades. Originally, they were thought to be a single species (*Lepsiella scobina*), but *Haustrum albomarginatum* is now considered its own species by many. Many Muricidae phylogenies and identifications have been contentious due to morphological convergence and plasticity (Zou *et al.* 2012). These two species are morphologically very similar making them difficult to distinguish from each other, particularly when shell features have been eroded (Tan 2003; Logan 2019). Whilst several morphological studies classified these two as separate species (Kitching & Lockwood 1974; Smith & McVeagh 1991), various subsequent studies have neglected to adopt this classification as two species, opting instead to refer to *Haustrum albomarginatum* as a subspecies (Tan 2003).

The relatively recent accessibility of DNA sequencing has allowed for a more confident assessment of species classifications, producing high resolution datasets which can help researchers distinguish between morphologically similar or cryptic species. Genetic studies have claimed to affirm the distinction of *Haustrum scobina* and *Haustrum albomarginatum* as separate species (Barco *et al.* 2015; Logan 2019). In the most recent of these *Haustrum albomarginatum* samples were inadvertently included in a population level phylogeographic study of *Haustrum scobina* which used the mitochondrial cytochrome oxidase subunit I (COI) gene (Logan 2019). Two other studies have classified the two as separate species using analysis of mitochondrial markers, nuclear markers and allozyme electrophoresis (Smith & McVeagh 1991; Barco *et al.* 2015).

Barco *et al.* (2010) developed a molecular phylogenetic framework for the Muricidae family using the mitochondrial 12S rRNA, 16S rRNA, COI and nuclear 18S rRNA markers. This study only included one sample each of *Haustrum haustorium*, *Haustrum lacunosum* and *Haustrum scobina* (using 12S, 16S and 18S markers) but no samples of *Haustrum albomarginatum*. Barco *et al.* (2015) included eight samples each of *Haustrum scobina* and *Haustrum albomarginatum* in their phylogenetic study of the Pagodulinae and Haustrinae sub-families. This study recovered the *Haustrum scobina* and *Haustrum albomarginatum* samples as species-level clades. These samples were all collected from the same locality (Hawera, Taranaki) which gives support to the recovered phylogeny. Barco *et al.* (2015) do not cover the potential

diversity of the genus though, which is widely abundant across the entire country (Morton & Miller 1973). Logan (2019) was the first to conduct a population-level analysis of *Haustrum scobina*. That study found that *Haustrum albomarginatum* was a species separate from *Haustrum scobina* based on a COI dataset, but it only includes 16 samples of *Haustrum albomarginatum* in a study with 376 *Haustrum scobina* samples.

Logan (2019) noted that one clade recovered from their *Haustrum scobina* dataset was exceptionally divergent from the rest of the sampled specimens (both *Haustrum scobina* and *Haustrum albomarginatum*) but fails to explain why this clade was present. This unexpected diversity could be a result of a processing error, marker and/or sample size issues or the presence of a cryptic species. The phylogeny of *Haustrum scobina* and *Haustrum albomarginatum* remains to be resolved, and as one of the most abundant species or group of species on New Zealand's rocky shores (Morton & Miller 1973), it should be an important issue. A more comprehensive nation-wide study representing both *Haustrum scobina* and *Haustrum albomarginatum* is necessary to do this.

Haustrum scobina lay eggs in irregular clutches, cemented to rock which hatch after about 10 weeks (Carrasco & Phillips 2014). Embryos are direct developing and crawl after hatching without any planktonic larval stage. Carrasco and Phillips (2014) observed a mean of 235 ± 17 (SE) propagules per egg capsule. Only 10 ± 1 (SE) juveniles hatched from each capsule, the remaining embryos are used as nurse embryos during development.

Figure 2.1 - Intra-capsular development of *Haustrum scobina*. Photographs were taken at weekly intervals. A, Four hours after collection. B–J, Weeks 2 to 10. Scale bars: A–C = 0.25 mm; D–J = 0.5 mm. Abbreviations: b, blastomere; ct, cephalic tentacle; cv, cephalic vesicle; e, eye; f, foot; h, head; lk, larval kidneys; mo, morula; ne, nurse embryo; o, operculum; pc, prismatic cells; s, shell; sp, siphon; v, velum. Figure taken from (Carrasco & Phillips 2014).

Without proper taxonomic assignment, it can be difficult to deduce life-history patterns and behaviour of a species from literature alone. This is evident when attempting to find reliable research describing *Haustrum albomarginatum*, as most studies to date refer to the species as *Haustrum scobina* whether *Haustrum albomarginatum* was included in the research or not. For the purpose of this thesis, current knowledge of *Haustrum scobina* was assumed to be roughly representative of *Haustrum albomarginatum*. Any notable difference in life-history patterns between them would have most likely lead to classification of the two as separate species from the start, so it is not entirely unreasonable to assume they share very similar life-history patterns. Anecdotal evidence suggests however that niche partitioning does occur between the two proposed species, vertically through the tidal zone and across features such as streams (K. Walton, B. A. Marshall pers. comm. 2019). Luckens (1975) also note that copulation of *Haustrum albomarginatum* occurred earlier than

for *Haustrum scobina*, and occurred again in spring only for *Haustrum albomarginatum*, although only at a single site in Christchurch.

Haustrum albomarginatum are commonly identified by their smooth shells, whereas *Haustrum scobina* tend to have pronounced axial costae around the shell. The spire angle of *Haustrum scobina* is typically narrower and the aperture is generally narrower, and more ovoid than the aperture of *Haustrum albomarginatum*. They attain comparable size at maturity (~35 mm height and ~20 mm width). Erosion of shell features makes identification of species by shell morphology alone unreliable, however. Genetic techniques can be particularly useful in assigning species in morphologically similar groups.

Genetic analysis can be a powerful way of distinguishing species, but these methods are not without shortcomings. Several problems can arise in phylogenetic and phylogeographic studies producing false signals about the evolutionary patterns. In some cases invertebrate mitochondrial DNA can even be indirectly selected for due to linkage disequilibrium with inherited micro-organisms (Hurst & Jiggins 2005). It may be tempting to classify highly divergent clades as cryptic species, particularly when considering the historic contention around a morphologically uniform and understudied group. It is important to note that such results may be caused by phenomena like incomplete lineage sorting in a marker. Collaboration nuclear DNA markers with mitochondrial DNA markers should be used to identify issues like these. One aim of this study was to clarify the taxonomy of *Haustrum scobina* and *Haustrum albomarginatum*. This was in addition to the overall phylogeographic goal of the project, which was to test for population structure and gene flow patterns in *Haustrum scobina* and *Haustrum albomarginatum* on a large scale around the New Zealand coastline. The results of such a study can be compared to previously studied and ecologically similar species and used to identify common environmental processes which significantly affect more than just one species (Teske *et al.* 2011).

With the taxonomic uncertainty in consideration, the samples in this study are referred to as *Haustrum scobina sensu lato*, meaning *Haustrum scobina* 'in the loose sense'. This terminology is used to avoid classifying samples to species-level before resolving taxonomy, and therefore avoid bias in taxonomic analysis. The immediate

outgroup to *Haustrum scobina sensu lato* is assumed to be *Haustrum haustorium*, with *Haustrum lacunosum* being the next closest species (Barco *et al.* 2010; Barco *et al.* 2015). *Haustrum scobina sensu lato* is found in high abundance on rocky shores in the low-tide to mid-littoral zones, where abundance only decreases in areas of excessively strong surf (Morton & Miller 1973; Logan 2019). The target species is not of known cultural significance, nor is it commercially important or of conservation concern. Studying abundant species may prove useful for inferring patterns in similar species which are more difficult to study due to cultural and conservation concerns or those which are difficult to sample (Walton 2017).

Marker Choice

A 610 bp portion of the mitochondrial COI gene was the primary marker for this study. COI was chosen to be consistent with previous New Zealand studies (Fleming *et al.* 2017; Dohner *et al.* 2018; Walton *et al.* 2018; Logan 2019), as well as other global studies. An 827 bp portion of the nuclear large sub-unit 28s rRNA locus was also used to support phylogenetic interpretation of the primary COI dataset. Large sub-unit 28S is a nuclear gene that undergoes recombination, so it has a larger effective population size and is less prone than COI to being affected by rapid divergence in a lineage, thus reducing the likelihood of incomplete lineage sorting. Nuclear genes are typically slower evolving than mitochondrial genes but tend to show a more conserved part of the phylogeny. Relative to other nuclear markers, 28S is considered 'fast evolving' and suitable for distinguishing intra-specific or close-relative phylogenies. The 28S marker has also been used in studies of the *Haustrum* genus, making direct comparisons to this study possible (Barco *et al.* 2010).

Phylogeographic studies are often limited in their approach to sampling efforts. Many studies take an *ad-hoc* approach to sampling, picking up samples where they can. This nearly always leads to *post-hoc* speculation about the findings and inconclusive results. Whilst there is nothing intentionally misleading with this approach, it can be an inefficient way of testing for population structure. Finding any results of significance can be hit and miss. Targeted studies may be able to identify a phylogeographic break but this requires many samples over a small area, so they are not often conducted. Typically breaks are found over gradients of faunal composition change and cannot be established with significant confidence (Dell 1962; Beu 2012; Walton 2017). Where a distinct break occurs, it is generally in the presence of a significant and obvious physical barrier such as the Cook Strait. Even in such cases, generalising break localities to large areas like the Cook Straight is somewhat uninformative. For varying taxa, these localities differ and are constantly changing. If we already know such a barrier will affect the distribution of a species, then it may not be useful to conduct fine-scale studies over these areas.

This thesis research was conducted with a broad nation-wide approach to sampling efforts, which can lead to better spent sampling efforts in subsequent investigations. The intention was to identify key areas where diversity is high or lacking, and where population connections or breaks might occur. This approach allows future investigations to concentrate on more localised areas where breaks, connections and diversity cannot easily be explained by common-sense thinking and literature research.

Study Aims

Building reliable seascape models of genetic flow is hardly a small undertaking, particularly in such varied environments as New Zealand. To date, a respectable archive of marine phylogeographic research has been reported, though as with most

modern fields of study there is still much more to be addressed. As the number and range of such studies increases, many questions are answered, and previous observations may be explained. However, with each new study comes further questions and recommendation of additional research to explain both surprising and expected results. Direct developing molluscs make ideal candidates for phylogeographic investigation, and are well represented in the field (Walton 2017). A range of studies in New Zealand have focused on direct developing molluscs, which were often ignored given the traditional paradigm that such species have little potential for dispersal and will not reveal any interesting processes or patterns.

Previous studies of direct developing gastropods in New Zealand suggest human-mediated transport may have influenced the genetic structure of such populations (Fleming *et al.* 2017). Common haplotypes shared between the Hauraki Gulf and Nelson have been found in *Cominella virgata* (Fleming *et al.* 2017) and evidence of long distance dispersal has been found in *Zeacumantus subcarinatus* (Keeney *et al.* 2013). It is likely such dispersal events might be facilitated naturally, via rafting on material such as bull-kelp (Bussolini & Waters 2015), floating as juveniles or egg masses (Adachi & Wada 1999; Dohner 2016) or even when specimens are carried by migratory birds (Miura *et al.* 2011). However drifting on kelp is unlikely to be a common occurrence for carnivorous species which are mostly associated with hard substrates (Donald *et al.* 2005) and laboratory experiments have shown long-distance floating of similar species to *Haustrum scobina sensu lato* (*Cominella maculosa*) to be unlikely (Dohner 2016). Dohner (2016) found that hatchlings can be dislodged and held in the water column but tend to sink in vertically circulating water columns. Local water currents may influence fine-scale dispersal, but hatchlings are unlikely to be carried great distances. In the context of a nation-wide phylogeographic study, short-distance dispersal has little influence. The shared haplotype found in *Cominella virgata* between the Hauraki Gulf and Nelson indicates that shipping may have a role to play in long-distance dispersal events of direct developing benthic invertebrates (Fleming *et al.* 2017). Species may be translocated as foul on ships or even attached to rocks (*Haustrum scobina sensu lato* egg capsules are cemented to rocks), which were used as ballast in shipping as late as the 20th century in New Zealand (Moore & Kenny 1986; Hewitt *et al.* 2009). Given the likelihood of shipping as a vector for species such as *Haustrum scobina*

sensu lato, sampling efforts were made at several ports significant distances from each other around New Zealand. The over-arching aim of this thesis was to build upon these studies and identify common patterns between them.

The main research questions of this thesis were as follows:

1 – Is there evidence for more than one species existing in *Haustrum scobina sensu lato*?

The null hypothesis is that there are two species in *Haustrum scobina sensu lato*, these are *Haustrum scobina* and *Haustrum albomarginatum*. This is consistent with current research (Barco *et al.* 2015).

There are two alternative hypotheses. Hypothesis one states *Haustrum scobina sensu lato* is a single species. Hypothesis two states *Haustrum scobina sensu lato* consists of more than two species. Hypothesis two was developed due to the unexplained lineage of the divergent haplotype found by Logan (2019).

2 – How has the *Haustrum scobina sensu lato* lineage evolved and dispersed over time?

3 – Are there any common patterns between the phylogeography of *Haustrum scobina sensu lato* and the phylogeography of ecologically similar New Zealand species (eg. *Cominella maculosa* and *Cominella virgata*)?

Methods & Materials

Sample Collection

Haustrum scobina sensu lato specimens were collected from fourteen locations around the New Zealand coastline by Kerry Walton, Meredith Walton (then Buchholz), Peter Ritchie, Claire Elliot, Nina Eastgate, George Rzoska and I between December 2018 and January 2020. Samples from six different locations were

borrowed from the Museum of New Zealand Te Papa Tongarewa (Wellington), as well as outgroup samples (*Haustrum haustorium* and *Haustrum lacunosum*) from five different locations. A table of these samples, their museum registration number, and collector can be found in the appendix (Supplementary Table 1). The unpublished COI dataset produced by Logan (2019) and 16 COI sequences from the published dataset by Barco *et al.* (2015) were also included in the analysis. A list of the sequences from published studies accessed via Genbank, and outgroup sequences is available in the results (Table 2.4). Where possible, specimens morphologically representing either *Haustrum scobina* or *Haustrum albomarginatum* were collected to get a proportionate representation of each putative species, though this is not a 100% reliable method of telling the two apart due to their morphologically similar nature. In locations where both *Haustrum scobina* and *Haustrum albomarginatum* (as morphologically defined) were present, both were collected. At many locations, only one morphotype was found. Initial analysis after field collection ignored species-level labelling of samples. Given the uncertainty of species classification in this genus any labelling of samples as either *Haustrum scobina* or *Haustrum albomarginatum* may impart bias in taxonomic analysis and classification. The collective group of samples is referred to as *Haustrum scobina sensu lato*.

Samples were preserved in >99 % ethanol. The soft body of each specimen was separated from the shell and stored in an individually marked 2 mL o-ring tube filled with >99 % EtOH and associated with the corresponding shell voucher. In some cases, the body had retreated into the upper whorls of the shell making it impossible to keep the shell intact. Crushed shells were discarded but recognisable pieces of shell were kept where possible. Specimens and their associated vouchers are currently stored at -20ºC in the Victoria University of Wellington molecular ecology lab but will shortly be assigned a museum registration number and transferred to the Museum of New Zealand Te Papa Tongarewa (Wellington) Natural History Collection.

<u>Extraction, PCR and Sequencing</u>

DNA extraction was performed using the rapid salt-extraction method (Aljanabi & Martinez 1997). Approximately 4 mm^3 of foot muscle tissue was clipped from behind the operculum of each sample. Ethanol was pressed out of the tissue before it was finely cut and placed in a 1.5 ml Eppendorf tube. Four hundred and eighty microlitres of DNA extraction buffer (0.4 M NaCl, 10 mM Tris-HCL pH 8.0 and 2 mM EDTA pH 8.0) was added and tubes were heated to 80°C for 5 minutes to de-activate enzymes. Samples were then cooled on ice before adding 10 µl of 10 mg/µl Proteinase K. They were then incubated while rotating at 37°C overnight. Digested samples were centrifuged at 16,000 rpm for 5 minutes before transferring the supernatant to a new 1.5 mL Eppendorf tube. Proteins were precipitated by adding 320 µL of 5 M NaCl solution to each tube before inverting 60 times. Tubes were then spun at maximum speed for 5 minutes, before the supernatant was transferred to a new tube. Wide bore pipette tips were used during the transfer of supernatants to avoid damaging DNA. DNA was precipitated by adding 525 µL of chilled isopropanol to each tube and inverting several times, before storing tubes at -20°C for at least one hour. Tubes were then centrifuged at 13,000 rpm and 4°C for 20 minutes. The isopropanol was removed and 1 mL of chilled 70% ethanol was added to each tube and inverted several times to wash excess salts from the DNA pellet. Tubes were again spun at 13,000 rpm and 4°C for 10 minutes before the ethanol was removed. Tubes were left in a heat block at 30°C for 25-30 minutes to evaporate any remaining ethanol. Forty microlitres of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) was added to each tube and left for one hour at 4°C to re-hydrate DNA. Samples were then stored at 4°C if PCR was performed soon after, or -20°C for storage longer than a day. DNA quantity and quality were checked using a NanoPhotometer$^®$ NP80 (Implen). An aliquot of each sample was taken and diluted to 25 ng/µL using molecular grade water for PCR.

A 710 bp portion of the mitochondrial COI gene and 1000 bp portion of the nuclear 28s gene were amplified by polymerase chain reaction (PCR). Universal Folmer primers were used for amplification of COI, these were LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-

TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer *et al.* 1994). Reactions were 25 µl in total, consisting of 67 mM Tris-HCl, 16 mM $(NH_4)_2SO_4$, 0.1% stabilizer, 3 mM MgCl, 0.6 mg/mL BSA, 0.1 µM each of forward and reverse primer, 0.4 mM dNTPs, 0.05 U/µL of Taq polymerase and 25 ng of template DNA, mixed in a 200 µL PCR tube. Tubes were placed in a thermocycler for the PCR. The initial denaturing step was at 95°C for three minutes, then 40 cycles of 95°C for 35 s, 50°C for 35 s, 72°C for 45 s were performed before a final extension phase of 72°C for 10 minutes.

A portion nuclear large sub-unit 28s rRNA was amplified using primers LSU5' (5'-TAG GTC GAC CCG CTG AAY TTA A-3') (Littlewood *et al.* 2000) and LSU1600R (5-'AGC GCC ATC CAT TTT CAG G-3') (Williams *et al.* 2003). Aliquots of DNA samples were diluted to 25 ng/µL and 1 µL of each sample was added to each reaction. Reactions were 25 µL in total, consisting of 67 mM Tris-HCl, 16 mM $(NH_4)_2SO_4$, 0.1 % stabilizer, 3 mM MgCl, 0.6 mg/ml BSA, 0.1 µM each of forward and reverse primer, 0.4 mM dNTPs, 0.05 U/µL of Taq polymerase and 25 ng of template DNA, mixed in a 200 µl PCR tube. Tubes were placed in a thermocycler, the initial denaturing step was at 95°C for three minutes, then 35 cycles of 95°C for 30 s, 52°C for 60 s, 72°C for 60 s were performed before a final extension phase of 72°C for five minutes.

Every PCR run included a control mix, in which the template DNA was replaced with molecular grade water. Controls were run on a 1.5% agarose gel to check for contamination of PCR reagents. PCR product quality and length was also checked by gel electrophoresis on 1.5% agarose. Gels were run at 400 mA and 80 V for 30 minutes. Each lane contained 2 µl of PCR product and 2 µl of Thermo Scientific 6X DNA Loading Dye. Each row of lanes contained one lane with 2 µl of Bioline Easyladder I (100 – 2000 bp) to directly compare the length of the amplicon.

Statistical Analysis

Sequence alignment

Raw sequences were imported into Geneious Prime version 2019.2.1 (Kearse *et al.* 2012). Chromatograms were inspected for each sequence and those deemed poor quality were discarded. COI sequences were trimmed to 610 bp (consistent with comparable studies) and 28S sequences were trimmed to 827 bp (the longest region of good quality in all samples used). Trimmed sequences were aligned using MUSCLE alignment (Edgar 2004) with default settings in Geneious Prime. The alignment files for each marker were then imported into MEGA X (Kumar *et al.* 2018) where the alignments were visual inspected and any insertions or deletions were removed and corrected.

Haplotype analysis and general summary statistics

Haplotype data files were produced using DnaSP v6 (Rozas *et al.* 2017). MEGA X (Kumar *et al.* 2018) was used to determine the most suitable nucleotide substitution model for haplotype data files, this was the HKY model (Hasegawa *et al.* 1985) with gamma-distributed rates for both datasets. Summary statistics including general haplotype and nucleotide diversity indices, Tajima's D test (Tajima 1989), Fu's Fs test (Fu 1997) Harpending's raggedness index (Harpending 1994), SSD and Tau values were produced in Arlequin v3.5.2 (Excoffier *et al.* 2006).

Clustering of COI haplotypes and corroboration with the 28S dataset

K-means clustering by pairwise distance between COI haplotypes was implemented in R Studio version 3.6.2. This was done to corroborate COI mtDNA clusters and 28S rRNA clades. Cluster analysis was performed using the factoextra package (Kassambara & Mundt 2017) and ggplot (Wickham & Wickham 2007) in RStudio v1.2 (Allaire 2012). Voucher photos of some 28S samples were clear-cut using Adobe Photoshop v5.2.0.

Haplotype networks and maps

Median-joining haplotype networks (Bandelt *et al.* 1999) were produced using PopART (Population Analysis with Reticulate Trees) v1.7 (Leigh & Bryant 2015). Maps of haplotype distribution were also produced in PopART v1.7. Median-joining networks and maps were edited using the open source programme Inkscape v1.0.

Phylogenetic trees

Bayesian phylogenetic trees were produced using MRBAYES v3.2.7a (Huelsenbeck & Ronquist 2001). Bayesian analysis was run using the HKY model of nucleotide substitution with gamma distribution rates for both the COI and 28S sequences. Analysis of the COI haplotypes was run for 1,500,000 generations, after which standard deviation of split frequencies reach 0.0099. Analysis of the 28S haplotypes was run for 1,500,000 generations, after which the standard deviation of split frequencies was 0.0079. The first 375,000 generations (25% of total) were discarded and the potential scale reduction factor for all convergence diagnostics was 1.00 in runs for the COI and 28S haplotypes. Maximum likelihood trees for both the COI and 28S datasets were produced in MEGA X (Kumar *et al.* 2018). Both maximum likelihood trees were produced using the HKY model of nucleotide substitution and for 1000 bootstrap replications. Bayesian and bootstrap consensus trees were annotated in FigTree v1.4.3 (Rambaut 2017).

