

Phylogeography of the oyster borer (*Haustrum scobina*)

Sebastian Logan



A thesis submitted to Victoria University of Wellington in partial
fulfilment of the requirements for the degree of Master
of Science in Ecology and Biodiversity

2019

Abstract:

An effective investigation of the underlying ecological processes that shape genetic diversity and connectivity typically requires comparisons among phylogeographic studies of multiple species. Phylogeographic studies of New Zealand's coastal marine benthos have historically relied on *post hoc* speculation rather than directed research questions to investigate ecological processes. There has also been a lack of studies on direct developing marine molluscs. Direct developers are expected to have a low potential for dispersal and thus show a pattern of genetic isolation by distance across their distributions. Recent research indicates that this assumption may frequently be violated by instances of long distance dispersal/translocation. The oyster borer (*Haustrum scobina*) is an endemic direct-developing marine mollusc found in high abundances at rocky intertidal environments across the entirety of New Zealand. This distribution and life history makes *H. scobina* an ideal target to study genetic connectivity in a species expected to show low realised dispersal and high population genetic structuring. This thesis research used 379 new DNA sequences from the mitochondrial gene cytochrome c oxidase subunit 1 (COI) to investigate the phylogeography of *H. scobina* across the southern North Island. In addition 16 new COI sequences were inadvertently sequenced from the morphologically similar congener *Haustrum albomarginatum*. Results from both species support the recently proposed division of *H. scobina* and *H. albomarginatum* as separate species. *H. scobina* populations show significant geographic structure and a lack of haplotype diversity across the southeastern North Island concordant with results of another previous study of a direct developer. This finding suggests that ecological processes may be producing similar population genetic structures for direct developers generally. Contrast between high and low haplotype diversities in northern and southern *H. scobina* populations respectively, indicates that southern *H. scobina* populations may have originated via recolonisation from northern populations following a range contraction during the Last Glacial Maximum. Evidence of multiple long distance dispersal/translocation events was found indicating that long distance dispersal via rafting and/or inadvertent human-mediated translocations may have occurred frequently. Results are then discussed with a view to inform further research in to New Zealand direct developers.

Acknowledgements:

To begin, I would like to thank my supervisors Peter Ritchie and Nicole Phillips for their guidance and support over the last two years and for giving me the opportunity to study such an interesting subject. Thank you both for sharing your experience, being so easy to talk to and making me feel welcome in what was an unfamiliar environment. Your knowledge, advice and good humour has been truly invaluable.

Secondly I'd like to thank my parents, without their constant support and encouragement in so many aspects of my life I never would have had the opportunity to keep pursuing my goals and dreams. The numerous collection trips to remote coasts and beaches that we went on are some of the best memories I have from these last few years.

Thanks must also go to my excellent lab mates for making my time in the lab such an enjoyable experience. Special mention goes to Angela Fleming for constantly being available to help with any issues I encountered in the lab and answering my incessant questions about PCR. A huge thanks must also go to Kerry Walton, from patiently training me in all the lab techniques, to providing valuable advice and comments on thesis ideas and my results, this thesis would not have been possible without his assistance.

My thanks also go to the authors of Barco *et al.* (2015) for providing the published sequences on GenBank that served as the basis for this research. Finally, I would like to thank Bruce Marshall for providing his advice and experience on interpreting my results.

Table of Contents

List of figures:.....	1
List of tables:	2
Chapter 1: General introduction.....	3
1.1: Population connectivity	3
1.2: Measuring and interpreting genetic connectivity.....	4
1.3: Patterns of genetic connectivity	7
1.4: Marker choice	9
1.5: Geography and oceanography of New Zealand.....	11
1.6: New Zealand marine phylogeography and population genetics	13
1.7: Study species: <i>Haustorium scobina</i> (Quoy and Gaimard 1833).....	15
1.8: Thesis aims	17
Chapter 2: Phylogeography of the oyster borer: <i>Haustorium scobina</i> (Quoy and Gaimard 1833)	18
2.1: Introduction	18
2.2: Methods	20
2.2.1: Sampling plan/collection/storage	20
2.2.2: DNA extraction, purification, amplification and sequencing	21
2.2.3: Phylogenetic tree.....	22
2.2.4: Statistical analyses.....	23
2.2.5: Demographic history	24
2.3: Results	25
2.4: Discussion.....	39
2.4.1: Distributions and sampling of <i>Haustorium scobina</i> and <i>Haustorium albomarginatum</i>	39
2.4.2: Genetic diversity and population structure	39
2.4.3: Placement of haplotype five.....	42
2.4.4: Long distance dispersal events.....	45
Chapter 3: Conclusions and future research	47
3.1: Conclusions	47
3.2: Future research	48
References:	51