Tests for non-neutral evolution and contamination

To account for selection, non-synonymous sites were removed from the COI dataset. A dataset with only synonymous and conserved sites should theoretically represent diversity that is neutral to selection. DnaSP v6 was used to identify and remove non-synonymous mutations in the COI dataset when testing for non-neutral selection. BLAST searches were performed in Geneious Prime version 2019.2.1 (Kearse *et al.* 2012) to test for contamination. Codon-based Z-tests for positive selection and neutral evolution were conducted using MEGA X (Kumar *et al.* 2018), with the Nei-Gojobori p-distance method (Nei & Gojobori 1986) and 1000 bootstrap replications.

Results

This study added 277 new 610 bp COI sequences of *Haustrum scobina sensu lato* from 19 locations to the dataset produced by Logan (2019) and part of the dataset produced by Barco *et al*. (2015). This enabled the formation of a dataset of 654 COI sequences. These featured 58 haplotypes, but the putative species could not be identified by these haplotypes. The addition of 26 new 827 bp 28S rRNA sequences to the single 28s rRNA sequence produced by Barco *et al*. (2010) enabled the formation of a dataset of 27 28S rRNA sequences of *Haustrum scobina sensu lato*. These featured seven haplotypes, which grouped into two reciprocally monophyletic groups. The 28S rRNA dataset supported the conclusion that there are two species in *Haustrum scobina sensu lato*; these are most likely *Haustrum scobina* and *Haustrum albomarginatum*.

Summary Statistics and Sampling Site Information

Table 2.1 displays sampling site information, site codes, haplotypes found at each location. Table 2.2 summarises genetic diversity indices, neutrality indices and indices of population expansion. All information in these tables refers to the primary COI dataset only.

Table 2.1 - Site codes, species and locations. All statistics refer to the primary COI dataset. n=sample size, s=number of segreating sites, Nh=number of haplotypes, Np=number of private haplotypes. Sites which consist or partially consist of samples from Logan (2019) are in bold, Hawera includes 16 samples from Barco *et al.* (2015) and the museum registration number is shown for these as well as any site which includes samples from Te Papa.

Species	Site Code	Location	Te Papa Registration Number (If Applicable)	Latitude (Decimal Degrees)	Longitude (E) (Decimal Degrees)	n	s	Nh	Haplotypes ('*' Denotes Private)
H. scobina sensu lato	AKA	Akaroa	TBA	-43.815	172.953	23	3	1	1
H. scobina sensu lato	CAS	**Castlepoint**	TBA	-40.870	176.230	29	0	2	5, 6*
H. scobina sensu lato	CHI	Chatham Islands	M.315671	-43.945	176.563	3	1	2	7, 8*
H. scobina sensu lato	COR	Opito Bay	TBA	-36.722	175.811	17	1	5	9*, 10*, 11*, 12*, 13*
H. scobina sensu lato	CPS	**Cape Palliser**	TBA	-41.610	175.290	24	0	1	14
H. scobina sensu lato	FPS	**Flat Point**	TBA	-41.240	175.960	23	0	6	5, 15*, 16*, 17*, 18*, 19*
H. scobina sensu lato	HAW	Hawera	M.301488/M.301489/TBA	-39.606	174.240	46	8	5	1, 20*, 21, 31*, 32*

H. scobina sensu lato	KIS	**Kawau Island**	TBA	-36.400	174.850	47	0	9	22, 23, 24, 25, 26*, 27*, 28, 29*, 30*
H. scobina sensu lato	KPS	**Kau Point**	TBA	-41.290	174.830	31	0	5	14, 28, 33*, 34*, 35*
H. scobina sensu lato	MAP	**Mahia Peninsula**	M.316155/TBA	-39.088	177.956	33	0	3	25, 36*, 37*
H. scobina sensu lato	MBS	**Makara Bay**	TBA	-41.220	174.710	30	0	2	14, 38*
H. scobina sensu lato	MOR	Moeraki	TBA	-45.358	170.855	8	1	1	1
H. scobina sensu lato	MWB	Mangawhai Beach	TBA	-36.081	174.598	30	2	5	7, 23, 39, 40*, 41*
H. scobina sensu lato	NEL	Nelson	M.129326	-41.200	173.326	1	1	1	1
H. scobina sensu lato	NPL	New Plymouth	TBA	-39.057	174.056	27	4	5	1, 21, 42, 43*, 44*
H. scobina sensu lato	ONE	Onehunga Bay	M.318156	-41.091	174.854	3	1	3	1, 14, 21
H. scobina sensu lato	OTH	Otago Harbour	TBA	-45.775	170.714	8	1	1	1
H. scobina sensu lato	PAI	Paterson Inlet	TBA	-45.905	168.124	8	1	2	1, 45*
H. scobina sensu lato	PAS	**Port Ahuriri**	TBA	-39.480	176.900	31	0	3	5, 22, 46*
H. scobina sensu lato	PBS	**Pourerere Beach**	TBA	-40.110	176.870	32	0	1	5
H. scobina sensu lato	PRB	Purau Bay	TBA	-43.623	172.738	24	4	3	1, 19, 47*
H. scobina sensu lato	RBH	Robin Hood Bay	TBA	-41.359	174.072	9	1	3	1, 14, 48*
H. scobina sensu lato	SPI	Spirits Bay	TBA	-34.420	172.859	20	1	5	5, 49*, 50*, 51*, 52*
H. scobina sensu lato	SSB	Sunset Beach	M.306340/M.306339	-37.402	174.705	5	2	4	2*,3*,4*,53*
H. scobina sensu lato	STI	Bathing Beach	TBA	-46.893	168.131	5	1	1	1
H. scobina sensu lato	TAS	**Tatapouri Point**	TBA	-38.650	178.150	32	0	4	5, 25, 54*, 55*

H. scobina sensu lato	TBS	**Titahi Bay**	TBA	-42.200	174.830	23	0	5	14, 21, 22, 56*, 57*
H. scobina sensu lato	URU	Urenui	TBA	-38.988	174.389	23	4	2	1, 42
H. scobina sensu lato	WHB	Whangapoua Beach	M.316768	-36.702	175.610	2	1	1	25
H. scobina sensu lato	WHS	**Whangaehu Beach**	TBA	-40.400	176.630	27	0	2	5, 25
H. scobina sensu lato	WPB	Waipu Beach	TBA	-36.030	174.509	30	3	5	7, 23, 24, 39, 58*
H. scobina sensu lato	-	All sample locations	TBA	-	-	654	41	58	-

Table 2.2 - Neutrality and mismatch indices. All statistics refer to the primary COI dataset. Hd = Haplotype diversity, n = Nucleotide diversity, k= average number of nucleotide substitutions. Numbers in bold indicate a significant value at that site for the given statistic at the 95% confidence level. Sites with no haplotype diversity are uninformative for this analysis and marked with a dash.

Species	Location	H_d (sd)	π (sd)	k (average number of nucleotide substitutions)	Fu's Fs (p-value)	Tajima's D (*p*-value)	Harpending's Raggedness Statistics (*p*-value)	SSD (*p*-value)	Tau (95% C.I)
H. scobina sensu lato	Akaroa	-	-		-	-	-	-	-
H. scobina sensu lato	Castlepoint	0.0690 (0.0632)	0.00125 (0.002)	0.138 (0.208)	-0.365 (0.167)	**-1.509 (0.046)**	0.876 (0.840)	0.0026 (0.110)	2.330 (0.000 - 85.250)
H. scobina sensu lato	Chathams	0.667 (0.314)	0.0181 (0.017)	2.00 (1.512)	1.609 (0.719)	0.000 (1.000)	1.000 (1.000)	**0.284 (0.028)**	3.489 (0.000 - 3.488)
H. scobina sensu lato	Opito Bay	0.5074 (0.1403)	0.0052 (0.005)	0.574 (0.488)	**-2.826 (0.002)**	**-1.577 (0.031)**	0.148 (0.494)	**0.0146 (0.060)**	0.531 (0.095 - 3.231)
H. scobina sensu lato	Cape Palliser	-	-	-	-	-	-	-	-
H. scobina sensu lato	Flat Point	0.395 (0.128)	0.009 (0.007)	0.949 (0.674)	**-2.195 (0.035)**	**-2.191 (0.001)**	0.237 (0.822)	0.008 (0.570)	1.520 (0.000 - 4.049)
H. scobina sensu lato	Hawera	0.574 (0.036)	0.0133 (0.067)	14.572 (6.645)	20.528 (1.000)	3.007 (1.000)	0.546 (0.224)	**0.174 (0.034)**	28.951 (20.600 - 35.426)
H. scobina sensu lato	Kawau Island	0.7734 (0.0481)	0.179 (0.089)	19.678 (8.861)	16.568 (1.000)	1.576 (0.953)	0.0838 (0.432)	0.045 (0.382)	37.425 (4.385 - 45.846)
H. scobina sensu lato	Kau Point	0.3978 (0.1066)	0.0386 (0.0219)	4.241 (2.162)	4.609 (0.946)	**-1.874 (0.019)**	0.199 (0.800)	0.010 (0.526)	31.862 (0.000 - 171.531)
H. scobina sensu lato	Mahia Peninsula	0.2273 (0.0933)	0.003 (0.003)	0.337 (0.344)	-0.630 (0.192)	-0.629 (0.250)	0.442 (0.898)	0.001 (0.336)	1.359 (0.000 - 85.250)
H. scobina sensu lato	Makara Bay	0.129 (0.079)	0.001 (0.002)	0.129 (0.201)	-0.439 (0.150)	-0.764 (0.215)	0.568 (0.902)	0.0001 (0.118)	0.132 (0.000 - 0.246)
H. scobina sensu lato	Moeraki	-	-	-	-	-	-	-	-
H. scobina sensu lato	Mangawhai Beach	0.5977 (0.059)	0.035 (0.020)	3.804 (1.970)	3.922 (0.928)	**-2.435 (0.00)**	0.440 (0.126)	**0.123 (0.024)**	39.728 (0.479 - 30.422)

H. scobina sensu lato	Nelson	-	-	-	-	-	-	-
H. scobina sensu lato	New Plymouth	0.6097 (0.088)	0.106 (0.055)	11.622 (5.432)	11.920 (0.998)	0.249 (0.690)	0.193 (0.574)	0.067 (0.358)	27.850 (1.383 - 42.638)
H. scobina sensu lato	Onehunga Bay	1.00 (0.272)	0.197 (0.151)	21.667 (13.302)	1.947 (0.540)	0.000 (1.000)	0.667 (0.962)	**0.268 (0.044)**	30.049 (16.612 - 27.547)
H. scobina sensu lato	Otago Harbour	-	-	-	-	-	-	-
H. scobina sensu lato	Paterson Inlet	0.250 (0.180)	0.002 (0.003)	0.250 (0.311)	-0.182 (0.198)	-1.055 (0.221)	0.3125 (0.888)	**0.001 (0.016)**	0.309 (0.000 - 172.875)
H. scobina sensu lato	Port Ahuriri	0.4280 (0.0964)	0.024 (0.015)	2.632 (1.445)	5.309 (0.974)	0.520 (0.742)	0.310 (0.742)	0.047 (0.350)	9.404 (0.000 - 90.854)
H. scobina sensu lato	Pourerere Beach	-	-	-	-	-	-	-
H. scobina sensu lato	Purau Bay	0.3587 (0.1096)	0.083 (0.044)	9.123 (4.349)	13.743 (1.00)	0.245 (0.663)	0.589 (0.772)	0.077 (0.312)	32.553 (0.000 - 170.500)
H. scobina sensu lato	Robin Hood Bay	0.4167 (0.1907)	0.072 (0.042)	7.944 (4.080)	6.306 (0.991)	**-1.933 (0.002)**	0.493 (0.720)	0.070 (0.500)	6.363 (0.000 - 10.186)
H. scobina sensu lato	Spirits Bay	0.616 (0.1056)	0.067 (0.036)	7.332 (3.581)	6.067 (0.985)	-1.099 (0.134)	0.277 (0.574)	0.068 (0.402)	3.354 (0.880 - 6.145)
H. scobina sensu lato	Sunset Beach	0.069 (0.0632)	0.236 (0.147)	26.000 (13.816)	3.232 (0.875)	0.795 (0.794)	0.190 (0.850)	0.105 (0.726)	28.270 (20.941 - 39.600)
H. scobina sensu lato	Bathing Beach	-	-	-	-	-	-	-	-
H. scobina sensu lato	Tatapouri Point	0.333 (0.100)	0.018 (0.012)	2.030 (1.172)	2.605 (0.880)	-0.280 (0.398)	0.441 (0.784)	0.033 (0.434)	7.690 (0.000 - 85.250)
H. scobina sensu lato	Titahi Bay	0.751 (0.051)	0.044 (0.025)	4.814 (2.439)	4.289 (0.959)	1.285 (0.935)	0.164 (0.558)	0.040 (0.408)	7.188 (4.072 - 10.589)
H. scobina sensu lato	Urenui	0.498 (0.053)	0.005 (0.004)	0.498 (0.441)	1.470 (0.711)	1.433 (0.946)	0.248 (0.286)	**0.027 (0.002)**	0.515 (0.123 - 49.027)
H. scobina sensu lato	Whangapoua Beach	-	-	-	-	-	-	-	-
H. scobina sensu lato	Whangaehu Beach	0.143 (0.086)	0.009 (0.007)	0.997 (0.694)	3.250 (0.915)	-1.367 (0.078)	0.776 (0.760)	0.015 (0.216)	7.710 (0.000 - 85.891)

H. scobina sensu lato	Waipu Beach	0.671 (0.062)	0.186 (0.094)	20.483 (9.308)	20.267 (1.000)	3.612 (1.000)	0.157 (0.634)	0.090 (0.118)	40.049 (28.101 - 53.993)
H. scobina sensu lato	All sample locations	0.8937 (0.0058)	0.171 (0.0837)	16.355	3.903 (N/A)	-0.129 (0.617)	0.302 (0.496)	0.05127 (0.19)	12.25 (3.22 - 50.51)

COI Phylogeny
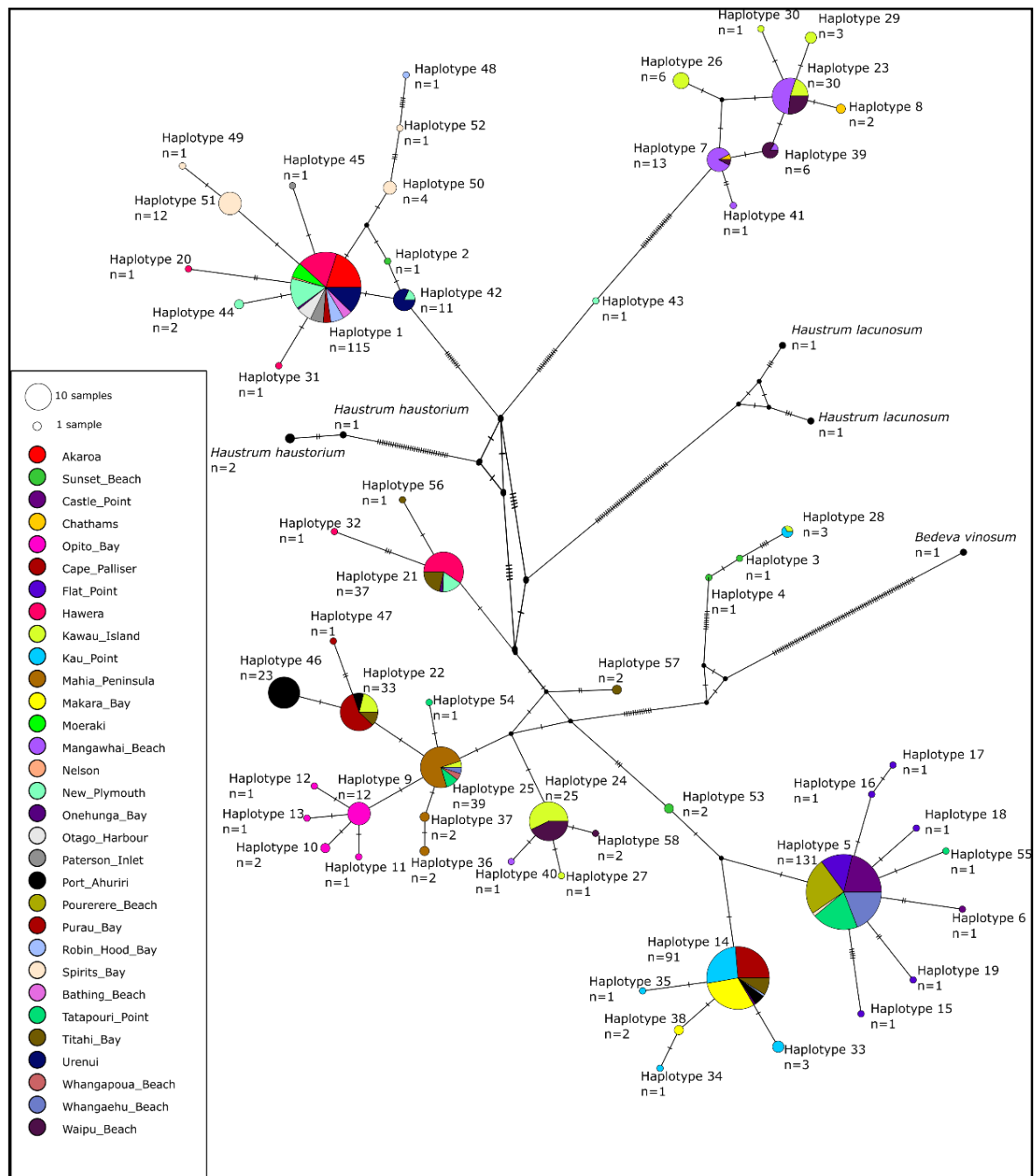
Figure 2.2 - Median-joining network of all *Haustrum scobina sensu lato* COI haplotypes. This figure represents the two putative in-group species. Outgroup species (*Haustrum haustorium*, *Haustrum lacunosum* and *Bedeva vinosum*) were obtained from GenBank where they were deposited by Barco *et al.* (2015) and Ayre *et al.* (2009); these are included to demonstrate expected levels of mutational distance between and within species. Hatch marks represent one mutational step. Note outgroup species are marked black, but this does not correspond to sample location.
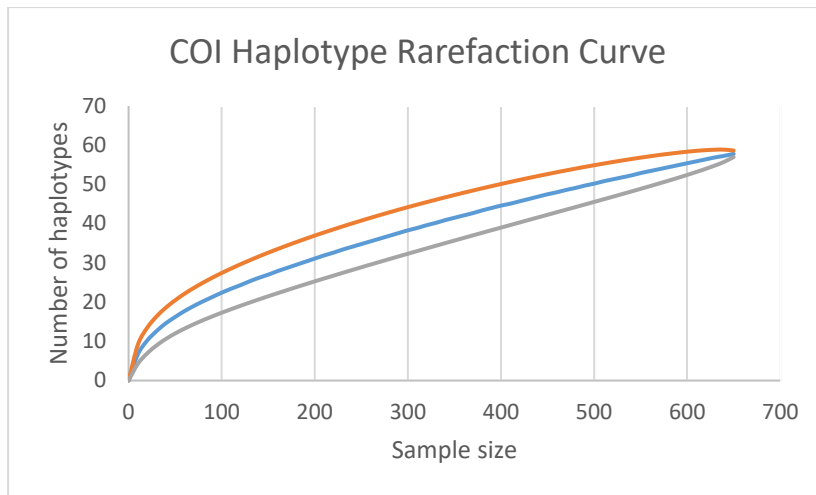
Figure 2.3 - COI haplotype rarefaction curve. The blue line indicates expected number of haplotypes for a given sample size and the orange and grey lines are at the upper and lower 95% confidence intervals respectively.
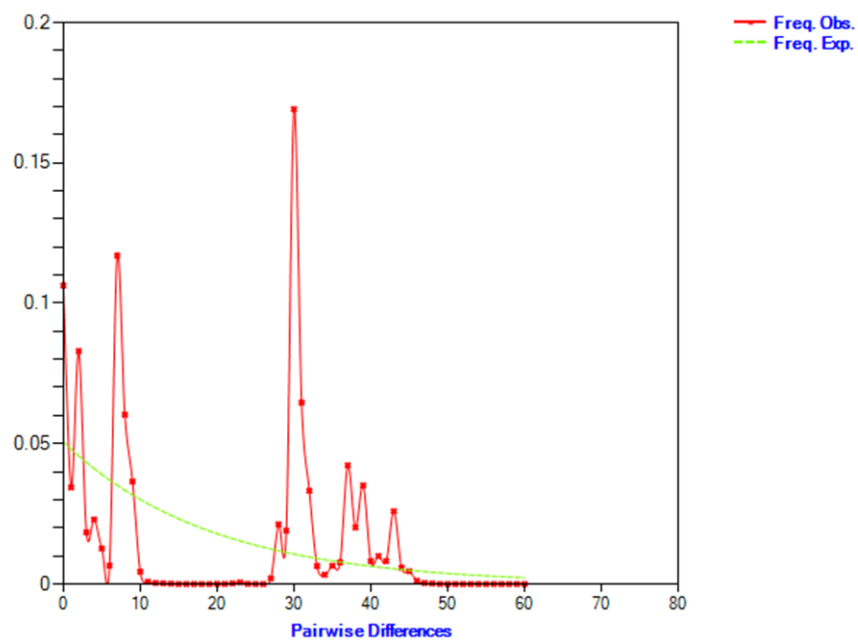
Figure 2.4 - Mismatch distribution of all COI haplotypes for all sites combined. The red line shows the observed haplotype frequencies and the green line shows expected distribution of haplotype frequencies.
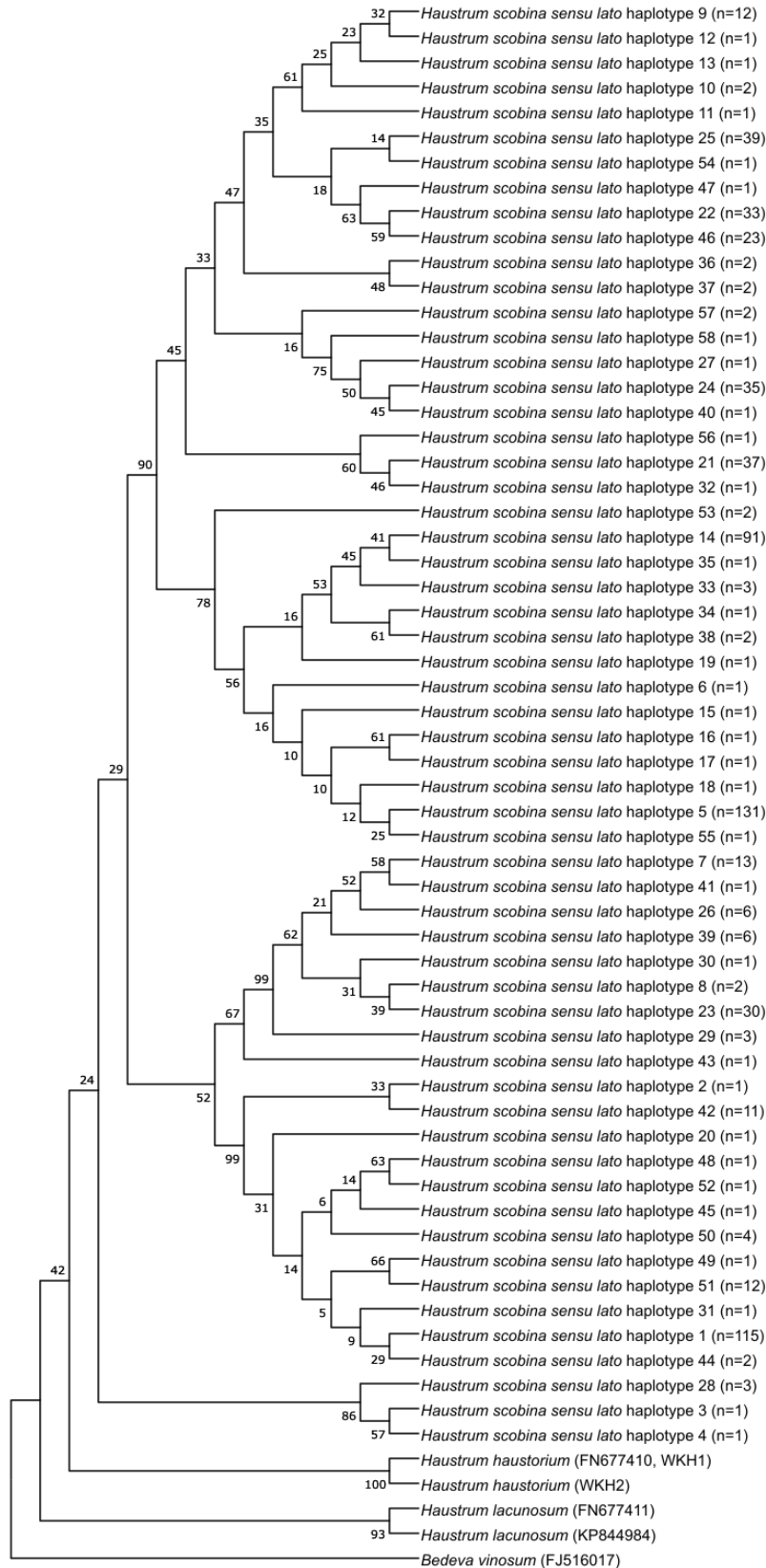
Figure 2.5 - Maximum likelihood consensus tree of all *Haustrum scobina sensu lato* COI haplotypes and outgroup species from Barco *et al.* (2015) and Ayre *et al.* (2009). Node values correspond to bootstrap support (percentage of replications) after 1000 replications.

The in-group formed four species-level clades in the COI haplotype median-joining network (Figure 2.2), when compared to the distance of outgroup species. Each clade also had a significant number of samples, except for haplotypes 3, 4 and 28. This suggested the distance between these clades was not due to sequencing or processing error. BLAST searches also confirmed that all clades belong to *Haustrum scobina sensu lato* and were not a result of contamination.

It was impossible to reliably determine how many species there are in *Haustrum scobina sensu lato* from this dataset alone. Nor was it possible to determine which clade represented which proposed species within *Haustrum scobina sensu lato*. Reciprocal monophyly of the in-group could not reliably be identified in the COI data. The maximum-likelihood tree (Figure 2.5) recovered the in-group as a monophyletic clade with low bootstrap support (24% of replications). A Bayesian tree for the COI dataset can be found in the appendix (Supplementary Figure 4). The Bayesian tree did not recover the in-group as a monophyletic clade with moderate support (posterior probability of 0.64).

This level of diversity in the COI dataset could indicate there are more than two species in *Haustrum scobina sensu lato*. The haplotype rarefaction curve (Figure 2.3) indicated significantly more haplotypes would be found with further sampling.

Neutrality Tests

Twenty-five non-synonymous sites were removed from all COI sequences. The resulting phylogeny was essentially the same as the full dataset, except that there were fewer rare haplotypes (see Supplementary Figure 1). It was still impossible to distinguish reciprocal monophyly that could correspond to expected species boundaries in the in-group. This suggested that selection acting on non-synonymous sites may not be confounding the result of the COI phylogeny.

A codon-based Z-test for neutrality returned a Z statistic of -5.69 and p-value of 0.00. The same test for positive selection returned a Z statistic of -5.63 and p-value of 1.00.

At the 95% confidence level this supported the conclusion that there was no selection acting directly on COI. It was not possible to rule out genetic hitchhiking, where the COI gene is directly linked to a gene under selection.

Nuclear 28S rRNA Dataset

The bi-modal histogram of mismatch distributions in the COI dataset (Figure 2.4) indicated there are only two species in the in-group, though these still could not be identified in the COI sequences. This unexpectedly diverse phylogeny might indicate there are issues such as incomplete lineage sorting, hybridisation or mitochondrial introgression in the COI gene or mitochondrial genome. To test for these issues, a dataset using an 827 bp portion of the nuclear large sub-unit 28S rRNA locus was produced. This smaller dataset consisted of representative samples from each major clade in the COI dataset.

The following figures summarise the phylogeny of the 28S dataset. This dataset was produced only to give phylogenetic support to the COI dataset, so indices of population expansion and diversity were not included.

Figure 2.6 - Haplotype network of all 28S rRNA sequences. Hatch marks represent one mutational step. Samples were chosen to represent each clade of the COI dataset. This haplotype network indicates there are two in-group species. Haplotypes 4, 5, 6, and 7 are most likely *Haustrum scobina sensu stricto*, and haplotypes 1, 2 and 3 are likely *Haustrum albomarginatum sensu stricto*. Sequences used for outgroup species are the same as those used and listed in the maximum likelihood tree (Figure 2.7).

Figure 2.7 - Maximum likelihood consensus tree of all 28S haplotypes. *Haustrum scobina* and *Haustrum albomarginatum* are labelled *sensu stricto*. Node values correspond to bootstrap support, where each number is the percentage of replications from 1000 that recovered the phylogeny. Outgroup sequences were obtained from GenBank where they were deposited by Barco *et al.* (2010) and Barco *et al.* (2015), GenBank accession numbers are shown beside them. Samples WAH1 and WDH2 were produced in this thesis research.

The 28S rRNA phylogeny showed a consistent grouping of the in-group samples to two distinct species (see Figure 2.6). These two species did not appear to be each other's closest relatives as was previously expected (Barco *et al*. 2015). *Haustrum haustorium* appeared to be as closely related to the in-group species as they were to each other. The maximum likelihood tree for the 28S dataset (Figure 2.7) recovered *Haustrum scobina sensu stricto* and *Haustrum albomarginatum sensu stricto* as a monophyletic clade with moderate support (58% of replications). A polytomy of the two putative in-group species and *Haustrum haustorium* in the Bayesian tree with moderate support (posterior probability of 0.53) (Supplementary Figure 4) however, means a relationship between the three species could not be reliably established. The 28S haplotype network (Figure 2.6) shows there were six mutational steps between haplotype 4 and *Haustrum haustorium*. There were eight mutational steps between haplotype 1 and each of *Haustrum haustorium* and haplotype 4, giving support to the phylogeny of the Bayesian tree.

Corroborating the 28S Dataset With COI Haplotype Clusters

The optimal number of clusters for the COI dataset was three (see Figure 2.8). Shell morphology suggested there are at least two species in *Haustrum scobina sensu lato* and it has been suggested that additional hitherto unrecognized species within the in-group were possible yet unlikely (K. Walton pers. comm, 2020). The 28S rRNA phylogeny indicated there are two species in the in-group (see Figure 2.6). This supports the null hypothesis in this study. COI clusters were therefore assigned with a limit of either two or three clusters.
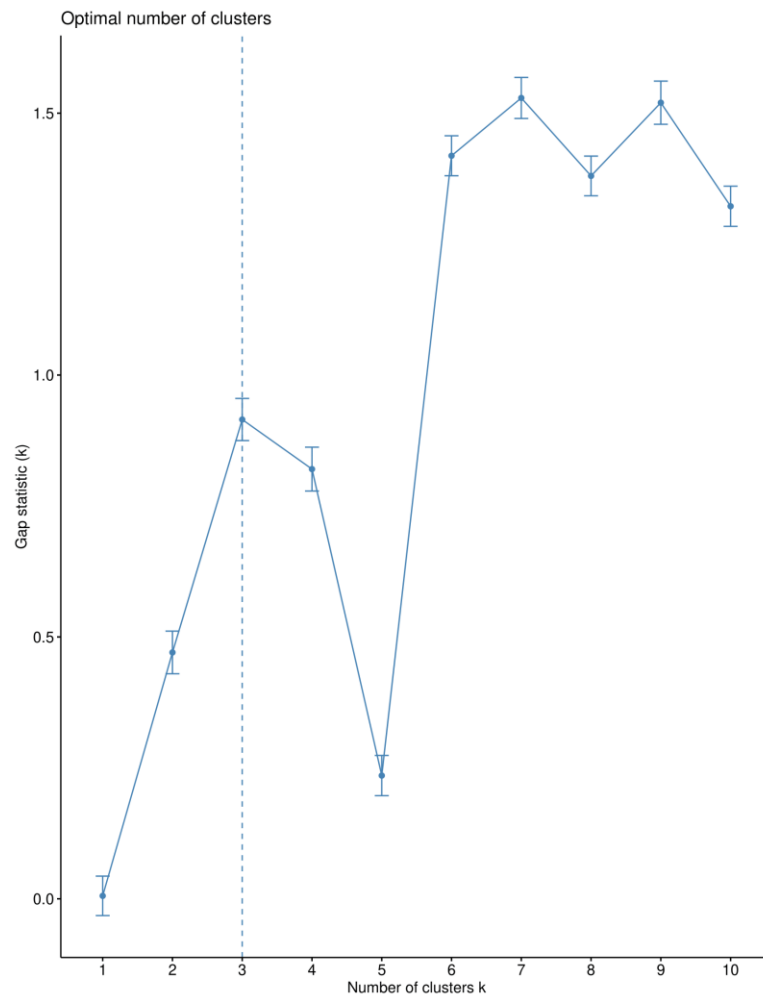
Figure 2.8 - Gap statistic (k) plot for COI haplotype clusters. Produced using the factoextra and ggplot packages in RStudio v3.
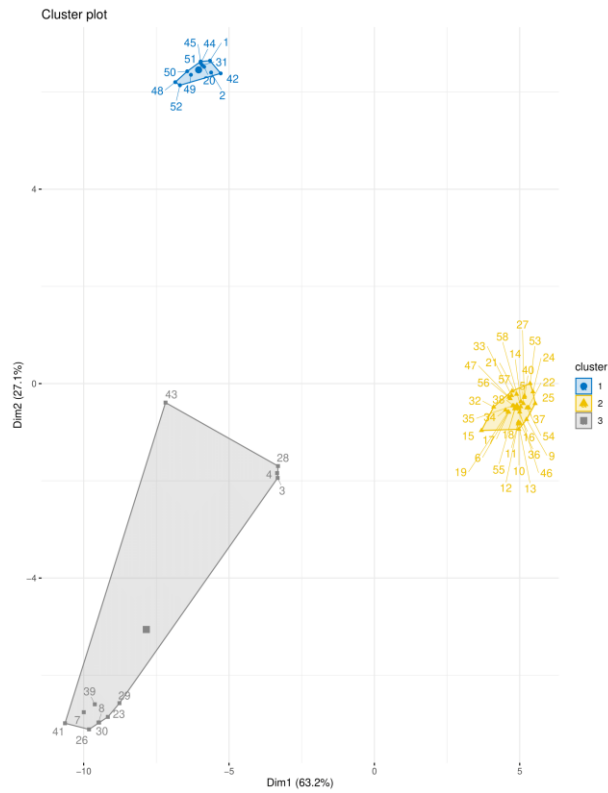
Figure 2.9 - COI haplotype cluster plot for three clusters. Produced using the factoextra and ggplot packages in RStudio v3.
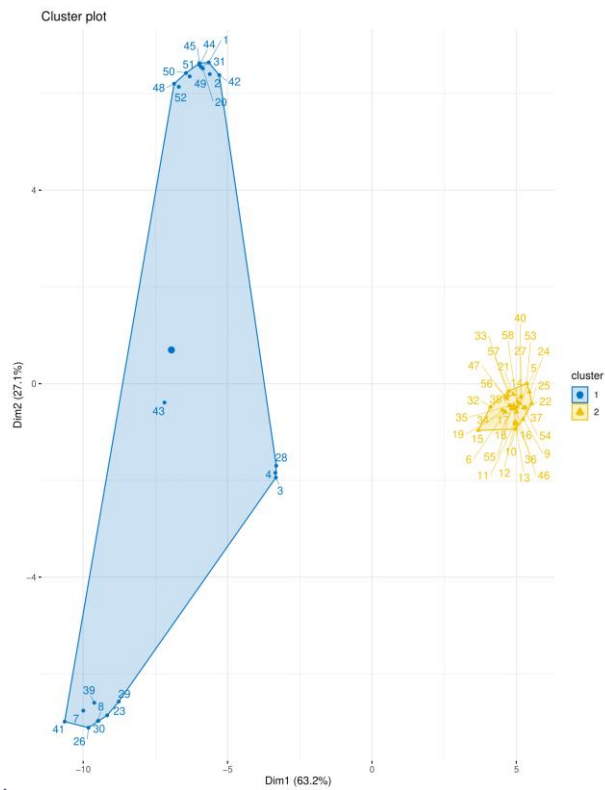
Figure 2.10 - COI haplotype cluster plot for two clusters. Produced using the factoextra and ggplot packages in RStudio v3.
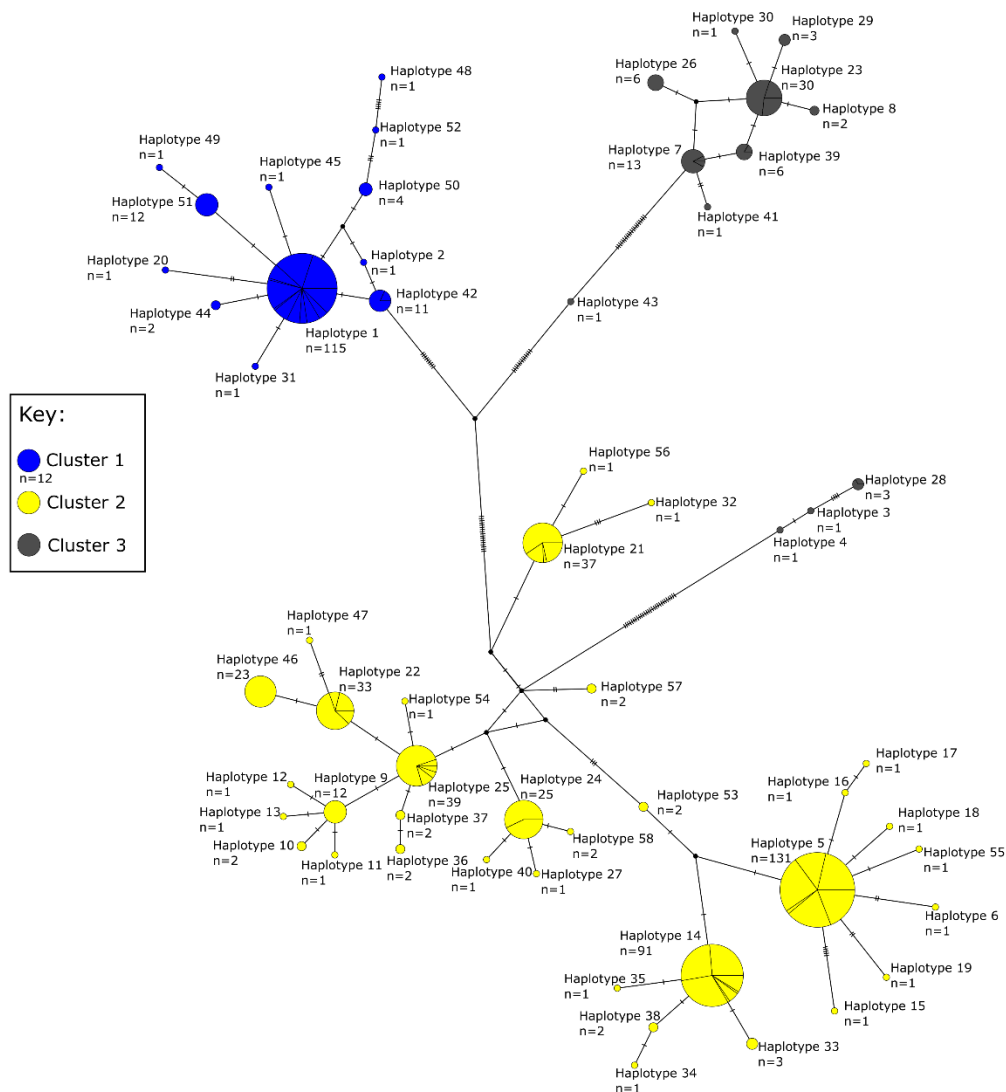
Figure 2.11 - Haplotype network coloured according to clusters. Hatch marks represent one mutational step. Note when samples are sorted into two clusters, clusters 1 and 3 form a single cluster and cluster 2 remains unchanged. Produced using the factoextra and ggplot packages in RStudio v3.

Cluster 2 consisted of the same haplotypes in both cluster plots. This cluster was considerably less diverse than the haplotypes that made up the rest of the dataset. It would be reasonable to assume that cluster 2 is a single species. Due to incongruence between the 28S rRNA phylogeny and the COI mtDNA phylogeny it was not possible to say whether cluster 2 is *Haustrum scobina*, *Haustrum albomarginatum* or a mixture of both.

Table 2.3 - List of 28S rRNA samples and corresponding COI mtDNA haplotypes and clusters. URU 3 was deleted from the COI mtDNA dataset due to poor quality and there is no COI sequence available for FN677467.

Group A	28S Haplotype ('*' Denotes Private)	COI Haplotype	COI Cluster (for 3 clusters)	Group B	28S Haplotype ('*' Denotes Private)	COI Haplotype	COI Cluster (for 3 clusters)
AKA 2	1	1	1	FN677467 (*H. scobina* Barco *et al.* 2010)	4	N/A	N/A
SSB 3	1	4	3	HAW 1	6*	21	2
SSB 4	1	3	3	HAW 10	4	21	2
HAW 30	3*	20	1	MWB 28	7*	40	2
NPL 19	2*	44	1	NPL 18	4	21	2
NPL 8	1	44	1	NPL 2	4	21	2
PRB 22	1	1	1	ONE 1	4	14	2
PRB 6	1	1	1	ONE 2	4	21	2
PRB 8	1	1	1	PRB 2	5*	22	2
STI 1	1	1	1	SSB1	4	53	2
URU 3	1	Deleted	Deleted	SSB2	4	53	2
WPB 11	1	24	2	WPB 12	4	24	2

WPB 14	1	23	3
WPB 25	1	23	3
WPB 6	1	24	2

Group A - *Haustrum albomarginatum*

Haustrum albomarginatum

NPL 8

Haustrum albomarginatum
AKA 2

10mm

Group B - *Haustrum scobina*

Haustrum scobina

NPL 18

Haustrum scobina
HAW 10

Figure 2.12 – Voucher specimens of Group A and Group B 28S rRNA samples.

Shell morphology of voucher specimens from 28S group A matched what was expected of *Haustrum albomarginatum* and voucher specimens of 28S group B matched what was expected of *Haustrum scobina* (Figure 2.12). There is a reasonable degree of error in identifying *Haustrum scobina* and *Haustrum albomarginatum* by shell type though due to morphological similarity and erosion of shell features.

Group B in the 28S rRNA dataset consisted of samples exclusively from cluster 2 in the COI mtDNA dataset (Table 2.3). This dataset cannot give full support to the idea that COI cluster 2 is a single species though, as samples from cluster 2 also appear in Group A of the 28S dataset (Table 2.3). This 28S rRNA dataset was not expanded upon due to inaccessibility to the lab during the COVID-19 pandemic.

There were three large haplogroups in the COI dataset with species-level distance between them when compared to outgroup species (Figure 2.11). *Haustrum haustorium* placed a similar genetic distance from the in-group as the large groups of haplotypes did from each other. The closest haplotype of cluster 3 (excluding outliers; haplotypes 43, 3, 4 and 28) to each of clusters 1 and 2 was haplotype 7 (Figure 2.2 & Figure 2.11). The minimum distance between haplotype 7 and the closest cluster 1 haplotype (42) was 42 mutational steps, (6.89% sequence difference). Between haplotype 7 and the closest cluster 2 haplotype (21) this was 37 mutational steps (6.07% sequence difference). The closest haplotypes to each other of clusters 1 and 2 were haplotypes 42 and 21, which were 27 mutational steps (3.44% sequence difference) from each other. The closest in-group haplotype to *Haustrum haustorium* was haplotype 21, from cluster 2. These were 39 mutational steps (6.39% sequence difference) away from each other. Moreover, this phylogeny could not be rooted with the expected immediate outgroup.

Cluster 3 was particularly diverse and difficult to place phylogenetically. One of its outliers, haplotype 43 was only 22 mutational steps (3.61% sequence difference) from haplotype 42 of cluster 1. The maximum distance between any two haplotypes from cluster 3 was 45 mutational steps (7.38% sequence difference) between haplotypes 28 and 29, though haplotype 28 is a significant outlier. Even within this cluster, there

was more sequence variation than between most clusters and the expected immediate outgroup (*Haustrum haustorium*).