List of figures:

Figure 1.1: Four common patterns of population structure.....	8
Figure 1.2: Map of New Zealand and its major coastal current systems.....	12
Figure 2.1: Map showing the geographic distribution of <i>Haustrum scobina</i> COI haplotypes.....	33
Figure 2.2: COI median-joining network for <i>Haustrum scobina</i>	34
Figure 2.3: Rarefaction curve for all 610bp <i>Haustrum scobina</i> COI haplotypes sampled.....	34
Figure 2.4: A) A linked-loci clustering of the 379 <i>Haustrum scobina</i> COI sequences produced using Bayesian Analysis of Population Structure (BAPS) showing six groups B) Admixture analysis of the linked loci-clustering showing five groups.....	35
Figure 2.5: Mismatch distributions for the individual sites and all sites combined.....	37
Figure 2.6: Maximum Likelihood tree for (610bp) COI sequences showing the 17 <i>Haustrum scobina</i> haplotypes and other Muricid sequences acquired from Barco <i>et al.</i> (2015) via GenBank.....	38
Figure 2.7: Front and rear views of the shells from the three samples with the highly divergent haplotype 5.....	42
Figure 2.8: Front and back views of a reference selection of <i>H. scobina</i> and <i>H. albomarginatum</i> shells collected from Kawau Island that had their assignment to a species confirmed via COI sequencing.....	44

List of tables:

Table 2.1: Table of sample-site locations and respective summary statistics for the 610bp <i>Haustrum scobina</i> COI sequences.....	26
Table 2.2: Table of AMOVA results from the combined <i>Haustrum scobina</i> COI dataset.....	27
Table 2.3: Pairwise ϕ_{ST} sample site comparison for the <i>Haustrum scobina</i> 610bp COI sequences.....	28
Table 2.4: Table of the sequences used in the maximum-likelihood phylogeny including those derived from Barco et al. (2015) via GenBank.....	29
Table 2.5: Haplotype composition for the complete <i>Haustrum scobina</i> COI dataset.....	35

Chapter 1: General introduction

1.1: Population connectivity

The contemporary distribution and population-genetic structure of marine invertebrate species can be determined by a diverse range of interacting factors including biological traits, environmental conditions and historic events. Of particular importance, is the level of dispersal between discrete populations which can be loosely termed “population connectivity” (Lowe and Allendorf 2010). Understanding the levels of population connectivity for a species of interest is essential as it can affect resistance to local extinction (Brown and Kodric-Brown 1977), recolonization rates (Turgeon and Kramer 2012), adaptation and speciation (Mayr 1970, Ovenden 2013). In this thesis, species are defined using the biological species concept adapted from Mayr (1942) where species are groups of potentially or actually interbreeding, natural populations that are reproductively incompatible and thus isolated from other such groups.

As with many biological terms, populations can be defined in a number of ways depending on the context. Waples and Gaggiotti (2006) noted that populations are usually defined using either an evolutionary or ecological paradigm. In an evolutionary context, reproductive interactions and the genetic consequences are the primary concern. Here, populations can be defined as a group of conspecific individuals living close enough together to facilitate regular mating. This contrasts with the ecological paradigm where demographic cohesion is considered and populations are defined as groups of conspecifics that co-occur in time and space to enable demographic interactions such as competition and social interaction.

The delineation of populations in an evolutionary context hinges on the levels of reproductively successful migrants between groups of conspecifics which is often termed gene flow. When gene flow is low or non-existent between two populations, the forces of genetic drift (leading to the elimination of alleles) and mutation (leading to the creation of new alleles) can act independently within each population. Over time this leads to the accumulation of genetic differences and potentially reproductive isolation i.e. speciation (Mayr 1970, Slatkin 1987). Conversely, high gene flow limits genetic differentiation and can, over time, homogenise alleles between genetically diverse groups. Gene flow relies on an

absolute number of reproductively successful migrants per generation and is relatively independent of the sizes of each population (Ovenden 2013). Surprisingly, as few as 10 migrants per generation may provide sufficient gene flow to maintain homogeneity between large and small populations (Lowe and Allendorf 2010) and this number may be even lower if the migrants have genes undergoing positive selection (Rieseberg and Burke 2001).