Geographical Distribution of Haplotypes
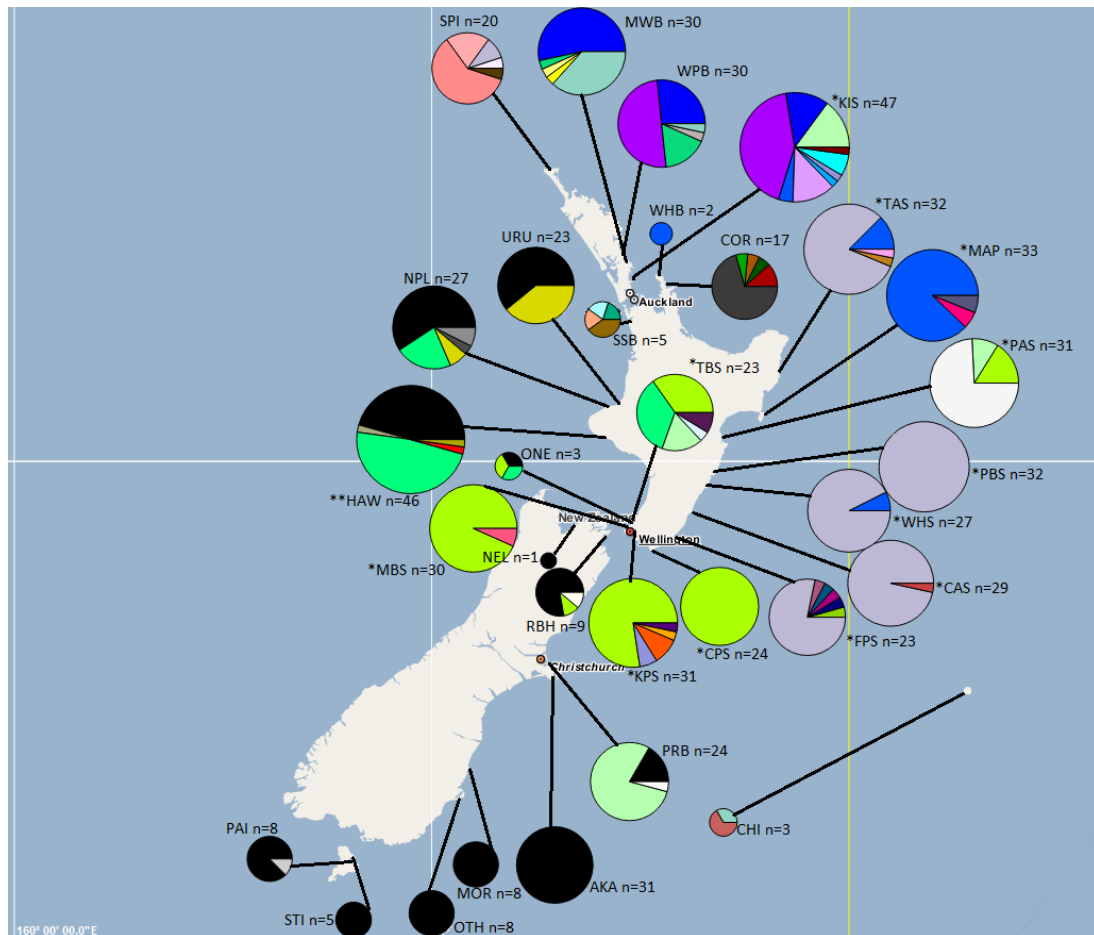
Figure 2.13 - Geographical distribution of all *Haustrum scobina sensu lato* COI haplotypes. This figure represents the two putative in-group species. Due to its large size, the key to haplotypes by colour is in the appendix (Supplementary Figure 1).

The geographical distribution of COI haplotypes (Figure 2.13) showed a lack of diversity in the South Island when compared to the North Island. There was association between the South Island population and the Taranaki population, as haplotype 1 was shared among these sites. There appeared to be a genetic disjunction between Cape Palliser and Flat Point in the south-eastern North Island. The significance of these patterns was difficult to determine without splitting the dataset

into its respective species. Sites with low diversity may only have one species, compared to sites which appear to have high diversity but may consist of two species.
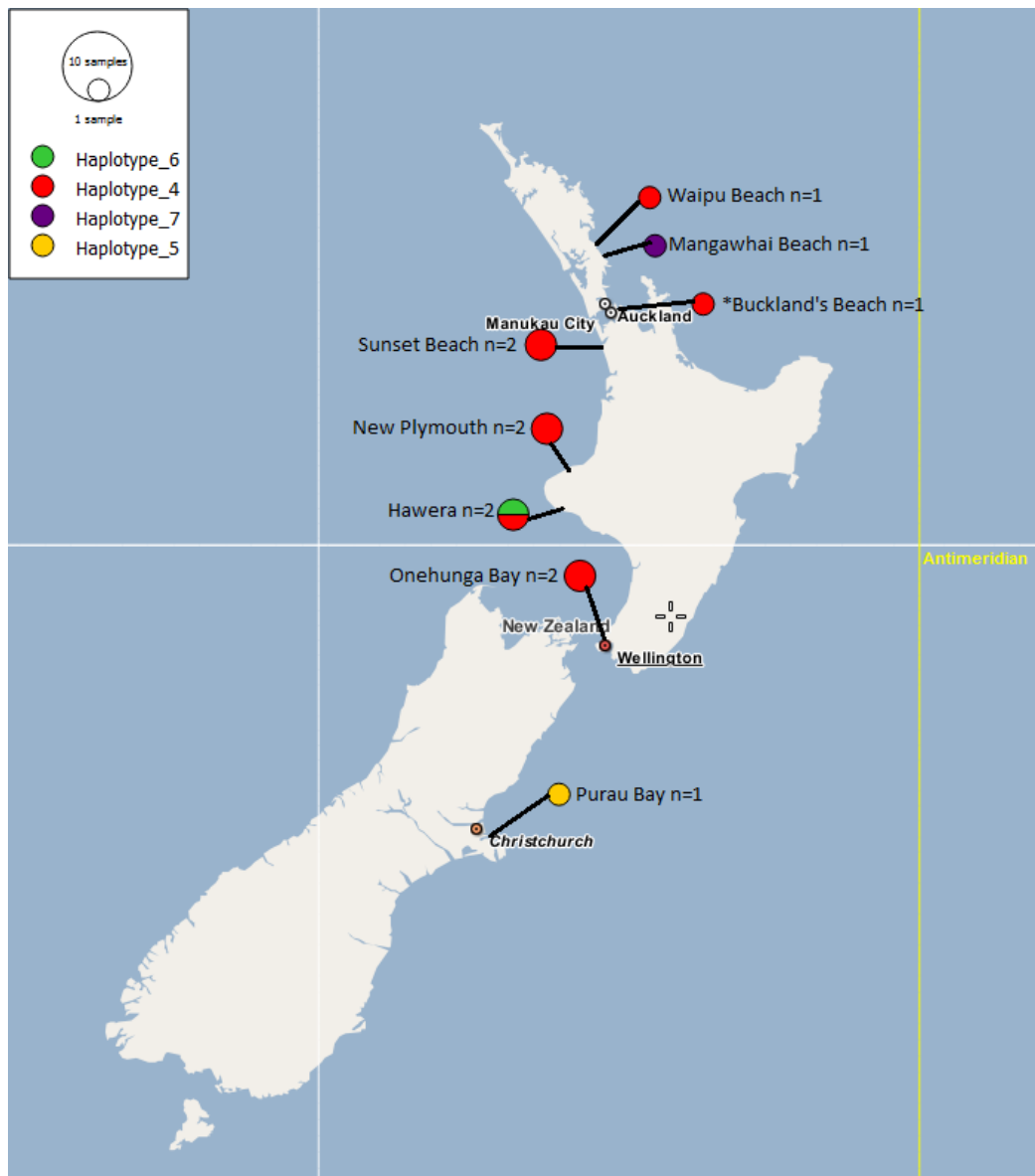
Figure 2.14 - Geographical distribution of all 28S group B haplotypes. These represent *Haustrum scobina sensu stricto*. An asterisk marks Buckland's Beach as it contains one sequences from Barco *et al*. (2010) (FN677467).
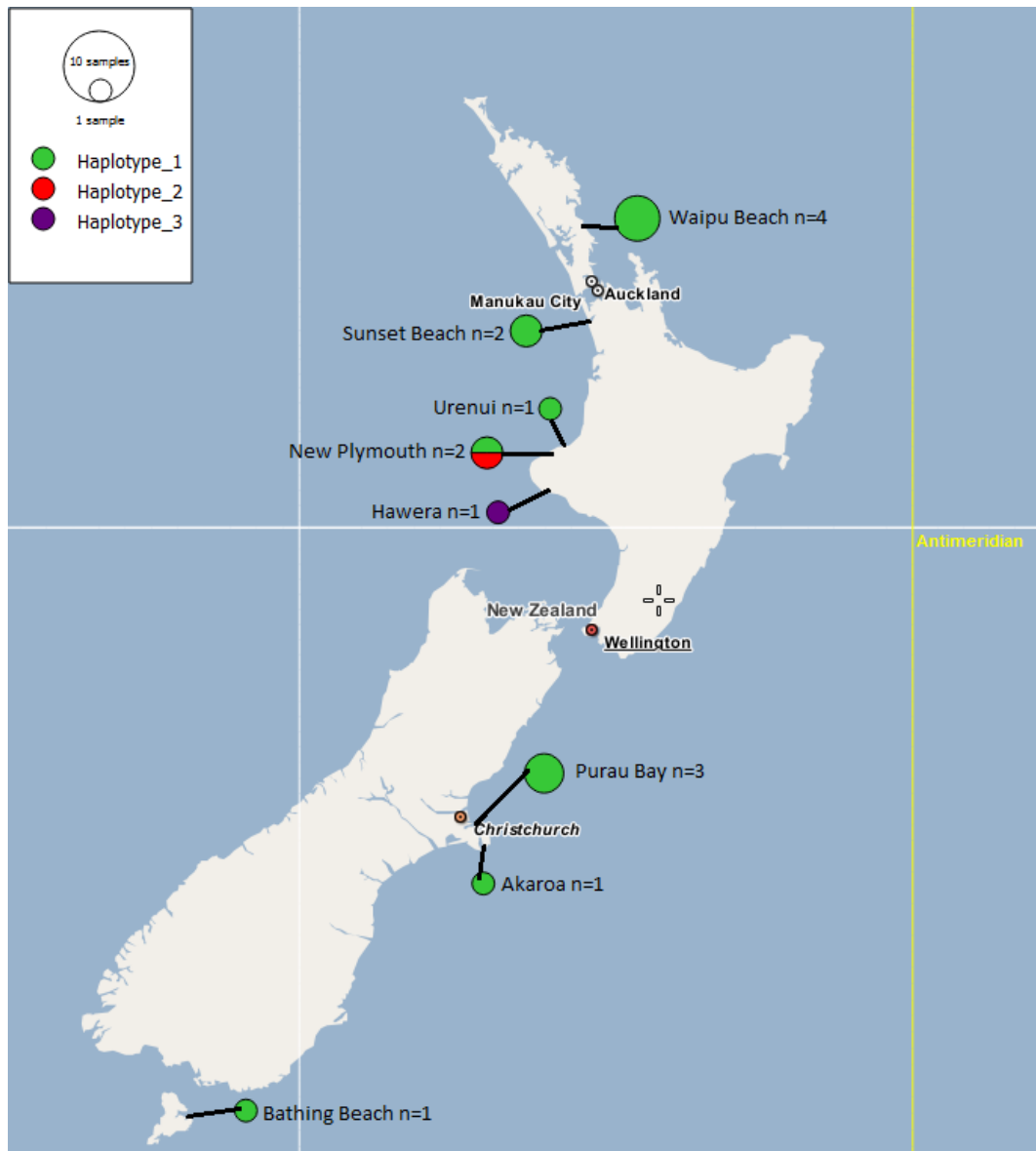
Figure 2.15 - Geographical distribution of 28S group A haplotypes. These represent *Haustrum albomarginatum sensu stricto*.

Haplotypes 4 and 1 for the 28S *Haustrum scobina* and *Haustrum albomarginatum sensu stricto* datasets respectively appeared throughout most of their respective ranges. This can give the appearance of connectivity between all regions, but without a larger sample size and comparable datasets this claim would be unsubstantiated, as it may be an artefact of a slow evolving nuclear marker.

Outgroup Species and Sequences Included in Molecular Analysis

Table 2.4 - Additional sequences used in molecular analysis. The GenBank accession number is listed unless samples were processed in this study (marked in bold). Sequences from Logan (2019) are not included here as they were not published and were obtained with permission to be used in the primary dataset. Published datasets include Ayre *et al.* (2009), Barco *et al.* (2010) and Barco *et al.* (2015).

Species	Genbank Accession # or site code	Marker	Location	Museum registration number
Bedeva vinosum	FJ516017	COI	South-eastern Australia	-
Haustrum lacunosum	KP844984	COI	Ringaringa, Stewart Island, Intertidal	M.285283/3
Haustrum lacunosum	FN677411	COI	St. Clair, Dunedin, Intertidal	UO LL2-SC
Haustrum haustorium	FN677410	COI	Pukerua Bay, Intertidal	UO HH-PB
Haustrum haustorium	**WKH1**	COI	Waikokopu, Intertidal	M.326577/1
Haustrum haustorium	**WKH2**	COI	Waikokopu, Intertidal	M.326577/2
Haustrum albomarginatum	KP844974	COI	SW of Hawera, Intertidal	M.301488/8
Haustrum albomarginatum	KP844975	COI	SW of Hawera, Intertidal	M.301488/6
Haustrum albomarginatum	KP844976	COI	SW of Hawera, Intertidal	M.301488/5
Haustrum albomarginatum	KP844977	COI	SW of Hawera, Intertidal	M.301488/4
Haustrum albomarginatum	KP844978	COI	SW of Hawera, Intertidal	M.301488/3
Haustrum albomarginatum	KP844979	COI	SW of Hawera, Intertidal	M.301488/2
Haustrum albomarginatum	KP844980	COI	SW of Hawera, Intertidal	M.301488/1
Haustrum albomarginatum	KP844981	COI	SW of Hawera, Intertidal	M.301488/7
Haustrum scobina	KP844985	COI	SW of Hawera, Intertidal	M.301489/5
Haustrum scobina	KP844986	COI	SW of Hawera, Intertidal	M.301489/7
Haustrum scobina	KP844987	COI	SW of Hawera, Intertidal	M.301489/8

Haustrum scobina	KP844988	COI	SW of Hawera, Intertidal	M.301489/4
Haustrum scobina	KP844989	COI	SW of Hawera, Intertidal	M.301489/3
Haustrum scobina	KP844990	COI	SW of Hawera, Intertidal	M.301489/2
Haustrum scobina	KP844991	COI	SW of Hawera, Intertidal	M.301489/1
Haustrum scobina	KP844992	COI	SW of Hawera, Intertidal	M.301489/6
Haustrum lacunosum	FN677465	28S	St. Clair, Dunedin, Intertidal	UO LL2-SC
Haustrum haustorium	FN677464	28S	St. Clair, Dunedin, Intertidal	UO HH-PB
Haustrum haustorium	**WAH1**	28S	Waihau Bay	M.326589
Haustrum haustorium	**WKH2**	28S	Waikokopu	M.326577/1
Haustrum scobina	FN677467	28S	Buckland's Beach, Intertidal	UO LS2-BB
Bedeva vinosum	FN677468	28S	Diana's Beach, Tasmania	AMS C458268

Discussion

<u>Identification of Species in COI Sequences</u>

To conduct reliable and robust phylogeographic analysis on any dataset, the recognised species must be identifiable. Conducting phylogeographic analysis on a dataset that consists of more than one species will give misleading results about the levels of diversity and how populations are differentiated. The COI mtDNA phylogeny produced in this study suggested there are as many as three species in *Haustrum scobina sensu lato*. If the tightest cluster, COI cluster 2, was separated and treated as one species, it is likely that some of the diversity in that species would be omitted. This is because some samples of that species fell into other clusters. Misleading indications of diversity might also result from a second, unidentified species being part of COI cluster 2. All group B samples (most likely *Haustrum scobina*) (see Figure 2.12), from the 28S rRNA dataset fell within cluster 2, but not all cluster 2 samples from the COI mtDNA dataset fell within 28S group B. Also, some voucher specimens which appeared to be *Haustrum scobina* based on morphology appeared in clusters 1 and 3 (Figure 2.11). This would mean that any results and conclusions gathered from analysis purely on COI cluster 2 could be misleading.

While the COI dataset cannot be divided into putative species and their populations analysed by methods such as AMOVA, it is still a useful dataset. Shared haplotypes between regions still give some indication of gene flow, and areas of low haplotypic diversity can be used to infer some level of restriction to gene flow. This study was intended as a survey of *Haustrum scobina sensu lato* over most of New Zealand, with the intention of identifying future research directions.

The *Haustrum* genus is one of the most abundant group of mobile macro-invertebrates on New Zealand rocky shores (Morton & Miller 1973). The northern-hemisphere analogue genus *Nucella* has been widely studied (Palmer 1990; Gibbs 1993), and COI has also been used to investigate phylogeographic patterns. COI sequences recovered two species in *Haustrum scobina sensu lato* for Barco *et al.* (2015) when 16 samples from Hawera, Taranaki were used. Logan (2019) generated a COI dataset with 379

sequences. During that study, the only suggestion that the haplotype network was complicated and difficult to interpret was the divergent haplotype 5, but this was only found in three samples, so it was assumed to be sequencing or processing error. Logan (2019) suggested a phylogeographic study of both *Haustrum scobina* and *Haustrum albomarginatum* in parallel using more comprehensive sampling throughout their distribution. COI was expected to be a suitable marker, but high haplotypic diversity, an insufficient sample size and incongruence with other markers only became apparent after a large dataset representing sites from across New Zealand was generated.

<u>Neutrality and Mismatch Analysis</u>

Significant negative Fu's Fs values were found for sample sites Opito Bay and Flat Point (see Table 2.2). Significant negative Tajima's D values were found for Opito Bay, Flat Point, Kau Point, Mangawhai Beach and Robin Hood Bay (see Table 2.2). Negative and significant Tajima's D and Fu's Fs values indicate an excessive number of alleles, which may be the result of either recent population expansion, selection or genetic hitchhiking. Harpending's raggedness index was non-significant for these sites, indicating a good fit of data to the expected diversity. However, the sum of squares deviation was significant for the Chatham Islands, Opito Bay, Hawera, Mangawhai Beach, Onehunga Bay, Paterson Inlet and Urenui (see Table 2.2). This indicates unexpectedly high levels of haplotype diversity. A significant number of rare haplotypes (see Figure 2.2) indicated population expansion at these sites was more likely than selection. This could be a result of more than one species being present at each site and must be given appropriate consideration under this condition.

<u>Is There Evidence for More Than One Species in *Haustrum scobina sensu lato*?</u>

The pooled samples showed high levels of COI diversity. Several samples fell into the aberrant clade found by Logan (2019); these were haplotypes 3, 4 and 28 with a total number of five samples. This suggested the clade reported in the previous study was not erroneous. Large sample sizes in all clusters of the COI mtDNA dataset indicated that the diversity found was a reasonably good representation of the population and not a result of sequencing or processing errors.

In the case where there are cryptic species in *Haustrum scobina sensu lato*, each cluster would be a single species and the outgroup species an expected species-level distance from the in-group species (hypothesis two). Even in this scenario, there were large distances between some haplotypes in cluster 3. The use of a single marker (COI) was not enough evidence to accept this was the case. Here the species problem becomes evident. Distinct groups of species cannot always be identified from a dataset analysis. Before these clades can be considered species or sub-species, common phenomena such as incomplete lineage sorting and introgression must be tested for.

The 28S dataset showed a more obvious set of species groups than the COI dataset (Figure 2.6). There were two reciprocally monophyletic clades; each is likely to be a separate species. *Haustrum lacunosum* placed a relatively large distance from the in-group species and *Haustrum haustorium*. This dataset suggested that there are in fact only two species in the in-group. The expected immediate outgroup (*Haustrum haustorium)* branched just as close to the two in-group clades as they did to each other (see Figure 2.6). This appeared to conflict with previous findings (Barco *et al.* 2015) which suggested members of *Haustrum scobina sensu lato* are each other's closest relatives, with *Haustrum haustorium* being more distantly related. This finding was supported by the COI dataset. It is worth noting however, that while this study included more samples than previous studies on the genus, it used fewer markers. The 28S dataset in this study also consisted of shorter sequences than those used by Barco *et al*. (2015). The shorter sequence length in this dataset was due to large low-quality regions being present at the end of each sequence, which had to be trimmed. The 28S

dataset supported the conclusion that there are two species in the in-group (*Haustrum scobina* and *Haustrum albomarginatum*). The null hypothesis, that there are two species in *Haustrum scobina sensu lato*, could not be rejected.

Is the 28S Dataset Reliable?

Twenty-seven 28S sequences (50% of all original 28S samples) were discarded from the final 28S dataset. As the 28S dataset was being used to resolve the complicated phylogeny recovered from the COI dataset, it was imperative that the sequences were high quality. Issues with weak priming on large sub-unit 28S rRNA have been encountered in other studies (McArthur & Koop 1999). Colgan *et al.* (2000) found large variation of evolutionary rates in 28S across various gastropod orders and found that much of the variability was in discrete areas of the sequence. Rosenberg *et al.* (1994) also found considerable variation in mutation rates of the 28S D6 region between taxa, even on an intra-generic level. McArthur and Koop (1999) required complex evolutionary models in analysis of 28S rRNA sequences across 32 gastropod genera. Despite potential complications, 28S was an appropriate choice as an independent marker to resolve the in-group phylogeny in this study as preliminary data was already available (Barco *et al.* 2010).

The 28S marker performed well for distinguishing intra-generic relationships in the *Haustrum* genus. Reasonable sequence diversity for a small sample size (seven haplotypes from 27 samples) also indicated it would be suitable for intra-specific phylogenetic and phylogeographic analysis. The closest haplotypes between 28S groups A and B were eight mutational steps apart (sequence difference of 0.97%) (see Figure 2.6). Between group A and *Haustrum haustorium* this was six mutational steps (sequence difference of 0.72%), and between group B and *Haustrum haustorium* there were eight mutational steps (sequence difference of 0.97%). This is still reasonably low for species-level distances, relative to other intra-generic molluscan studies. Komaru *et al.* (2012) found 28S rRNA sequence variation rates between 1.1% and 2.68% between species of *Corbicula* clams.

There was no indication in this study that the 28S dataset should not be trusted. A more extensive study using this approach would be a valuable next step for the research. The small 28S rRNA dataset reported here is not large enough to confidently recover the phylogeny of the in-group. Work may be required to optimise primers for the *Haustrum* genus if 28S were to be used for a larger sample size. However, additional sampling might show that 28S rRNA is more conserved than expected or indicated in this study.

Corroborating the COI mtDNA and Nuclear 28S rRNA Phylogenies

The intent of using a 28S dataset was to corroborate findings with the COI dataset and identify species in the COI haplotypes. However, as samples from all COI clusters appeared across both in-group clusters in the 28S dataset it was difficult to identify species in COI clusters and haplotypes. This made reliable population structure analysis of the COI dataset difficult to assess. A more extensive genetic dataset will likely be required (e.g. SNPs from genome sequencing or whole mitochondrial genomes) for a better level of resolution. Whole mitochondrial genome sequences might yield interesting results if the reason for the convoluted COI phylogeny can be identified.

Increasing the COI sample size by an order of magnitude could show a more complete picture of the COI haplotype diversity. The goal in this approach is to either fill in the large distances between haplogroups or provide a better assessment of how this lineage-level diversified. It is possible that 'intermediate' haplotypes between clusters exist, the way haplotype 43 of COI cluster 3 sat more centralised to the dataset than other haplotypes in cluster 3 (Figure 2.6). However, to discover these would require an impractically large number of samples and is unlikely to be achievable. Whole mitochondrial genome sequences would give more resolution and enable a test for whether COI is hitchhiking on a linked gene. One of the key goals in this study was to identify areas of high diversity and taxonomic anomalies. This study still achieved that.

Rare haplotypes branching off the most common haplotypes (1 and 4) in the 28S median-joining haplotype network (Figure 2.6) suggested there has been recent radiation in each of the in-group lineages. This result is consistent with many other phylogenetic studies of similar molluscs (Fleming *et al.* 2017; Dohner *et al.* 2018; Walton *et al.* 2018). However, the 28S dataset did not contain enough samples at each site to support phylogeographic interpretation of population migration patterns.

Many phylogeographic studies support the idea that populations have shifted back and forth between areas of glacial refugia during glacial maxima and the open coastline after glaciation (Fleming *et al.* 2017; Dohner *et al.* 2018; Walton *et al.* 2018). There was some evidence this may also be the case in *Haustrum scobina sensu lato*. It is important to remember that glacial periods can last just as long as inter-glacial periods. Areas of 'glacial refugia' could also be considered a species' normal distribution, while range expansion during inter-glacial periods might be the ab-normal condition.

Lack of COI haplotype diversity at southern sample sites could be the result of recent range expansion into southern areas as a series of founder events, as suggested for several other taxa (Fleming *et al.* 2017). However, as it was not possible to identify species by sequences from the COI data, the presence of a single species (likely *Haustrum albomarginatum* based on voucher specimens) in southern regions could also give the appearance of reduced diversity. High haplotype diversity levels in north-eastern North Island populations indicated this could have been an area with glacial refugia (Maggs *et al.* 2008). Haplotype 1 was shared between the South Island, Onehunga Bay and Taranaki. High frequencies of haplotype 1 and higher levels of diversity in Taranaki indicated this region may have acted as glacial refugia for South Island populations. During the LGM when sea levels were 120 m lower than today, a land bridge most likely connected Farewell Spit and South Taranaki (Lewis *et al.* 1994). Populations may have been connected along the coastline from Golden Bay to Taranaki, and subsequently migrated south as the climate warmed. It is also worth

noting that range expansion after glacial periods may not be a function of latitude or temperature but of suitable habitat. Much of the South Island was covered in glacial ice down to the coast during the LGM, particularly on the western side of the Southern Alps (Suggate & Almond 2005). This would have reduced the available habitat, but rocky habitats free from ice scour likely still existed, as they do in Antarctica today. Reduced sea levels may have resulted in coastlines being predominantly mud or sand at latitudes where rocky reef is found at the coast today. This could have further reduced suitable habitat for *Haustrum scobina sensu lato* in certain areas.

It is likely that recent environmental processes such as the LGM have consistently had similar influence across a broad range of species. Populations can be forced to move north or south if they are sensitive to thermal change, or lack of suitable habitat. Environmental forces during glacial cycles can have a direct effect on the distribution and evolution of species (Wardle 1963). The southward range of *Haustrum scobina* appeared to be limited in comparison to *Haustrum albomarginatum* (K. Walton pers. comm. 2019; pers. obs. 2019). The southernmost record of *Haustrum scobina* comes from Purau Bay, Lyttleton Harbour. This was found during sampling for this study. *Haustrum albomarginatum* have been observed as far south as Stewart Island. *Haustrum albomarginatum* might have greater tolerance to cold conditions or may have expanded southwards more quickly after the LGM. It is entirely possible *Haustrum scobina* is still expanding southwards since the end of the LGM ~17.65 kya. (Suggate & Almond 2005), and by chance *Haustrum albomarginatum* may have persisted in southern regions during the LGM. Range observations are based on shell features and should be considered with a degree of scepticism.

The Purau Bay samples shared haplotype 22 with Titahi Bay, Port Ahuriri and Kawau Island, but there was nowhere between these locations where haplotype 22 was found (see Figure 2.13). Given Purau Bay is the most southern record to date of *Haustrum scobina* (based on morphology) and happens to be in the heavily trafficked Lyttelton Harbour it is likely this population was established by human-mediated transportation. Titahi Bay, Port Ahuriri and Kawau Island are all in busy harbours or ports. This result is consistent with findings from Fleming *et al.* (2017) who attributed association between the Hauraki Gulf and Nelson in *Cominella virgata* to human-

mediated transport. *Haustrum scobina* may have been transported to Lyttelton harbour as foul on ships or more likely, with the rocks used as ballast in 19th/20th century ships (Hewitt *et al.* 2009).

The Chatham Islands samples shared haplotype 7 with some Northland samples (see Figure 2.13). The East Auckland Current and East Cape Current could carry rafting egg capsules to the Chatham Rise and Subtropical Convergence zone. Rafting on southern bull-kelp (*Durvillaea antarctica*) has been proposed as a potential rafting vector for several New Zealand species, as it has been shown to have genetic structuring consistent with New Zealand currents (Bussolini & Waters 2015). Other potential vessels include pumice and driftwood but *Haustrum scobina* has not been observed to raft on any of these, including bull-kelp.

There appeared to be a genetic disjunction between the south Wellington/Wairarapa coast and the eastern Wairarapa coast between Castlepoint and Flat Point (see Figure 2.13). A similar pattern was reported in *Cominella maculosa* by Fleming *et al.* (2017). This region has been subject to several major events of sudden uplift throughout the Holocene (McSaveney *et al.* 2006). Hay (2020) and Parvizi *et al.* (2019) give evidence for uplift in the region structuring the population of southern bull-kelp and causing discrete areas where the species is absent. Uplift events may have structured populations of benthic invertebrates in similar ways, but comprehensive sampling across the region would be required to give support to this idea.

Potential Causes of the COI Complications

The COI dataset suggested considerable levels of intermixing between species. This could be a result of mitochondrial introgression, or incomplete lineage sorting since the time they diverged into different species. Central common haplotypes of each species with rare haplotypes branching in a star-like pattern in the 28S median-joining haplotype network (Figure 2.6) indicated there may have been recent separation and radiation; if this event occurred rapidly, it is likely that incomplete lineage sorting has resulted in a complication and unresolved COI dataset. Hills *et al.* (2011) suggest COI

to be an appropriate marker for gastropod taxa that are not overly species dense and separated by 1 to 50 ma. *Haustrum scobina* and *Haustrum haustorium* are recorded in Castlecliffian strata, which is dated between 1.63 and 0.34 ma (Beu & Maxwell 1990; Hollis *et al.* 2010). No research mentions *Haustrum albomarginatum* in the fossil record, though this is likely due to its morphological similarity to *Haustrum scobina* rather than an absence. This leaves the possibility of the *Haustrum albomarginatum* lineage having only recently diverged, though *Haustrum scobina* fossils would need to be reviewed and confirmed as such to give any evidence for this.

Another explanation is that high haplotypic diversity in 28S was a result of introgression between previously isolated groups. Given that 28S samples from both species appeared across the entire COI dataset, introgression is difficult to rule out. Mitochondrial DNA retains a genetic pattern of population history much longer than nuclear DNA because it does not undergo recombination. A mixture of genetic signals from one or more population-wide events could be retained in the mitochondrial genome from before the *Haustrum scobina sensu lato* lineage was still actively bifurcating.

Significant glacial cycles are controlled by Milankovitch cycles on 26 ka, 41 ka and 100 ka cycles (Berger 1988). Axial precession cycles over 26 ka periods do not lead to the great changes in ice cover which are seen on 100k a time scales due to eccentricity changes but do occur more frequently. Frequent glaciation and ice melt repeatedly covers and exposes suitable habitat. This can isolate and re-connect populations repeatedly. Allopatric speciation under such conditions might be interrupted frequently and certain genes might be more prone to incomplete lineage sorting as a result. This could also result in introgression if hybrids crossed back and forth between species. Both *Haustrum scobina* and *Haustrum albomarginatum* are found across New Zealand in great abundance. If allopatric speciation resulted in their divergence, then a major nation-wide population expansion of at least one species must have occurred.

The possibility of selection acting on the COI gene cannot be ruled out either. While analysis indicated the COI gene is unlikely to be under selection, it is possible that COI is genetically hitchhiking on another mitochondrial gene which is under the

influence of selection. This could result in retention of an old signal which has been selected for before the *Haustrum scobina* and *Haustrum albomarginatum* lineages bifurcated. This signal is likely to spread through both populations if allopatric speciation is interrupted regularly.

Haustrum Genus Taxonomy – What New Information Is There?

It was clear from both the COI and the 28S datasets that *Haustrum lacunosum* is a distinct species and the immediate outgroup species to *Haustrum haustorium* and *Haustrum scobina sensu lato*. *Haustrum haustorium* is normally assumed to be the immediate outgroup species to *Haustrum scobina sensu lato,* and clades within *Haustrum scobina sensu lato* were predicted to be each other's closest relatives (Barco *et al*. 2015; K. Walton pers. comm. 2019). Both datasets in this study showed little evidence that this is the case. The 28S Bayesian tree recovered a polytomy between *Haustrum haustorium*, *Haustrum scobina* and *Haustrum albomarginatum* (Supplementary Figure 4). The COI Bayesian tree placed a large number of haplotypes in a clade separate to the rest of the haplotypes and *Haustrum* species (Supplementary Figure 4). Members of *Haustrum scobina sensu lato* did not form a monophyletic clade in either dataset from this study.

Morphology of intertidal molluscs can be highly plastic and is at times a poor method of distinguishing between species (Zou *et al.* 2012). There has been a lot of research on *Nucella lapillus* (a Northern Hemisphere analogue to *Haustrum* spp.) populations affected by interrupted breeding in due to sterilisation by TBT (tributyltin) pollution. One study transplanted *Nucella lapillus* from the exposed Cornish coast of England to two sheltered inlets near Devon, where native populations had stopped breeding due to TBT poisoning (Gibbs 1993). Transplanted individuals had a thin shell and large operculum, which is typical in areas of strong surf where a broad foot is required for adhesion. Progeny were typical of species more heavily predated with a thicker shell and narrower operculum. This differential gene expression occurred over just one generation showing significant phenotypic plasticity. It is possible many Muricid

lineages have the potential to exhibit a diverse range of morphological characteristics even within a small gene pool.

Haustrum haustorium tends to grow larger than *Haustrum scobina sensu lato* but prior to adult stages these groups are morphologically similar. One of the most noticeable difference between the two groups is shell thickness and operculum width (pers. obs. 2020). *Haustrum haustorium* generally have a proportionately thinner shell and wider operculum relative to *Haustrum scobina sensu lato*. It is possible that predation on the smaller *Haustrum scobina sensu lato* could be part of the reason for differences in shell thickness and operculum width. Significant variation has been observed in phenotypes of *Nucella lapillus* under controlled conditions where juvenile snails were raised exposed to either crab (their main predator) and/or fish effluent or a control (Palmer 1990). The experiment included juveniles of exposed shore and protected shore phenotypes. Both laboratory and field results indicated that the exposed shore population was more labile morphologically than the protected shore population. Almost all specimens in the risk treatment category displayed adaptive variation, which included increased size-adjusted shell weight, thicker shell lips, increased retractability and change in apertural tooth height. High levels of phenotypic plasticity in Muricids are possible in a small gene pool and appears to be dependent on external environmental factors. Differing environmental niches may have more influence on the morphological differences between *Haustrum haustorium* and *Haustrum scobina sensu lato* than genetics.

Most Muricidae species, (including *Haustrum scobina sensu lato*) exhibit an outer calcite layer on their shells as calcite is less readily dissolved in cold water than aragonite (Taylor & Reid 1990; Tan 2003). *Haustrum haustorium* lacks this feature though (Tan 2003), which may give some indication as to why they have thinner shells. Operculum size tends to increase in areas of strong surf to give gastropods a better hold on their substrate (Gibbs 1993). Morton and Miller (1973) note that while *Haustrum scobina* is abundant in areas of strong surf, *Haustrum haustorium* is less frequently found on exposed coasts. This suggests that if environmental influences are the main driver of morphological differences between *Haustrum haustorium* and *Haustrum scobina sensu lato*, ecological interactions have a greater effect than physical factors, given *Haustrum haustorium* exhibits morphological features more

typical of species found on exposed coasts. *Haustrum haustorium* is also omnivorous, whilst *Haustrum scobina sensu lato* is carnivorous (Novak 2013). Calcite and aragonite are both commonly found forms of $CaCO_3$ in invertebrate shells and either form can often be selected for by the animal (Falini *et al.* 1996). Dietary preference can be behavioural and investigation into gut morphology of these species might reveal this. With the current evidence, it is difficult to explain the polytomy between *Haustrum scobina, Haustrum albomarginatum* and *Haustrum haustorium*.

Chapter 3 - Conclusions and Further Study

Study Aims

The aim of this study was to investigate the phylogeny and phylogeography of the *Haustrum scobina sensu lato* lineage. However, the dataset was unable to resolve the phylogeny of the expected species in the COI dataset. Several explanations were proposed for the findings reported in this thesis research. These included mechanisms such as incomplete lineage sorting or genetic hitchhiking. The 28S dataset suggested that *Haustrum scobina sensu lato* consists of two species. These are most likely *Haustrum scobina* and *Haustrum albomarginatum*. The null hypothesis of *Haustrum scobina sensu lato* consisting of two species could not be rejected in this case giving support to the findings of Barco *et al.* (2015).

Low haplotype diversity in the South Island population suggested that there have been recent colonisation events of these areas. A significant number of rare haplotypes in the COI dataset and a moderate number of rare haplotypes in the 28S dataset indicated recent population expansion. This gave support to the idea that populations have moved and re-colonised southern areas after the LGM. Haplotype 1 was the only haplotype found at several South Island populations but was also found in Wellington and Taranaki. The prevalence of haplotype 1 in Taranaki populations, as well as more diverse populations than in the South Island, indicated that Taranaki may have acted as an area of glacial refugia during the LGM for *Haustrum scobina sensu lato*.

Haplotype 22 was shared between Purau Bay samples and samples in Port Ahuriri, Kawau Island and Titahi Bay with no intermediate locations. Support is given here to Fleming *et al.* (2017), who proposed shipping as a vector for *Cominella virgata* translocation between ports. The Chatham Islands population also shared haplotype 7 with Mangawhai and Waipu populations. This may be due to rafting or human-mediated transport, but identifying an exact vector of transport is difficult.

Inability to divide the COI dataset into clearly delimited species limited the extent to which phylogeographic scenarios could be tested. Regional haplotype associations

may turn out to be a single species with very little COI diversity. For such reasons, little support can be given to these phylogeographic interpretations until the COI phylogeny is understood.

Perhaps the most surprising result to come from this research was the phylogenetic placement of *Haustrum haustorium*. The 28S Bayesian tree placed *Haustrum haustorium* in a polytomy with *Haustrum scobina* and *Haustrum albomarginatum*. It did not recover *Haustrum scobina sensu lato* as a monophyletic clade. This was also indicated by the median-joining haplotype network of the 28S dataset, but contradicted by both maximum likelihood trees, which recovered *Haustrum scobina sensu lato* as a monophyletic clade.

Morphology is highly plastic in Muricids (Palmer 1990; Gibbs 1993; Zou *et al.* 2012) and morphological convergence has led to many misleading phylogenetic results in Muricid lineages. *Haustrum scobina sensu lato* has been a notoriously difficult lineage to resolve. The hypothesis that *Haustrum scobina sensu lato* is not a monophyletic lineage may have limited support from other aspects of the *Haustrum* genus. However, it should not be immediately disregarded. The Muricidae family has proven to be a very complicated group for malacologists to resolve taxonomically.

Future Research

As with many Muricidae lineages, the *Haustrum* genus and *Haustrum scobina sensu lato* in particular are complex and highly diverse. Any phylogeographic studies on *Haustrum scobina sensu lato* will need to prioritise extensive sampling and multiple gene sequences. Determining the phylogenetic position of *Haustrum haustorium* is crucial to interpreting the relationship between members of *Haustrum scobina sensu lato*.

A 'brute force' approach, as mention in Chapter 2, is an approach to get a better understanding of the phylogeny. A common problem with phylogeographic studies is that the best sample size is best determined after a reasonably widespread initial

survey of the study species (Wong *et al.* 2011). This study appears to have encountered a complex set of species and populations that required large sample sizes to resolve. The COI haplotype rarefaction curve (Figure 2.3) indicated significantly more haplotype diversity would be discovered with further sampling. This curve may begin to flatten off with a sample size in the thousands, but it is difficult to determine exactly how many samples will be required. If all haplotype diversity in the COI lineage is discovered, large distances between some COI haplotype clusters may be filled in. This could reveal that part of the dataset is just one species with high diversity. Further COI sequencing should target areas in which these phylogenetic issues have been most prominent. The localities from the north-eastern North Island which comprise most of Cluster 3 (Mangawhai, Waipu and Kawau Island) are worth investigating further.

Given the incongruence between the COI phylogeny, shell morphology types and the 28S phylogeny though, it is quite likely the COI gene in *Haustrum scobina sensu lato* is influenced by complex past environmental issues. COI may not be phylogenetically informative with small sample sizes. Investing the time and resources required to expand on this dataset risks gaining little new insight without a very large investment. Sequencing the whole mitochondrial genome for a subsample of those included in this study could reveal why the COI lineage is producing such a complex signal.

Tan (2003) conclude that morphological evidence separating *Haustrum scobina* and *Haustrum albomarginatum* is trivial and not sufficient to differentiate the two as species. Audzijonyte *et al.* (2012) state they found several cases in deep-sea clams (Vesicomyidae and Pliocardiinae) where morphologically described species were simply representative of morphological plasticity but no cases in those clams where species-level COI differences were not accompanied by at least some degree of morphological variation in cryptic species. This point both gives validity to and undermines conclusions by Tan (2003). Morphological characteristics alone should not be considered enough evidence for resolution of species differences in molluscs. Genetic data is more likely to provide an accurate marker for detecting species-level differences. With that in mind, morphological studies of the *Haustrum* genus may have already provided all the information they can. Future studies would likely benefit most from focusing on the genetic aspect of species differences in the genus.

Johnson and Cumming (1995) found specimens with shell characteristics intermediate of two Muricids; *Drupella fragum* and *Drupella rugosa.* They did not observe any interbreeding between the two species though. Assigning species as hybrids based on shell morphology would be futile in the *Haustrum* genus and even in the wider *Muricidae* family given morphological convergence and plasticity. Holding live specimens of *Haustrum scobina* and *Haustrum albomarginatum* in captivity to observe whether they interbreed would be a relatively simple study requiring little investment of time or resources. This is worth investigating in future studies to explore the possibility of hybrid species.

The portion of the large sub-unit 28S rRNA locus used in this study proved effective for delimiting intra-genus relationships in *Haustrum* and appeared to have sufficient diversity for intraspecific studies too. Phylogeographic studies of the *Haustrum* genus may opt to use an entirely different marker than COI in the future. Amplification and sequencing of 28S proved to be somewhat problematic in this study, but refined techniques may be able to solve these issues. However, added complexity of analysing nuclear markers and fewer studies to compare results with are drawbacks.

There was evidence in this study, and in previous comparable studies (Fleming *et al.* 2017; Walton *et al.* 2018) that glacial cycles have influenced the distribution of direct developing benthos. Reduced genetic diversity in southern populations when compared to northern populations indicated post-glacial re-colonisation of southern sites. On-going colonisation of southern sites might be expected as the climate continues to warm. Further to this, species may actually be displaced from the northern limits of their range (or southern in the Northern Hemisphere) if sea temperatures exceed their thermal maximum, as is a common occurrence now in many coral species (Bridge *et al.* 2014). The average temperature resulting in 50% mortality (LT50) of *Haustrum scobina* is 38.2°C (Moffitt 2019). Moffitt (2019) suggest future warming will negatively impact north-eastern Auckland populations of *Haustrum scobina*. Oxygen content in water decreases with increased temperature too, making conditions harsher for some species (Davenport & Davenport 2007). Part of the motivation for studying the contemporary distribution and inferring historical migration of a species, is to better inform ourselves of how they will react to future change. Experiments to determine the upper thermal tolerance of a species could

prove extremely valuable when planning to mitigate loss of biodiversity. Assessing the thermal tolerance of *Haustrum scobina* and *Haustrum albomarginatum* across their entire distribution could reveal at-risk areas. Direct developing gastropods are by their nature slow to migrate. While they may have migrated south in New Zealand over almost 20 ka since the LGM, northern populations may not have the capacity to migrate southwards fast enough under accelerated warming (Yao *et al.* 2017).

Closing Remarks

"It still remains necessary to emphasize that tentative conclusions are hypotheses for testing and that the gulf between raw data and paleogeographic conclusions has often been crossed by a rather flimsy bridge of projection and extrapolation… In the very inexact science of biogeography, hypotheses can seldom be directly proved, though they are tested by new data at every step, and in Charles Darwin's words 'the doctrine must sink or swim according as it groups and explains phenomena" (Fleming 1975).