Demographic cohesion also relies on gene flow but rather than an absolute number of migrants it depends on the proportion of the population exchanged as successful migrants (e.g. 10% of a population). Population sizes can therefore affect the level of demographic interdependence. In the marine environment many species have very large population sizes meaning demographic interdependence may require thousands more migrants than would be required for genetic connectivity. This means populations can be connected genetically but not demographically which (Ovenden 2013) refers to as “crinkled populations”.

Demographically uncoupled or “crinkled” populations must be treated as separate entities as they will respond to different stressors independently, regardless of any apparent genetic connectivity. Direct tracking methods such as capture-mark-recapture (CMR) and information on dispersal potential can be combined with genetic data to aid in characterising demographic uncoupling within genetically connected populations (Lowe and Allendorf 2010). Determining the levels of demographic interdependence is beyond the scope of this thesis and inferences must be limited to those that can be based solely on genetic connectivity (e.g. the location of phylogeographic breakpoints).

1.2: Measuring and interpreting genetic connectivity

Many marine benthic invertebrates have a biphasic life history with a demersal adult stage (Scheltema 1986) meaning the majority of migration occurs in early life history stage as small planktonic propagules. The small size of these propagules and the vastness of marine environments can prohibit the use of conventional, direct measurements of population connectivity such as satellite tracking of migrants. A combination of artificial batch marking using either paint (Henry and Jarne 2007) or immersion of larvae in chemical tags such as calcein (Thorrold *et al.* 2002) with traditional CMR methods may increase the likelihood of successfully observing dispersal directly. However, the high fecundity and mortality typical

of marine benthic invertebrates can make this difficult to achieve in practice (Thorrold *et al.* 2002). Another approach could be to use biochemical profiling of shells to attempt to track an individual back to locations it may have developed within (West *et al.* 2006). This method however requires prior knowledge of the isotope compositions present across a species' distribution to be able to accurately track an individual's previous movement and such information may not be readily available. In many cases this leaves indirect genetic methods as the most appropriate option for investigating dispersal in many marine benthic invertebrate taxa which can limit the potential inferences that can be made about demographic cohesion. The use of alternative, direct methods such as the lab/field testing of dispersal potential and measurements of vital rates in combination with genetic analyses may better elucidate contemporary demographic processes.

Phylogeography is the study of the underlying processes that determine how genetic diversity is distributed geographically (Avice 2000). As mentioned, the level of gene flow among populations is a primary factor that determines the level of genetic differentiation. As gene flow is determined by factors such as dispersal potential (Weersing and Toonen 2009), habitat availability and geographic barriers to dispersal (Colson and Hughes 2004), the observed phylogeography can provide information about the underlying causes. Dispersal potential during the larval stage is often cited as the key predictor of genetic structure for marine benthic invertebrates (Palumbi 1994). Species without a pelagic larval phase (direct-developers) are thought to have markedly less potential for dispersal than those with one (indirect-developers)(Scheltema 1986). Additionally, within species with pelagic larvae, pelagic larval duration (PLD) and feeding method (lecithotrophic vs planktotrophic) appears to also affect the potential for dispersal. Longer time spent in the water column may provide more opportunity for ocean currents to move the larvae between sites (Wares *et al.* 2001). A study by Weersing and Toonen (2009) reviewed a number of papers examining PLD and found almost no correlation between PLD and genetic differentiation. Accurate predictions about genetic differentiation cannot be made from simple life history traits alone.

Ocean currents, habitat availability and geographic barriers will affect the movement and reproductive success of larvae/juveniles and this influence may change depending on mode of development or PLD (Wares *et al.* 2001). Selection also plays a role in structuring