The technology available to researchers has come a long way since the 1970's, though these words from Charles Fleming, one of New Zealand's most esteemed natural scientists remain true. Even with the remarkable resolution available to us today, we can still only chip away at the mass of information that is the natural world, gathering small amounts of information from increasingly large datasets at every step.

References

Adachi, N. & Wada, K. (1999). Distribution in relation to life history in the direct–developing gastropod Batillaria cumingi (Batillariidae) on two shores of contrasting substrata. Journal of Molluscan Studies, 65, 275-288.

Aljanabi, S.M. & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research, 25, 4692-4693.

Allaire, J. (2012). RStudio: integrated development environment for R. Boston, MA, 770, 394.

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

Apte, S. & Gardner, J.P. (2001). Absence of population genetic differentiation in the New Zealand greenshell mussel Perna canaliculus (Gmelin 1791) as assessed by allozyme variation. Journal of experimental marine biology and ecology, 258, 173-194.

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.

Audzijonyte, A., Krylova, E.M., Sahling, H. & Vrijenhoek, R.C. (2012). Molecular taxonomy reveals broad trans-oceanic distributions and high species diversity of deep-sea clams (Bivalvia: Vesicomyidae: Pliocardiinae) in chemosynthetic environments. Systematics and Biodiversity, 10, 403-415.

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

Ballard, J.W.O. & Whitlock, M.C. (2004). The incomplete natural history of mitochondria. Molecular ecology, 13, 729-744.

Bandelt, H.-J., Forster, P. & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution, 16, 37-48.

Barco, A., Claremont, M., Reid, D., Houart, R., Bouchet, P., Williams, S. et al. (2010). A molecular phylogenetic framework for the Muricidae, a diverse family of carnivorous gastropods. Molecular Phylogenetics and Evolution, 56, 1025-1039.

Barco, A., Marshall, B., Houart, R. & Oliverio, M. (2015). Molecular phylogenetics of Haustrinae and Pagodulinae (Neogastropoda: Muricidae) with a focus on New Zealand species. Journal of Molluscan Studies, 81, 476-488.

Berger, A. (1988). Milankovitch theory and climate. Reviews of geophysics, 26, 624-657.

Beu, A. (2012). Marine Mollusca of the last 2 million years in New Zealand. Part 5. Summary. Journal of the Royal Society of New Zealand, 42, 1-47.

Beu, A., Alloway, B., Pillans, B., Naish, T. & Westgate, J. (2004). Marine Mollusca of oxygen isotope stages of the last 2 million years in New Zealand. Part 1: Revised generic positions and recognition of warm-water and cool-water migrants. Journal of the Royal Society of New Zealand, 34, 111-265.

Beu, A. & Maxwell, P. (1990). Cenozoic Mollusca of New Zealand: New Zealand Geological Survey Palaeontological Bulletin 58. New Zealand, 75-454.

Blackmore, G. (2000). Field evidence of metal transfer from invertebrate prey to an intertidal predator, Thais clavigera (Gastropoda: Muricidae). Estuarine, Coastal and Shelf Science, 51, 127-139.

Bridge, T.C., Ferrari, R., Bryson, M., Hovey, R., Figueira, W.F., Williams, S.B. et al. (2014). Variable responses of benthic communities to anomalously warm sea temperatures on a high-latitude coral reef. PloS one, 9, e113079.

Buhle, E.R. & Ruesink, J.L. (2009). Impacts of invasive oyster drills on Olympia oyster (Ostrea lurida Carpenter 1864) recovery in Willapa Bay, Washington, United States. Journal of Shellfish Research, 28, 87-97.

Bussolini, L.T. & Waters, J.M. (2015). Genetic analyses of rafted macroalgae reveal regional oceanographic connectivity patterns. Journal of Biogeography, 42, 1319-1326.

Carrasco, S. & Phillips, N. (2014). Encapsulation and development of three New Zealand neogastropods with contrasting embryo packaging and maternal provisioning. New Zealand Journal of Zoology, 41, 171-186.

Carriker, M.R. (1981). Shell penetration and feeding by naticacean and muricacean predatory gastropods: a synthesis. Malacologia, 20, 403-422.

Chiswell, S.M. & Booth, J.D. (1999). Rock lobster Jasus edwardsii larval retention by the Wairarapa Eddy off New Zealand. Marine Ecology Progress Series, 183, 227-240.

Chiswell, S.M., Bostock, H.C., Sutton, P.J. & Williams, M.J. (2015). Physical oceanography of the deep seas around New Zealand: a review. New Zealand Journal of Marine and Freshwater Research, 49, 286-317.

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

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

Clark, K., Nissen, E., Howarth, J., Hamling, I., Mountjoy, J., Ries, W. et al. (2017). Highly variable coastal deformation in the 2016 Mw7. 8 Kaikōura earthquake reflects rupture

complexity along a transpressional plate boundary. Earth and Planetary Science Letters, 474, 334-344.

Colgan, D.J., Ponder, W.F. & Eggler, P.E. (2000). Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. Zoologica Scripta, 29, 29-63.

Cong, Q., Shen, J., Borek, D., Robbins, R.K., Opler, P.A., Otwinowski, Z. et al. (2017). When COI barcodes deceive: complete genomes reveal introgression in hairstreaks. Proceedings of the Royal Society B: Biological Sciences, 284, 20161735.

Cowen, R.K. & Sponaugle, S. (2009). Larval dispersal and marine population connectivity.

Craw, D., Upton, P., Burridge, C.P., Wallis, G.P. & Waters, J.M. (2016). Rapid biological speciation driven by tectonic evolution in New Zealand. Nature Geoscience, 9, 140-144.

Davenport, J. & Davenport, J.L. (2007). Interaction of thermal tolerance and oxygen availability in the eurythermal gastropods Littorina littorea and Nucella lapillus. Marine Ecology Progress Series, 332, 167-170.

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic biology, 56, 879-886.

Dell, R. (1962). New Zealand marine provinces—do they exist. Tuatara, 10, 43-52.

Dohner, M. (2016). Crawling to connectivity? The direct-developing journey of the spotted whelk (Cominella maculosa).

Dohner, M., Phillips, N.E. & Ritchie, P.A. (2018). Fine-scale genetic structure across a New Zealand disjunction for the direct-developing intertidal whelk Cominella maculosa (Gastropoda: Buccinidae). Biological Journal of the Linnean Society, 123, 593-602.

Donald, K.M., Kennedy, M. & Spencer, H.G. (2005). The phylogeny and taxonomy of austral monodontine topshells (Mollusca: Gastropoda: Trochidae), inferred from DNA sequences. Molecular Phylogenetics and Evolution, 37, 474-483.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32, 1792-1797.

Excoffier, L., Laval, G. & Schneider, S. (2006). An integrated software package for population genetics data analysis. Computational and Molecular Population Genetics Lab (CMPG), Institute of Zoology, University of Berne, Switzerland.

Falini, G., Albeck, S., Weiner, S. & Addadi, L. (1996). Control of aragonite or calcite polymorphism by mollusk shell macromolecules. Science, 271, 67-69.

Fleming, A.M., Dohner, M.M., Phillips, N.E. & Ritchie, P.A. (2017). Genetic connectivity among populations of two congeneric direct-developing whelks varies across spatial scales. New Zealand Journal of Marine and Freshwater Research, 52, 100-117.

Fleming, C.A. (1975). The geological history of New Zealand and its biota. In: Biogeography and ecology in New Zealand. Springer, pp. 1-86.

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

Fox, A. (1977). Pa of the Auckland Isthmus: an archaeological analysis. Records of the Auckland Institute and Museum, 1-24.

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

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

Gardner, J., Bell, J., Constable, H., Hannan, D., Ritchie, P. & Zuccarello, G. (2010). Multi-species coastal marine connectivity: a literature review with recommendations for further research. NZ Aquatic Environment Biodiversity Report, 58, 1-47.

Ghiselli, F., Maurizii, M.G., Reunov, A., Ariño-Bassols, H., Cifaldi, C., Pecci, A. et al. (2019). Natural heteroplasmy and mitochondrial inheritance in bivalve molluscs. Integrative and comparative biology, 59, 1016-1032.

Gibbs, P. (1993). Phenotypic changes in the progeny of Nucella lapillus (Gastropoda) transplanted from an exposed shore to sheltered inlets. Journal of Molluscan Studies, 59, 187-194.

Gillespie, G.D. (1985). Hybridization, introgression, and morphometric differentiation between Mallard (Anas platyrhynchos) and Grey Duck (Anas superciliosa) in Otago, New Zealand. The Auk, 102, 459-469.

Grapes, R. (1999). Geomorphology of faulting: The Wairarapa Fault, New Zealand. Zeitschrift für Geomorphologie Supplement Volumes, 191-217.

Hare, M.P. (2001). Prospects for nuclear gene phylogeography. Trends in Ecology & Evolution, 16, 700-706.

Harpending, H. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human biology, 591-600.

Hasegawa, M., Kishino, H. & Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of molecular evolution, 22, 160-174.

Hay, C.H. (2020). Seashore uplift and the distribution of the bull kelp Durvillaea willana Lindauer in New Zealand. New Zealand Journal of Botany, 58, 94-117.

Heath, R. (1985). A review of the physical oceanography of the seas around New Zealand—1982. New Zealand journal of marine and freshwater research, 19, 79-124.

Hebert, P.D., Cywinska, A., Ball, S.L. & Dewaard, J.R. (2003a). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270, 313-321.

Hebert, P.D., Ratnasingham, S. & De Waard, J.R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270, S96-S99.

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

Hewitt, C.L., Gollasch, S. & Minchin, D. (2009). The vessel as a vector–biofouling, ballast water and sediments. In: Biological invasions in marine ecosystems. Springer, pp. 117-131.

Hills, S.F., Trewick, S.A. & Morgan-Richards, M. (2011). Phylogenetic information of genes, illustrated with mitochondrial data from a genus of gastropod molluscs. Biological Journal of the Linnean Society, 104, 770-785.

Hoeh, W.R., Blakley, K.H. & Brown, W.M. (1991). Heteroplasmy suggests limited biparental inheritance of Mytilus mitochondrial DNA. Science, 251, 1488-1490.

Hollis, C.J., Crundwell, M., Beu, A.G., Morgans, H., Crampton, J., Raine, J. et al. (2010). Calibration of the New Zealand Cretaceous-Cenozoic timescale to GTS2004.

Howarth, J.D., Fitzsimons, S.J., Norris, R.J., Langridge, R. & Vandergoes, M.J. (2016). A 2000 yr rupture history for the Alpine fault derived from Lake Ellery, South Island, New Zealand. Bulletin, 128, 627-643.

Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17, 754-755.

Hurst, G.D. & Jiggins, F.M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences, 272, 1525-1534.

Johnson, M. & Cumming, R. (1995). Genetic distinctness of three widespread and morphologically variable species of Drupella (Gastropoda, Muricidae). Coral Reefs, 14, 71-78.

Jones, M. & Ross, P. (2018). Recovery of the New Zealand muricid dogwhelk Haustrum scobina from TBT-induced imposex. Marine pollution bulletin, 126, 396-401.

Kassambara, A. & Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses, 76.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28, 1647-1649.

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

Kitching, J. & Lockwood, J. (1974). Observations on shell form and its ecological significance in thaisid gastropods of the genus Lepsiella in New Zealand. Marine Biology, 28, 131-144.

Komaru, A., Houki, S., Yamada, M., Miyake, T., Obata, M. & Kawamura, K. (2012). 28S rDNA haplotypes of males are distinct from those of androgenetic hermaphrodites in the clam Corbicula leana. Development genes and evolution, 222, 181-187.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35, 1547-1549.

Kyle, C. & Boulding, E. (2000). Comparative population genetic structure of marine gastropods (Littorina spp.) with and without pelagic larval dispersal. Marine Biology, 137, 835-845.

Laird, M. & Bradshaw, J. (2004). The break-up of a long-term relationship: the Cretaceous separation of New Zealand from Gondwana. Gondwana Research, 7, 273-286.

Leigh, J.W. & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. Methods in Ecology and Evolution, 6, 1110-1116.

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

Littlewood, D., Curini-Galletti, M. & Herniou, E.A. (2000). The interrelationships of Proseriata (Platyhelminthes: Seriata) tested with molecules and morphology. Molecular Phylogenetics and Evolution, 16, 449-466.

Logan, S. (2019). Phylogeography of the oyster borer (Haustrum scobina).

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

Luckens, P.A. (1975). Predation and intertidal zonation of barnacles at Leigh, New Zealand. New Zealand journal of marine and freshwater research, 9, 355-378.

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., Kelly, J. et al. (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. Ecology, 89, S108-S122.

Mallet, J. & Willmott, K. (2003). Taxonomy: renaissance or Tower of Babel? Trends in Ecology & Evolution, 18, 57-59.

Mayr, E. (1963). Animal species and evolution. Animal species and evolution.

McArthur, A.G. & Koop, B.F. (1999). Partial 28S rDNA sequences and the antiquity of hydrothermal vent endemic gastropods. Molecular Phylogenetics and Evolution, 13, 255-274.

McSaveney, M.J., Graham, I.J., Begg, J.G., Beu, A.G., Hull, A.G., Kim, K. et al. (2006). Late Holocene uplift of beach ridges at Turakirae Head, south Wellington coast, New Zealand. New Zealand Journal of Geology and Geophysics, 49, 337-358.

Miura, O., Torchin, M.E., Bermingham, E., Jacobs, D.K. & Hechinger, R.F. (2011). Flying shells: historical dispersal of marine snails across Central America. Proceedings of the Royal Society B: Biological Sciences, 279, 1061-1067.

Moffitt, J. (2019). The direct and indirect effects of acute thermal stress on Haustrum scobina: A thermal tolerance assessment of three intertidal invertebrates. ResearchSpace@ Auckland.

Moore, P. & Kenny, J.A. (1986). Exotic rocks on the northern New Zealand coast: a probable association with early shipwrecks. Tane, 31, 125-131.

Mortimer, N., Campbell, H.J., Tulloch, A.J., King, P.R., Stagpoole, V.M., Wood, R.A. et al. (2017). Zealandia: Earth's hidden continent. GSA today, 27, 27-35.

Morton, B. (1999). Competitive grazers and the predatory whelk Lepsiella flindersi (Gastropoda: Muricidae) structure a mussel bed (Xenostrobus pulex) on a southwest Australian shore. Journal of molluscan studies, 65, 435-452.

Morton, B. (2006). Scavenging behaviour by Ergalatax contractus (Gastropoda: Muricidae) and interactions with Nassarius nodifer (Gastropoda: Nassariidae) in the Cape d'Aguilar Marine Reserve, Hong Kong. Journal of the Marine Biological Association of the United Kingdom, 86, 141-152.

Morton, J.E. & Miller, M. (1973). The New Zealand Sea Shore. Collins.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular biology and evolution, 3, 418-426.

Newnham, R.M., Lowe, D.J. & Williams, P.W. (1999). Quaternary environmental change in New Zealand: a review. Progress in Physical Geography, 23, 567-610.

Novak, M. (2013). Trophic omnivory across a productivity gradient: intraguild predation theory and the structure and strength of species interactions. Proceedings of the Royal Society B: Biological Sciences, 280, 20131415.

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

Palmer, A.R. (1990). Effect of crab effluent and scent of damaged conspecifics on feeding, growth, and shell morphology of the Atlantic dogwhelk Nucella lapillus (L.). In: Progress in Littorinid and Muricid Biology. Springer, pp. 155-182.

Parvizi, E., Craw, D. & Waters, J.M. (2019). Kelp DNA records late Holocene paleoseismic uplift of coastline, southeastern New Zealand. Earth and Planetary Science Letters, 520, 18-25.

Pillans, B. (2017). Quaternary stratigraphy of Whanganui Basin—a globally significant archive. In: Landscape and quaternary environmental change in New Zealand. Springer, pp. 141-170.

Pinsky, M.L., Palumbi, S.R., Andréfouët, S. & Purkis, S.J. (2012). Open and closed seascapes: Where does habitat patchiness create populations with high fractions of self-recruitment? Ecological Applications, 22, 1257-1267.

Proctor, R. & Carter, L. (1989). Tidal and sedimentary response to the late Quaternary closure and opening of Cook Strait, New Zealand: results from numerical modeling. Paleoceanography, 4, 167-180.

Puritz, J.B., Keever, C.C., Addison, J.A., Barbosa, S.S., Byrne, M., Hart, M.W. et al. (2017). Life-history predicts past and present population connectivity in two sympatric sea stars. Ecology and evolution, 7, 3916-3930.

Quoy, J.C.R. & Gaimard, P. (1833) Voyage de découverts de l'Astrolabe, exécuté par ordre du Roi pendant les années 1826-1827-1828-1829, sous le commandement de M.J. Dumont d'Urville. Zoologie. Mollusques, 2(2), Tastu, Paris.

Rambaut, A. (2017). FigTree-version 1.4. 3, a graphical viewer of phylogenetic trees. Computer program distributed by the author, website: http://tree. bio. ed. ac. uk/software/figtree.

Rosenberg, G., Kuncio, R., Davis, G., Harasewych, M. & Tillier, S. (1994). Ribosomal RNA phylogeny of selected gastropod and unionacean bivalve mollusks. pp. 111-121.

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

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular biology and evolution, 34, 3299-3302.

Shaw, K.L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in

Hawaiian crickets. Proceedings of the National Academy of Sciences, 99, 16122-16127.

Sheppard, J. (2017). Breeding ecology and productivity of mallards and mallard-grey duck hybrids in New Zealand. University of Auckland Auckland, New Zealand.

Smith, P. & McVeagh, M. (1991). Widespread organotin pollution in New Zealand coastal waters as indicated by imposex in dogwhelks. Marine Pollution Bulletin, 22, 409-413.

Sokal, R.R. & Crovello, T.J. (1970). The biological species concept: a critical evaluation. The American Naturalist, 104, 127-153.

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

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

Stevens, K. & Vella, P. (1981). Palaeoclimatic interpretation of stable isotope ratios in molluscan fossils from middle Pleistocene marine strata, Wanganui, New Zealand. Palaeogeography, Palaeoclimatology, Palaeoecology, 34, 257-265.

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

Suggate, R.P. & Almond, P.C. (2005). The last glacial maximum (LGM) in western South Island, New Zealand: implications for the global LGM and MIS 2. Quaternary Science Reviews, 24, 1923-1940.

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

Tan, K. (2003). Phylogenetic analysis and taxonomy of some southern Australian and New Zealand Muricidae (Mollusca: Neogastropoda). Journal of Natural History, 37, 911-1028.

Taylor, J.D. & Reid, D.G. (1990). Shell microstructure and mineralogy of the Littorinidae: ecological and evolutionary significance. In: Progress in Littorinid and Muricid Biology. Springer, pp. 199-215.

Templeton, A.R. (1989). The meaning of species and speciation: a genetic perspective. The units of evolution: Essays on the nature of species, 159-183.

Teske, P.R., Von der Heyden, S., McQuaid, C.D. & Barker, N.P. (2011). A review of marine phylogeography in southern Africa. South African Journal of Science, 107, 43-53.

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

Vaïtilingon, D., Eeckhaut, I., Fourgon, D. & Jangoux, M. (2004). Population dynamics, infestation and host selection of Vexilla vexillum, an ectoparasitic muricid of echinoids, in Madagascar. Diseases of aquatic organisms, 61, 241-255.

Van Valen, L. (1971). Adaptive zones and the orders of mammals. Evolution, 420-428.

Van Valen, L. (1976). Ecological species, multispecies, and oaks. Taxon, 233-239.

Verry, A.J., Walton, K., Tuck, I.D. & Ritchie, P.A. (2020). Genetic structure and recent population expansion in the commercially harvested deep-sea decapod, Metanephrops challengeri (Crustacea: Decapoda). New Zealand Journal of Marine and Freshwater Research, 54, 251-270.

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

Walton, K. (2017). Phylogeography of the New Zealand whelks Cominella maculosa and C. virgata.

Walton, K., Marshall, B.A., Phillips, N.E., Verry, A.J. & Ritchie, P.A. (2018). Phylogeography of the New Zealand whelks Cominella maculosa and C. virgata (Gastropoda: Neogastropoda: Buccinoidea: Buccinidae). Biological Journal of the Linnean Society, 126, 178-202.

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

Wardle, P. (1963). Evolution and distribution of the New Zealand flora, as affected by Quaternary climates. New Zealand journal of botany, 1, 3-17.

Waters, J.M., Fraser, C.I. & Hewitt, G.M. (2013). Founder takes all: density-dependent processes structure biodiversity. Trends in ecology & evolution, 28, 78-85.

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

Wickham, H. & Wickham, M.H. (2007). The ggplot package.

Williams, S., Reid, D. & Littlewood, D. (2003). A molecular phylogeny of the Littorininae (Gastropoda: Littorinidae): unequal evolutionary rates, morphological parallelism, and biogeography of the Southern Ocean. Molecular phylogenetics and evolution, 28, 60-86.

Wilmshurstl, J.M. & Higham, T.F. (2004). Using rat-gnawed seeds to independently date the arrival of Pacific rats and humans in New Zealand. The Holocene, 14, 801-806.

Winkworth, R., Robertson, A., Ehrendorfer, F. & Lockhart, P. (1999). The importance of dispersal and recent speciation in the flora of New Zealand. Journal of Biogeography, 26, 1323-1325.

Wong, Y., Meier, R. & Tan, K. (2011). High haplotype variability in established Asian populations of the invasive Caribbean bivalve Mytilopsis sallei (Dreissenidae). Biological Invasions, 13, 341-348.

Wright, J.M. & Bentzen, P. (1995). Microsatellites: genetic markers for the future. In: Molecular genetics in fisheries. Springer, pp. 117-121.

Yadav, A., Jain, A., Sahu, J., Dubey, A., Gadpayle, R., Barwa, D.K. et al. (2019). An overview on species hybridization in animals. International Journal of Fauna and Biological Studies

Yao, S.-L., Luo, J.-J., Huang, G. & Wang, P. (2017). Distinct global warming rates tied to multiple ocean surface temperature changes. Nature Climate Change, 7, 486-491.

Zachos, F.E. (2016). Species concepts in biology. Springer.

Zou, S., Li, Q. & Kong, L. (2012). Multigene barcoding and phylogeny of geographically widespread muricids (Gastropoda: Neogastropoda) along the coast of China. Marine Biotechnology, 14, 21-34.

Appendix

Suplementary Figure 1 – *Haustrum scobina sensu lato* COI haplotype distribution key. 'Haplotype' is abbreviated to 'Hap'
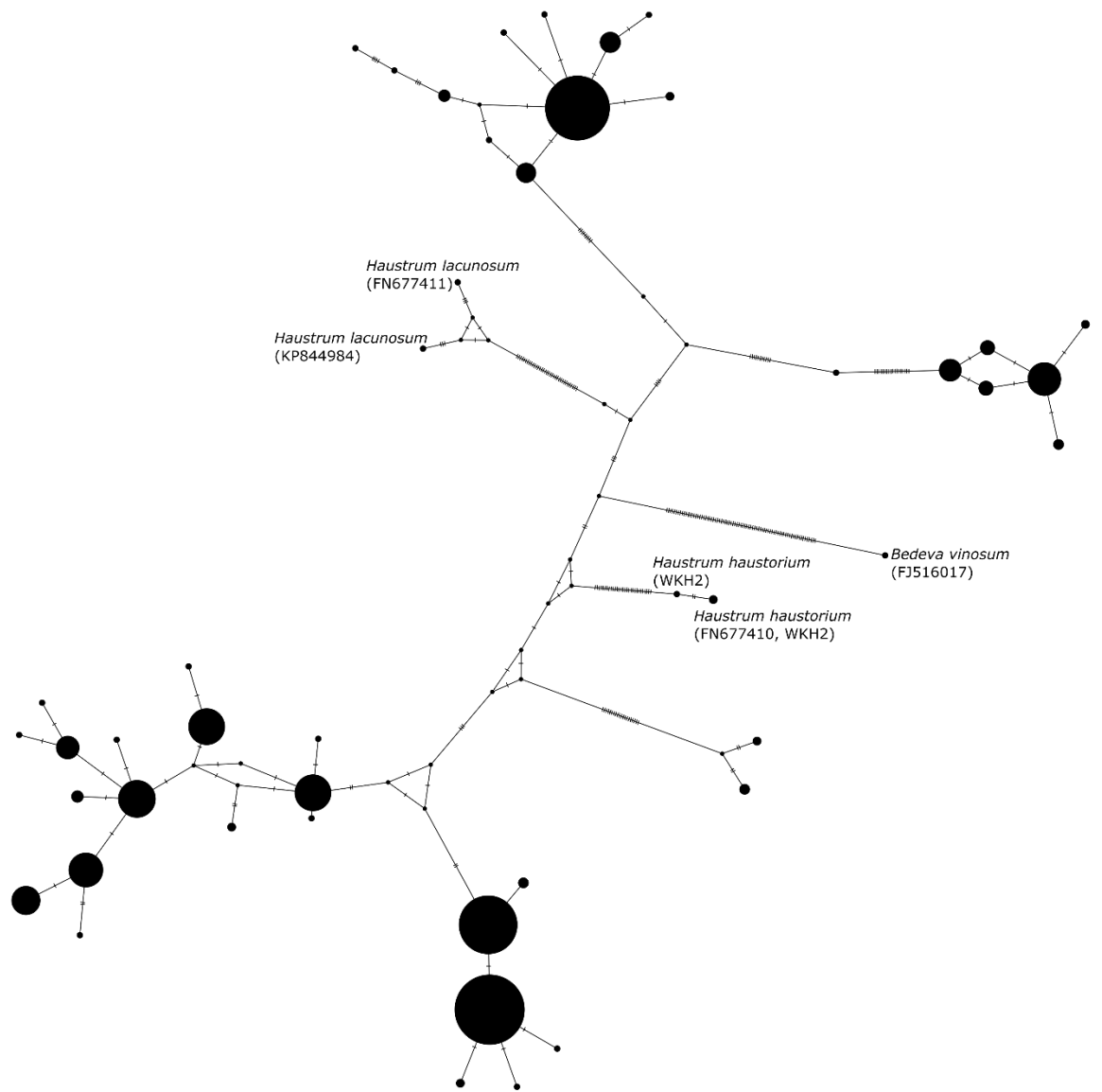
Suplementary Figure 2 - Median-joining network of all *Haustrum scobina sensu lato* COI sequences and outgroup species from Barco *et al.* (2015) and Ayre *et al.* (2009) with non-synonymous sites removed (585 bp). Hatch marks represent one mutational step. COI haplotypes are not labelled as they differ from those in the primary COI dataset, and this dataset is not discussed further.
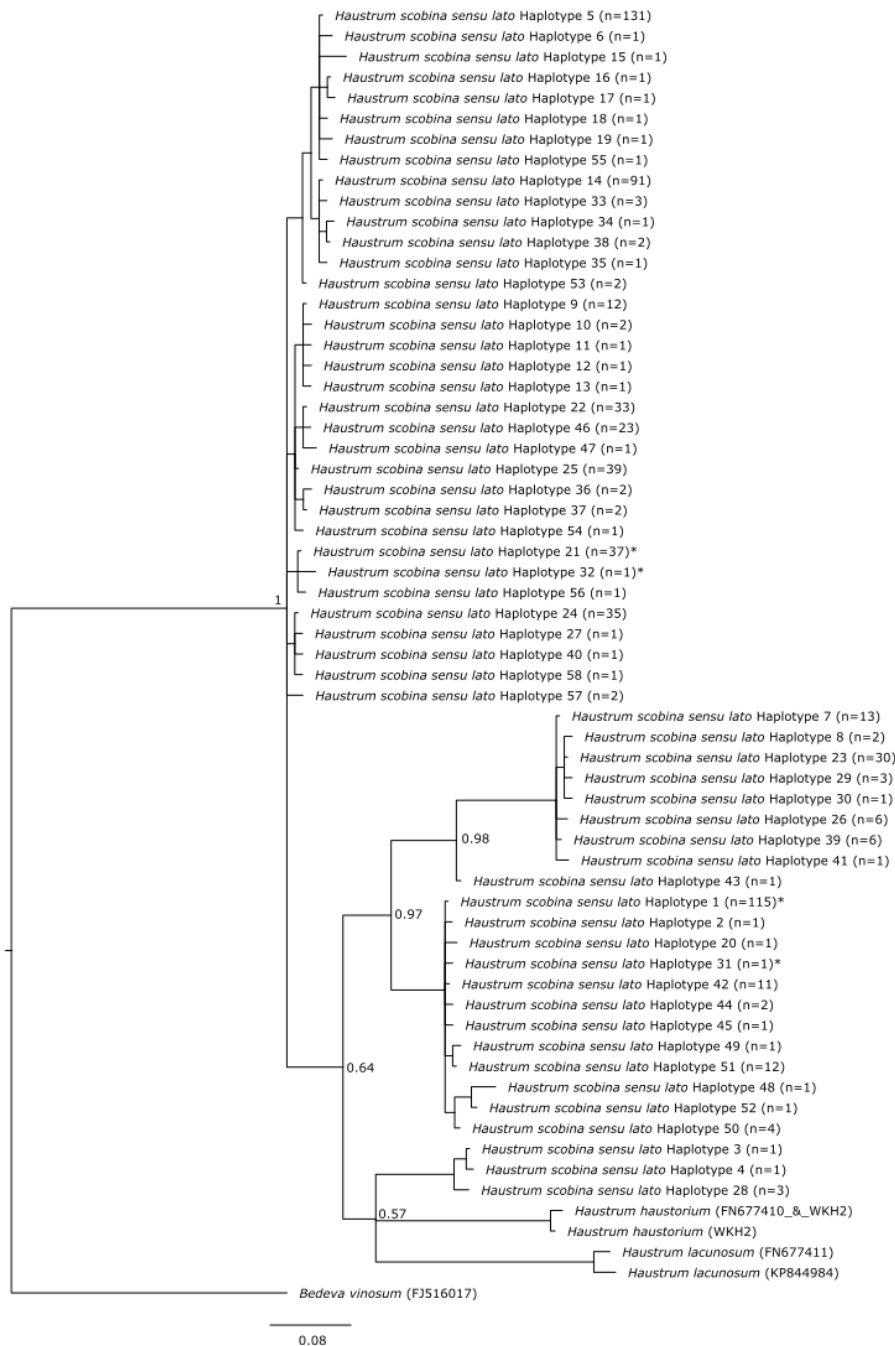
Suplementary Figure 3 - Bayesian tree of all *Haustrum scobina sensu lato* COI haplotypes and outgroup species. Where sequences were obtained from GenBank the accession number is shown beside them. Haplotypes marked by an asterisk include sequences from Barco *et al.* (2015). All others were produced as part of this research or by Logan (2019). Node values indicate posterior probabilities. The scale bar below the tree shows mutational distance.

Suplementary Figure 4 - Bayesian consensus tree of all *Haustrum scobina sensu stricto and Haustrum albomarginatum sensu stricto* 28S haplotypes. Haplotype 4 is marked by an asterisk and includes one sequence from Barco *et al*. (2010) (FN677467). Node labels indicate posterior probability. Outgroup sequences were taken GenBank where they had been deposited by Barco *et al*. (2010). All other sequences were generated in this thesis research.

Samples borrowed from the Te Papa Natural History Collection and sequenced as part of this research.

Species	Museum Registration Number	Location	Latitude (Decimal Degrees)	Longitude (E) (Decimal Degrees)	Code	Collected by
Haustrum scobina sensu lato	M.129326	Nelson	-41.200	173.326	NEL 1	Walton, Kerry
					NEL 2	
					NEL 3	
Haustrum scobina sensu lato	M.318156	Onehunga Bay	-41.091	174.854	ONE 1	Walton, Kerry
					ONE 2	
					ONE 3	
Haustrum scobina sensu lato	M.306340/306339	Sunset Beach	-37.402	174.705	SSB 1	Keil, Fred
					SSB 2	
					SSB 3	
					SSB 4	
					SSB 5	
Haustrum scobina sensu lato	M.315671	Chatham Island's	-43.945	176.563	CHI 1	Walton, Kerry
					CHI 2	
					CHI 3	
Haustrum scobina sensu lato	M.316768	Whangapoua Beach	-36.702	175.610	WHB 1	Marshall, Bruce with Plimmer, Kristelle
					WHB 2	

Haustrum scobina sensu lato	M.316155	Mahia Peninsula	-39.088	177.956	MAP 1	
					MAP 2	
					MAP 3	
Haustrum scobina sensu lato	M.326589	Waihau Bay	-37.615	177.910	WAH 1	Walton, Kerry with Scott, Eric & Scott, Heather
					WAH 2	
Haustrum scobina sensu lato	M.326577/1	Waikokopu	-39.071	177.829	WKH 1	Walton, Kerry with Scott, Eric & Scott, Heather
					WKH 2	
Haustrum scobina sensu lato	M.146000	Albion Inlet	N/A	N/A	ALL 1	N/A
Haustrum scobina sensu lato	M.321173/1	Ward Beach	-41.843	174.187	WDL 1	Marshall, Bruce with Walton, Kerry
					WDL 2	
Haustrum scobina sensu lato	M.085148/1	Ringaringa Beach	-46.905	168.145	RRL 1	Raven, Jenny

Supplementary Table 2 – Pairwise differences between COI haplotypes by mutational distance (below), with standard deviation above. 'Haplotype is abbreviated to 'Hap'.

	Hap_1	Hap_2	Hap_3	Hap_4	Hap_5	Hap_6	Hap_7	Hap_8	Hap_9	Hap_10
Hap_1		1.4013	5.0713	5.0452	4.671	4.7635	5.1177	5.1381	4.7184	4.7635
Hap_2	2		5.0713	5.0452	4.671	4.7635	5.1177	5.1381	4.7184	4.7635
Hap_3	41	41		0.9954	4.7184	4.8062	5.1734	5.1566	4.671	4.7184
Hap_4	40	40	1		4.7635	4.8467	5.1566	5.1381	4.6211	4.671
Hap_5	30	30	31	32		1.4013	5.0172	4.9873	2.7236	2.7236
Hap_6	32	32	33	34	2		5.0713	5.0452	3.0151	3.0151
Hap_7	43	43	46	45	39	41		1.7083	4.9873	4.9873
Hap_8	44	44	45	44	38	40	3		4.9553	4.9553
Hap_9	31	31	30	29	8	10	38	37		0.9954
Hap_10	32	32	31	30	8	10	38	37	1	
Hap_11	32	32	31	30	9	11	39	38	1	2
Hap_12	32	32	31	30	9	11	39	38	1	2
Hap_13	32	32	31	30	9	11	39	38	1	2
Hap_14	30	30	31	32	2	4	39	38	8	9
Hap_15	35	35	36	37	5	7	43	42	13	13
Hap_16	31	31	32	33	1	3	40	39	9	9
Hap_17	32	32	33	34	2	4	41	40	10	10
Hap_18	31	31	31	32	1	3	40	39	9	9
Hap_19	32	32	33	34	2	4	41	40	10	10
Hap_20	2	4	41	40	32	34	45	46	31	32
Hap_21	28	28	31	30	7	9	37	36	5	6
Hap_22	31	31	32	31	8	10	40	39	2	3
Hap_23	43	43	44	43	37	39	2	1	36	36
Hap_24	30	30	31	30	7	9	41	40	3	4
Hap_25	30	30	31	30	7	9	39	38	1	2
Hap_26	45	45	44	43	39	41	2	3	38	38
Hap_27	31	31	32	31	8	10	42	41	4	5
Hap_28	41	41	4	5	31	33	48	47	32	33
Hap_29	42	42	43	42	36	38	3	2	35	35
Hap_30	44	44	45	44	38	40	3	2	37	37
Hap_31	1	3	40	39	31	33	44	45	32	33
Hap_32	31	31	34	33	10	12	40	39	8	9
Hap_33	30	30	32	33	3	5	40	39	9	10
Hap_34	32	32	33	34	4	6	41	40	10	11
Hap_35	31	31	32	33	3	5	40	39	9	10
Hap_36	32	32	33	32	9	9	39	38	3	4
Hap_37	31	31	32	31	8	8	40	39	2	3
Hap_38	31	31	32	33	3	5	40	39	9	10
Hap_39	42	42	45	44	38	40	1	2	37	37
Hap_40	31	31	32	31	8	10	42	41	4	5
Hap_41	45	45	48	47	41	43	2	5	40	40
Hap_42	1	1	40	39	29	31	42	43	30	31
Hap_43	23	23	44	43	31	33	20	23	32	32
Hap_44	1	3	42	41	31	33	44	45	32	33
Hap_45	1	3	42	41	31	33	44	45	32	33
Hap_46	32	32	33	32	9	11	41	40	3	4
Hap_47	31	31	34	33	10	12	42	41	4	5
Hap_48	6	8	45	44	34	36	47	48	35	36
Hap_49	2	4	43	42	32	34	43	44	33	34
Hap_50	2	2	43	42	32	34	43	44	33	34
Hap_51	1	3	42	41	31	33	44	45	32	33
Hap_52	5	5	44	43	33	35	44	45	34	35
Hap_53	28	28	31	30	2	4	39	38	6	7
Hap_54	31	31	32	31	8	10	40	39	2	3
Hap_55	31	31	32	33	1	3	40	39	9	9
Hap_56	29	29	32	31	8	8	38	37	6	7
Hap_57	30	30	31	30	9	11	41	40	5	6
Hap_58	31	31	32	31	8	10	40	39	4	5

	Hap_11	Hap_12	Hap_13	Hap_14	Hap_15	Hap_16	Hap_17	Hap_18	Hap_19	Hap_20
Hap_1	4.7635	4.7635	4.7635	4.671	4.885	4.7184	4.7635	4.7184	4.7635	1.4013
Hap_2	4.7635	4.7635	4.7635	4.671	4.885	4.7184	4.7635	4.7184	4.7635	1.9633
Hap_3	4.7184	4.7184	4.7184	4.7184	4.9212	4.7635	4.8062	4.7184	4.8062	5.0713
Hap_4	4.671	4.671	4.671	4.7635	4.9553	4.8062	4.8467	4.7635	4.8467	5.0452
Hap_5	2.8747	2.8747	2.8747	1.4013	2.1847	0.9954	1.4013	0.9954	1.4013	4.7635
Hap_6	3.1464	3.1464	3.1464	1.9633	2.5602	1.7083	1.9633	1.7083	1.9633	4.8467
Hap_7	5.0172	5.0172	5.0172	5.0172	5.1177	5.0452	5.0713	5.0452	5.0713	5.1566
Hap_8	4.9873	4.9873	4.9873	4.9873	5.0955	5.0172	5.0452	5.0172	5.0452	5.1734
Hap_9	0.9954	0.9954	0.9954	2.7236	3.3858	2.8747	3.0151	2.8747	3.0151	4.7184
Hap_10	1.4013	1.4013	1.4013	2.8747	3.3858	2.8747	3.0151	2.8747	3.0151	4.7635
Hap_11		1.4013	1.4013	2.8747	3.4955	3.0151	3.1464	3.0151	3.1464	4.7635
Hap_12	2		1.4013	2.8747	3.4955	3.0151	3.1464	3.0151	3.1464	4.7635
Hap_13	2	2		2.8747	3.4955	3.0151	3.1464	3.0151	3.1464	4.7635
Hap_14	9	9	9		2.5602	1.7083	1.9633	1.7083	1.9633	4.7635
Hap_15	14	14	14	7		2.3817	2.5602	2.3817	2.5602	4.9553
Hap_16	10	10	10	3	6		0.9954	1.4013	1.7083	4.8062
Hap_17	11	11	11	4	7	1		1.7083	1.9633	4.8467
Hap_18	10	10	10	3	6	2	3		1.7083	4.8062
Hap_19	11	11	11	4	7	3	4	3		4.8467
Hap_20	32	32	32	32	37	33	34	33	34	
Hap_21	6	6	6	7	11	8	9	8	9	30
Hap_22	3	3	3	8	13	9	10	9	10	31
Hap_23	37	37	37	37	41	38	39	38	39	45
Hap_24	4	4	4	7	12	8	9	8	9	30
Hap_25	2	2	2	7	12	8	9	8	9	30
Hap_26	39	39	39	39	43	40	41	39	41	47
Hap_27	5	5	5	8	13	9	10	9	10	31
Hap_28	33	33	31	31	36	32	33	31	33	41
Hap_29	36	36	36	36	40	37	38	37	38	44
Hap_30	38	38	38	38	42	39	40	39	40	46
Hap_31	33	33	33	31	36	32	33	31	33	3
Hap_32	9	9	9	10	14	9	8	11	12	33
Hap_33	10	10	10	1	8	4	5	4	5	32
Hap_34	11	11	11	2	9	5	6	5	4	34
Hap_35	10	10	10	1	8	4	5	4	5	33
Hap_36	4	4	4	9	13	10	11	10	11	32
Hap_37	3	3	3	8	13	9	10	9	10	31
Hap_38	10	10	10	1	8	4	5	4	3	33
Hap_39	38	38	38	38	42	39	40	39	40	44
Hap_40	5	5	5	8	13	9	10	9	10	31
Hap_41	41	41	41	41	45	42	43	42	43	47
Hap_42	31	31	31	29	34	30	31	30	31	3
Hap_43	33	33	33	31	36	32	33	32	33	25
Hap_44	33	33	33	31	36	32	33	32	33	3
Hap_45	33	33	33	31	36	32	33	32	33	3
Hap_46	4	4	4	9	14	10	11	10	11	32
Hap_47	4	5	5	10	15	11	12	11	12	31
Hap_48	35	36	36	34	39	35	36	35	36	8
Hap_49	34	34	34	30	37	33	34	33	34	4
Hap_50	34	34	34	32	37	33	34	33	34	4
Hap_51	33	33	33	31	36	32	33	32	33	3
Hap_52	34	35	35	33	38	34	35	34	35	7
Hap_53	7	7	7	2	7	3	4	3	4	30
Hap_54	3	3	3	8	13	9	10	9	10	31
Hap_55	10	10	10	3	6	2	3	2	3	33
Hap_56	7	7	7	8	12	9	10	9	10	31
Hap_57	6	6	6	9	14	10	11	10	11	30
Hap_58	5	5	5	8	13	9	10	9	10	31