populations, differing environmental conditions may provide strong selective pressures that increase the reproductive success of locally recruited juveniles/larvae (due to better local adaptation) relative to individuals recruited from more distant sites (Hellberg *et al* 2002). Selection may also act upon the marker used, potentially showing a different level of connectivity if the marker provided adaptive benefits than would be observed with a neutral marker. An examples of this issue is seen in two studies looking at genetic discontinuity in the oyster *Crassostrea virginica* across Cape Canaveral. Reeb and Avise (1990) found significant genetic discontinuity using mtDNA markers which conflicts with a previous study (Buroker 1983) using allozyme markers that showed no break. This could be explained by selection acting to homogenise adaptive allozyme diversity and not neutral mtDNA diversity. Finally, historical effects must also be considered as estimates of genetic connectivity typically rely on an equilibrium existing between gene flow and genetic drift. Successive founder events during recolonization/range expansion may produce a more genetically homogeneous pattern than would be expected from contemporary demographic processes (Benzie and Williams 1997). Furthermore, a combination of gene-surfing (Barton *et al.* 2013) and competitive exclusion may also produce genetic homogeneity that can persist over time in a species undergoing range expansion (Waters *et al.* 2013). A large effective population size (N_e) would slow down genetic drift in newly established populations potentially masking the effect of a low migration rate until equilibrium is again reached between migration and N_e . The time taken for equilibrium between genetic drift and migration to be reached will be longer in species with larger N_e and lower rates of migration such as many direct-developing marine invertebrates (Hellberg *et al.* 2002, Marko 2004).

Indeed, care must be taken in assigning causal links to observed phylogeography as many different processes may be interacting synergistically. Independent evidence such as the lab/field dispersal tests of Dohner (2016) and field observations of habitat choice and availability may be used to assess the individual influence of these factors. To determine if markers are under selection one option is to sequence multiple markers. If any markers conflict with a geographic pattern seen consistently in other loci this could indicate that this marker may be under selection. Another approach is to compare studies of heterospecifics that differ in life history traits to determine whether these traits can affect genetic connectivity (Fleming *et al.* 2018). Environmental barriers to dispersal such as geographic

features and ocean currents may be identified when phylogeographic breaks are congruent between heterospecifics with similar life histories. In previous studies of New Zealand marine fauna, the ability to make unbiased comparisons has been hindered by the use of different markers, incomplete sampling and sampling across different spatial scales (Ross *et al.* 2009). In future, study design should promote unbiased comparisons by using the same markers, sampling scales and ranges as the studies intending to be compared to. By removing as many possible sources of bias, the influence of individual factors involved in determining genetic connectivity may be identified.

1.3: Patterns of genetic connectivity

While determining the causal factors may be problematic, common phylogeographic patterns can reveal historical and contemporary patterns of gene flow. Hellberg *et al.* (2002) identified a number of common patterns of genetic differentiation ranging from closed populations with little or no gene flow to completely open or “panmictic” populations. Examples of these are provided below and in Fig. 1.1.

Panmixia (Fig. 1.1A): Ongoing, long-distance gene flow can unify the effects of genetic drift across distant, discrete habitats homogenising genetic variation. This pattern is more common amongst marine invertebrates with pelagic larvae that can regularly engage in long-distance dispersal. In panmictic populations private haplotypes (unique haplotypes limited to a single site) will be rare or non-existent.

Phylogeographic break (Fig. 1.1B): When a geographic feature or ocean currents prevents migration genetic drift can act independently on either side of a phylogeographic break leading to the accumulation of genetic differences and eventually reciprocal monophyly. The maternal inheritance of mtDNA means that it has a lower effective population size than comparable nuclear markers, thus it will experience more genetic drift and reach reciprocal monophyly across a phylogeographic break more rapidly. A phylogeographic break may also appear if a species had recently expanded beyond a historic barrier (recolonization or range expansion). In this case there would much less genetic diversity in the newly established populations and over time this transition would become less defined as genetic drift and migration reach equilibrium.