	Hap_21	Hap_22	Hap_23	Hap_24	Hap_25	Hap_26	Hap_27	Hap_28	Hap_29	Hap_30
Hap_1	4.5687	4.7184	5.1177	4.671	4.671	5.1566	4.7184	5.0713	5.0955	5.1381
Hap_2	4.5687	4.7184	5.1177	4.671	4.671	5.1566	4.7184	5.0713	5.0955	5.1381
Hap_3	4.7184	4.7635	5.1381	4.7184	4.7184	5.1381	4.7635	1.9633	5.1177	5.1566
Hap_4	4.671	4.7184	5.1177	4.671	4.671	5.1177	4.7184	2.1847	5.0955	5.1381
Hap_5	2.5602	2.7236	4.9553	2.5602	2.5602	5.0172	2.7236	4.7184	4.9212	4.9873
Hap_6	2.8747	3.0151	5.0172	2.8747	2.8747	5.0713	3.0151	4.8062	4.9873	5.0452
Hap_7	4.9553	5.0452	1.4013	5.0713	5.0172	1.4013	5.0955	5.2014	1.7083	1.7083
Hap_8	4.9212	5.0172	0.9954	5.0452	4.9873	1.7083	5.0713	5.1883	1.4013	1.4013
Hap_9	2.1847	1.4013	4.9212	1.7083	0.9954	4.9873	1.9633	4.7635	4.885	4.9553
Hap_10	2.3817	1.7083	4.9212	1.9633	1.4013	4.9873	2.1847	4.8062	4.885	4.9553
Hap_11	2.3817	1.7083	4.9553	1.9633	1.4013	5.0172	2.1847	4.8062	4.9212	4.9873
Hap_12	2.3817	1.7083	4.9553	1.9633	1.4013	5.0172	2.1847	4.8062	4.9212	4.9873
Hap_13	2.3817	1.7083	4.9553	1.9633	1.4013	5.0172	2.1847	4.7184	4.9212	4.9873
Hap_14	2.5602	2.7236	4.9553	2.5602	2.5602	5.0172	2.7236	4.7184	4.9212	4.9873
Hap_15	3.1464	3.3858	5.0713	3.2697	3.2697	5.1177	3.3858	4.9212	5.0452	5.0955
Hap_16	2.7236	2.8747	4.9873	2.7236	2.7236	5.0452	2.8747	4.7635	4.9553	5.0172
Hap_17	2.8747	3.0151	5.0172	2.8747	2.8747	5.0713	3.0151	4.8062	4.9873	5.0452
Hap_18	2.7236	2.8747	4.9873	2.7236	2.7236	5.0172	2.8747	4.7184	4.9553	5.0172
Hap_19	2.8747	3.0151	5.0172	2.8747	2.8747	5.0713	3.0151	4.8062	4.9873	5.0452
Hap_20	4.671	4.7184	5.1566	4.671	4.671	5.1883	4.7184	5.0713	5.1381	5.1734
Hap_21		2.1847	4.885	1.9633	1.9633	4.9553	2.1847	4.7184	4.8467	4.9212
Hap_22	5		4.9873	1.7083	0.9954	5.0452	1.9633	4.7635	4.9553	5.0172
Hap_23	35	38		5.0172	4.9553	1.4013	5.0452	5.1734	0.9954	0.9954
Hap_24	4	3	39		1.4013	5.0713	0.9954	4.7184	4.9873	5.0452
Hap_25	4	1	37	2		5.0172	1.7083	4.7184	4.9212	4.9873
Hap_26	37	40	2	41	39		5.0955	5.1734	1.7083	1.7083
Hap_27	5	4	40	1	3	42		4.7635	5.0172	5.0172
Hap_28	31	32	46	31	31	46	32		5.1566	5.1883
Hap_29	34	37	1	38	36	3	39	45		1.4013
Hap_30	36	39	1	40	38	3	39	47	2	
Hap_31	29	32	44	31	31	44	32	40	43	45
Hap_32	3	8	38	7	7	40	8	34	37	39
Hap_33	8	9	38	8	8	40	9	32	37	39
Hap_34	9	10	39	9	9	41	10	33	38	40
Hap_35	8	9	38	8	8	40	9	32	37	39
Hap_36	4	3	37	4	2	39	5	33	36	38
Hap_37	5	2	38	3	1	40	4	32	37	39
Hap_38	8	9	38	8	8	40	9	32	37	39
Hap_39	36	39	1	40	38	3	41	47	2	2
Hap_40	5	4	40	1	3	42	2	32	39	41
Hap_41	39	42	4	43	41	4	44	50	5	5
Hap_42	27	30	42	29	29	44	30	40	41	43
Hap_43	31	34	22	33	33	22	34	46	21	23
Hap_44	29	32	44	31	31	46	32	42	43	45
Hap_45	29	32	44	31	31	46	32	42	43	45
Hap_46	6	1	39	4	2	41	5	33	38	40
Hap_47	7	2	40	5	3	42	6	34	39	41
Hap_48	32	35	47	34	34	49	35	45	46	48
Hap_49	30	33	43	32	32	45	33	43	42	44
Hap_50	30	33	43	32	32	45	33	43	42	44
Hap_51	29	32	44	31	31	46	32	42	43	45
Hap_52	31	34	44	33	33	46	34	44	43	45
Hap_53	5	6	37	5	5	39	6	31	36	38
Hap_54	5	2	38	3	1	40	4	32	37	39
Hap_55	8	9	38	8	8	40	9	32	37	39
Hap_56	1	6	36	5	5	38	6	32	35	37
Hap_57	4	5	39	4	4	41	5	31	38	40
Hap_58	5	4	38	1	3	40	2	32	37	39

	Hap_31	Hap_32	Hap_33	Hap_34	Hap_35	Hap_36	Hap_37	Hap_38	Hap_39	Hap_40
Hap_1	0.9954	4.7184	4.671	4.7635	4.7184	4.7635	4.7184	4.7184	5.0955	4.7184
Hap_2	1.7083	4.7184	4.671	4.7635	4.7184	4.7635	4.7184	4.7184	5.0955	4.7184
Hap_3	5.0452	4.8467	4.7635	4.8062	4.7635	4.8062	4.7635	4.7635	5.1566	4.7635
Hap_4	5.0172	4.8062	4.8062	4.8467	4.8062	4.7635	4.7184	4.8062	5.1381	4.7184
Hap_5	4.7184	3.0151	1.7083	1.9633	1.7083	2.8747	2.7236	1.7083	4.9873	2.7236
Hap_6	4.8062	3.2697	2.1847	2.3817	2.1847	2.8747	2.7236	2.1847	5.0452	3.0151
Hap_7	5.1381	5.0452	5.0452	5.0713	5.0452	5.0172	5.0452	5.0452	0.9954	5.0955
Hap_8	5.1566	5.0172	5.0172	5.0452	5.0172	4.9873	5.0172	5.0172	1.4013	5.0713
Hap_9	4.7635	2.7236	2.8747	3.0151	2.8747	1.7083	1.4013	2.8747	4.9553	1.9633
Hap_10	4.8062	2.8747	3.0151	3.1464	3.0151	1.9633	1.7083	3.0151	4.9553	2.1847
Hap_11	4.8062	2.8747	3.0151	3.1464	3.0151	1.9633	1.7083	3.0151	4.9873	2.1847
Hap_12	4.8062	2.8747	3.0151	3.1464	3.0151	1.9633	1.7083	3.0151	4.9873	2.1847
Hap_13	4.8062	2.8747	3.0151	3.1464	3.0151	1.9633	1.7083	3.0151	4.9873	2.1847
Hap_14	4.7184	3.0151	0.9954	1.4013	0.9954	2.8747	2.7236	0.9954	4.9873	2.7236
Hap_15	4.9212	3.4955	2.7236	2.8747	2.7236	3.3858	3.3858	2.7236	5.0955	3.3858
Hap_16	4.7635	2.8747	1.9633	2.1847	1.9633	3.0151	2.8747	1.9633	5.0172	2.8747
Hap_17	4.8062	2.7236	2.1847	2.3817	2.1847	3.1464	3.0151	2.1847	5.0452	3.0151
Hap_18	4.7184	3.1464	1.9633	2.1847	1.9633	3.0151	2.8747	1.9633	5.0172	2.8747
Hap_19	4.8062	3.2697	2.1847	1.9633	2.1847	3.1464	3.0151	1.7083	5.0452	3.0151
Hap_20	1.7083	4.8062	4.7635	4.8467	4.8062	4.7635	4.7184	4.8062	5.1381	4.7184
Hap_21	4.6211	1.7083	2.7236	2.8747	2.7236	1.9633	2.1847	2.7236	4.9212	2.1847
Hap_22	4.7635	2.7236	2.8747	3.0151	2.8747	1.7083	1.4013	2.8747	5.0172	1.9633
Hap_23	5.1381	4.9873	4.9873	5.0172	4.9873	4.9553	4.9873	4.9873	0.9954	5.0452
Hap_24	4.7184	2.5602	2.7236	2.8747	2.7236	1.9633	1.7083	2.7236	5.0452	0.9954
Hap_25	4.7184	2.5602	2.7236	2.8747	2.7236	1.4013	0.9954	2.7236	4.9873	1.7083
Hap_26	5.1381	5.0452	5.0452	5.0713	5.0452	5.0172	5.0452	5.0452	1.7083	5.0955
Hap_27	4.7635	2.7236	2.8747	3.0151	2.8747	2.1847	1.9633	2.8747	5.0713	1.4013
Hap_28	5.0452	4.8467	4.7635	4.8062	4.7635	4.8062	4.7635	4.7635	5.1883	4.7635
Hap_29	5.1177	4.9553	4.9553	4.9873	4.9553	4.9212	4.9553	4.9553	1.4013	5.0172
Hap_30	5.1566	5.0172	5.0172	5.0452	5.0172	4.9873	5.0172	5.0172	1.4013	5.0713
Hap_31		4.7635	4.7184	4.8062	4.7635	4.8062	4.7635	4.7635	5.1177	4.7635
Hap_32	32		3.1464	3.2697	3.1464	2.5602	2.7236	3.1464	5.0172	2.7236
Hap_33	31	11		1.7083	1.4013	3.0151	2.8747	1.4013	5.0172	2.8747
Hap_34	33	12	3		1.7083	3.1464	3.0151	0.9954	5.0452	3.0151
Hap_35	32	11	2	3		3.0151	2.8747	1.4013	5.0172	2.8747
Hap_36	33	7	10	11	10		0.9954	3.0151	4.9873	2.1847
Hap_37	32	8	9	10	9	1		2.8747	5.0172	1.9633
Hap_38	32	11	2	1	2	10	9		5.0172	2.8747
Hap_39	43	39	39	40	39	38	39	39		5.0713
Hap_40	32	8	9	10	9	5	4	9	41	
Hap_41	46	42	42	43	42	41	42	42	3	44
Hap_42	2	30	29	31	30	31	30	30	41	30
Hap_43	24	34	32	33	32	35	34	32	21	34
Hap_44	2	32	31	33	32	33	32	32	43	32
Hap_45	2	32	31	33	32	33	32	32	43	32
Hap_46	33	9	10	11	10	4	3	10	40	5
Hap_47	32	10	11	12	11	5	4	11	41	6
Hap_48	7	35	34	36	35	36	35	35	46	35
Hap_49	3	33	30	32	31	34	33	31	42	33
Hap_50	3	33	32	34	33	34	33	33	42	33
Hap_51	2	32	31	33	32	33	32	32	43	32
Hap_52	6	34	33	35	34	35	34	34	43	34
Hap_53	29	8	3	4	3	7	6	3	38	6
Hap_54	32	8	9	10	9	3	2	9	39	4
Hap_55	32	11	4	5	4	10	9	4	39	9
Hap_56	30	4	9	10	9	3	4	9	37	6
Hap_57	31	7	10	11	10	6	5	10	40	5
Hap_58	32	8	9	10	9	5	4	9	39	2

	Hap_41	Hap_42	Hap_43	Hap_44	Hap_45	Hap_46	Hap_47	Hap_48	Hap_49	Hap_50
Hap_1	5.1566	0.9954	4.2651	0.9954	0.9954	4.7635	4.7184	2.3817	1.4013	1.4013
Hap_2	5.1566	0.9954	4.2651	1.7083	1.7083	4.7635	4.7184	2.7236	1.9633	1.4013
Hap_3	5.2014	5.0452	5.1381	5.0955	5.0955	4.8062	4.8467	5.1566	5.1177	5.1177
Hap_4	5.1883	5.0172	5.1177	5.0713	5.0713	4.7635	4.8062	5.1381	5.0955	5.0955
Hap_5	5.0713	4.6211	4.7184	4.7184	4.7184	2.8747	3.0151	4.8467	4.7635	4.7635
Hap_6	5.1177	4.7184	4.8062	4.8062	4.8062	3.1464	3.2697	4.9212	4.8467	4.8467
Hap_7	1.4013	5.0955	4.0452	5.1381	5.1381	5.0713	5.0955	5.1883	5.1177	5.1177
Hap_8	2.1847	5.1177	4.2651	5.1566	5.1566	5.0452	5.0713	5.2014	5.1381	5.1381
Hap_9	5.0452	4.671	4.7635	4.7635	4.7635	1.7083	1.9633	4.885	4.8062	4.8062
Hap_10	5.0452	4.7184	4.7635	4.8062	4.8062	1.9633	2.1847	4.9212	4.8467	4.8467
Hap_11	5.0713	4.7184	4.8062	4.8062	4.8062	1.9633	1.9633	4.885	4.8467	4.8467
Hap_12	5.0713	4.7184	4.8062	4.8062	4.8062	1.9633	2.1847	4.9212	4.8467	4.8467
Hap_13	5.0713	4.7184	4.8062	4.8062	4.8062	1.9633	2.1847	4.9212	4.8467	4.8467
Hap_14	5.0713	4.6211	4.7184	4.7184	4.7184	2.8747	3.0151	4.8467	4.671	4.7635
Hap_15	5.1566	4.8467	4.9212	4.9212	4.9212	3.4955	3.5992	5.0172	4.9553	4.9553
Hap_16	5.0955	4.671	4.7635	4.7635	4.7635	3.0151	3.1464	4.885	4.8062	4.8062
Hap_17	5.1177	4.7184	4.8062	4.8062	4.8062	3.1464	3.2697	4.9212	4.8467	4.8467
Hap_18	5.0955	4.671	4.7635	4.7635	4.7635	3.0151	3.1464	4.885	4.8062	4.8062
Hap_19	5.1177	4.7184	4.8062	4.8062	4.8062	3.1464	3.2697	4.9212	4.8467	4.8467
Hap_20	5.1883	1.7083	4.3952	1.7083	1.7083	4.7635	4.7184	2.7236	1.9633	1.9633
Hap_21	5.0172	4.5136	4.7184	4.6211	4.6211	2.3817	2.5602	4.7635	4.671	4.671
Hap_22	5.0955	4.671	4.8467	4.7635	4.7635	0.9954	1.4013	4.885	4.8062	4.8062
Hap_23	1.9633	5.0955	4.1952	5.1381	5.1381	5.0172	5.0452	5.1883	5.1177	5.1177
Hap_24	5.1177	4.6211	4.8062	4.7184	4.7184	1.9633	2.1847	4.8467	4.7635	4.7635
Hap_25	5.0713	4.6211	4.8062	4.7184	4.7184	1.4013	1.7083	4.8467	4.7635	4.7635
Hap_26	1.9633	5.1381	4.1952	5.1734	5.1734	5.0713	5.0955	5.2127	5.1566	5.1566
Hap_27	5.1381	4.671	4.8467	4.7635	4.7635	2.1847	2.3817	4.885	4.8062	4.8062
Hap_28	5.2223	5.0452	5.1734	5.0955	5.0955	4.8062	4.8467	5.1566	5.1177	5.1177
Hap_29	2.1847	5.0713	4.122	5.1177	5.1177	4.9873	5.0172	5.1734	5.0955	5.0955
Hap_30	2.1847	5.1177	4.2651	5.1566	5.1566	5.0452	5.0713	5.2014	5.1381	5.1381
Hap_31	5.1734	1.4013	4.3317	1.4013	1.4013	4.8062	4.7635	2.5602	1.7083	1.7083
Hap_32	5.0955	4.671	4.8467	4.7635	4.7635	2.8747	3.0151	4.885	4.8062	4.8062
Hap_33	5.0955	4.6211	4.7635	4.7184	4.7184	3.0151	3.1464	4.8467	4.671	4.7635
Hap_34	5.1177	4.7184	4.8062	4.8062	4.8062	3.1464	3.2697	4.9212	4.7635	4.8467
Hap_35	5.0955	4.671	4.7635	4.7635	4.7635	3.0151	3.1464	4.885	4.7184	4.8062
Hap_36	5.0713	4.7184	4.885	4.8062	4.8062	1.9633	2.1847	4.9212	4.8467	4.8467
Hap_37	5.0955	4.671	4.8467	4.7635	4.7635	1.7083	1.9633	4.885	4.8062	4.8062
Hap_38	5.0955	4.671	4.7635	4.7635	4.7635	3.0151	3.1464	4.885	4.7184	4.8062
Hap_39	1.7083	5.0713	4.122	5.1177	5.1177	5.0452	5.0713	5.1734	5.0955	5.0955
Hap_40	5.1381	4.671	4.8467	4.7635	4.7635	2.1847	2.3817	4.885	4.8062	4.8062
Hap_41		5.1381	4.1952	5.1734	5.1734	5.1177	5.1381	5.2127	5.1566	5.1566
Hap_42	44		4.1952	1.4013	1.4013	4.7184	4.671	2.5602	1.7083	1.7083
Hap_43	22	22		4.3317	4.3317	4.885	4.8467	4.6211	4.2651	4.2651
Hap_44	46	2	24		1.4013	4.8062	4.7635	2.5602	1.7083	1.7083
Hap_45	46	2	24	2		4.8062	4.7635	2.3817	1.7083	1.7083
Hap_46	43	31	35	33	33		1.7083	4.9212	4.8467	4.8467
Hap_47	44	30	34	32	32	3		4.9212	4.8062	4.8062
Hap_48	49	7	29	7	6	36	36		2.7236	2.7236
Hap_49	45	3	23	3	3	34	33	8		1.9633
Hap_50	45	3	23	3	3	34	33	8	4	
Hap_51	46	2	24	2	2	33	32	7	1	3
Hap_52	46	6	26	6	5	35	35	5	7	3
Hap_53	41	27	31	29	29	7	8	32	30	30
Hap_54	42	30	34	32	32	3	4	35	33	33
Hap_55	42	30	32	32	32	10	11	35	33	33
Hap_56	40	28	32	30	30	7	8	33	31	31
Hap_57	43	29	33	31	31	6	7	34	32	32
Hap_58	42	30	32	32	32	5	6	35	33	33

	Hap_51	Hap_52	Hap_53	Hap_54	Hap_55	Hap_56	Hap_57	Hap_58
Hap_1	0.9954	2.1847	4.5687	4.7184	4.7184	4.6211	4.671	4.7184
Hap_2	1.7083	2.1847	4.5687	4.7184	4.7184	4.6211	4.671	4.7184
Hap_3	5.0955	5.1381	4.7184	4.7635	4.7635	4.7635	4.7184	4.7635
Hap_4	5.0713	5.1177	4.671	4.7184	4.8062	4.7184	4.671	4.7184
Hap_5	4.7184	4.8062	1.4013	2.7236	0.9954	2.7236	2.8747	2.7236
Hap_6	4.8062	4.885	1.9633	3.0151	1.7083	2.7236	3.1464	3.0151
Hap_7	5.1381	5.1381	5.0172	5.0452	5.0452	4.9873	5.0713	5.0452
Hap_8	5.1566	5.1566	4.9873	5.0172	5.0172	4.9553	5.0452	5.0172
Hap_9	4.7635	4.8467	2.3817	1.4013	2.8747	2.3817	2.1847	1.9633
Hap_10	4.8062	4.885	2.5602	1.7083	2.8747	2.5602	2.3817	2.1847
Hap_11	4.8062	4.8467	2.5602	1.7083	3.0151	2.5602	2.3817	2.1847
Hap_12	4.8062	4.885	2.5602	1.7083	3.0151	2.5602	2.3817	2.1847
Hap_13	4.8062	4.885	2.5602	1.7083	3.0151	2.5602	2.3817	2.1847
Hap_14	4.7184	4.8062	1.4013	2.7236	1.7083	2.7236	2.8747	2.7236
Hap_15	4.9212	4.9873	2.5602	3.3858	2.3817	3.2697	3.4955	3.3858
Hap_16	4.7635	4.8467	1.7083	2.8747	1.4013	2.8747	3.0151	2.8747
Hap_17	4.8062	4.885	1.9633	3.0151	1.7083	3.0151	3.1464	3.0151
Hap_18	4.7635	4.8467	1.7083	2.8747	1.4013	2.8747	3.0151	2.8747
Hap_19	4.8062	4.885	1.9633	3.0151	1.7083	3.0151	3.1464	3.0151
Hap_20	1.7083	2.5602	4.671	4.7184	4.8062	4.7184	4.671	4.7184
Hap_21	4.6211	4.7184	2.1847	2.1847	2.7236	0.9954	1.9633	2.1847
Hap_22	4.7635	4.8467	2.3817	1.4013	2.8747	2.3817	2.1847	1.9633
Hap_23	5.1381	5.1381	4.9553	4.9873	4.9873	4.9212	5.0172	4.9873
Hap_24	4.7184	4.8062	2.1847	1.7083	2.7236	2.1847	1.9633	0.9954
Hap_25	4.7184	4.8062	2.1847	0.9954	2.7236	2.1847	1.9633	1.7083
Hap_26	5.1734	5.1734	5.0172	5.0452	5.0452	4.9873	5.0713	5.0452
Hap_27	4.7635	4.8467	2.3817	1.9633	2.8747	2.3817	2.1847	1.4013
Hap_28	5.0955	5.1381	4.7184	4.7635	4.7635	4.7635	4.7184	4.7635
Hap_29	5.1177	5.1177	4.9212	4.9553	4.9553	4.885	4.9873	4.9553
Hap_30	5.1566	5.1566	4.9873	5.0172	5.0172	4.9553	5.0452	5.0172
Hap_31	1.4013	2.3817	4.6211	4.7635	4.7635	4.671	4.7184	4.7635
Hap_32	4.7635	4.8467	2.7236	2.7236	3.1464	1.9633	2.5602	2.7236
Hap_33	4.7184	4.8062	1.7083	2.8747	1.9633	2.8747	3.0151	2.8747
Hap_34	4.8062	4.885	1.9633	3.0151	2.1847	3.0151	3.1464	3.0151
Hap_35	4.7635	4.8467	1.7083	2.8747	1.9633	2.8747	3.0151	2.8747
Hap_36	4.8062	4.885	2.5602	1.7083	3.0151	1.7083	2.3817	2.1847
Hap_37	4.7635	4.8467	2.3817	1.4013	2.8747	1.9633	2.1847	1.9633
Hap_38	4.7635	4.8467	1.7083	2.8747	1.9633	2.8747	3.0151	2.8747
Hap_39	5.1177	5.1177	4.9873	5.0172	5.0172	4.9553	5.0452	5.0172
Hap_40	4.7635	4.8467	2.3817	1.9633	2.8747	2.3817	2.1847	1.4013
Hap_41	5.1734	5.1734	5.0713	5.0955	5.0955	5.0452	5.1177	5.0955
Hap_42	1.4013	2.3817	4.5136	4.671	4.671	4.5687	4.6211	4.671
Hap_43	4.3317	4.4558	4.7184	4.8467	4.7635	4.7635	4.8062	4.7635
Hap_44	1.4013	2.3817	4.6211	4.7635	4.7635	4.671	4.7184	4.7635
Hap_45	1.4013	2.1847	4.6211	4.7635	4.7635	4.671	4.7184	4.7635
Hap_46	4.8062	4.885	2.5602	1.7083	3.0151	2.5602	2.3817	2.1847
Hap_47	4.7635	4.885	2.7236	1.9633	3.1464	2.7236	2.5602	2.3817
Hap_48	2.5602	2.1847	4.7635	4.885	4.885	4.8062	4.8467	4.885
Hap_49	0.9954	2.5602	4.671	4.8062	4.8062	4.7184	4.7635	4.8062
Hap_50	1.7083	1.7083	4.671	4.8062	4.8062	4.7184	4.7635	4.8062
Hap_51		2.3817	4.6211	4.7635	4.7635	4.671	4.7184	4.7635
Hap_52	6		4.7184	4.8467	4.8467	4.7635	4.8062	4.8467
Hap_53	29	31		2.3817	1.7083	2.3817	2.5602	2.3817
Hap_54	32	34	6		2.8747	2.3817	2.1847	1.9633
Hap_55	32	34	3	9		2.8747	3.0151	2.8747
Hap_56	30	32	6	6	9		2.1847	2.3817
Hap_57	31	33	7	5	10	5		2.1847
Hap_58	32	34	6	4	9	6	5	