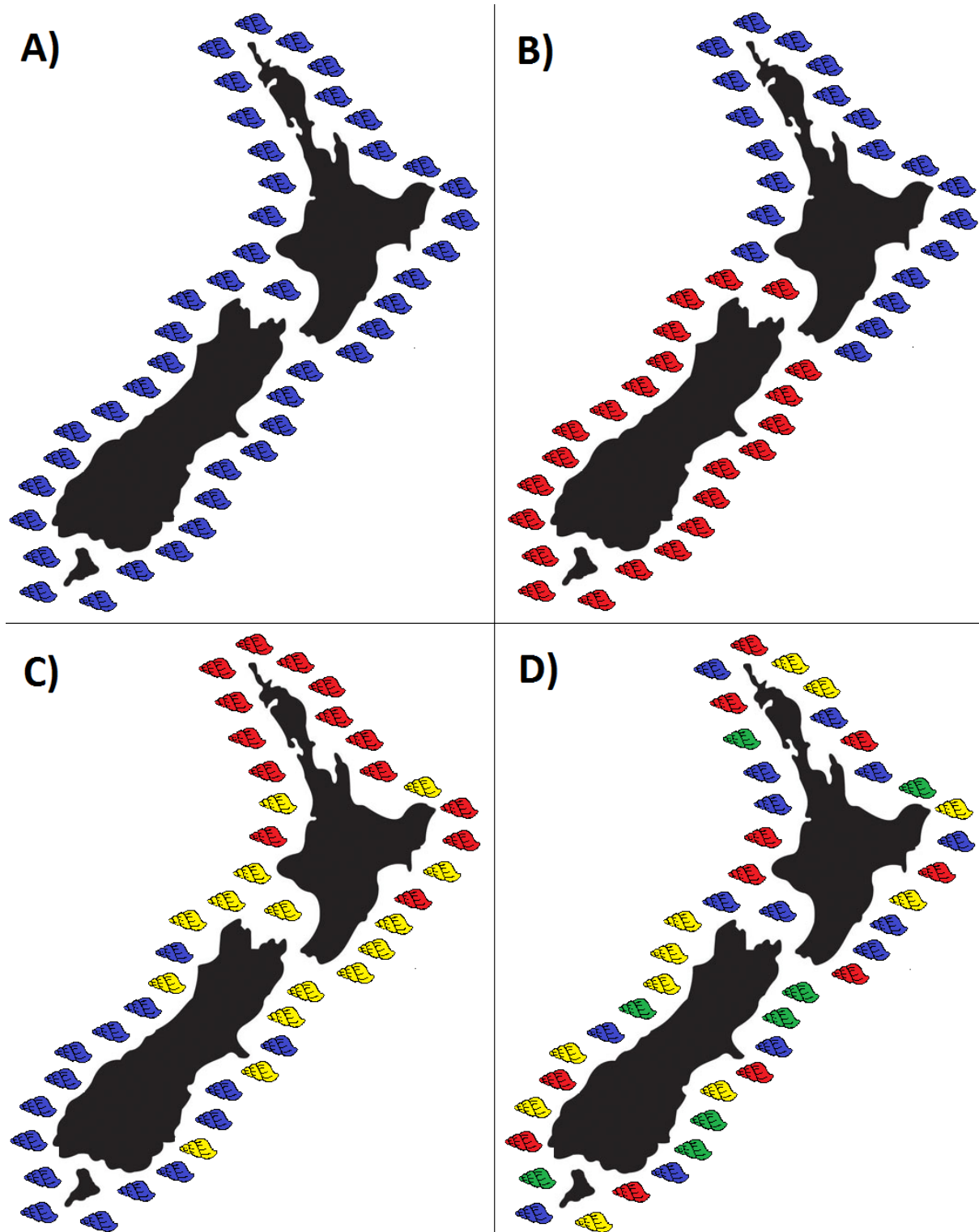


Figure 1.1: Four common patterns of population structure. A) Panmixia where high gene flow maintains genetic homogeneity, B) Phylogeographic break where a genetic divergence between populations coincides with a geographical feature, C) Isolation by distance (stepping stone model) where limited dispersal creates increasing genetic divergence with distance D) Chaotic genetic patchiness potentially caused by fine-scale oceanographic variation interacting with pelagic larvae NB: the colour of each shell represents a distribution of sequences at a site rather than a specific sequence

Isolation by distance or stepping-stone model (Fig. 1.1C): This pattern is most common in species with restricted dispersal such as direct-developers or species with short-lived pelagic larvae. In this case populations exchange migrants most frequently with sites that are in close proximity creating a pattern of increasing genetic differentiation (F_{st}) with distance.

Geographic clines: When environmental conditions that change across a geographic axis exert strong selective pressures on a marker, a pattern of progressive genetic change can result. Though it may appear similar to IBD allele frequencies along the geographic cline will change progressively rather than the more random distribution seen in IBD. Another difference from IBD is that this pattern can occur in species with potential for long distance dispersal (e.g. those with a long PLD) if strong selective forces favour local recruitment and natal fidelity.

Chaotic genetic patchiness (Fig. 1.1D): In some marine invertebrate species with pelagic larvae, adult populations may show low genetic diversity yet individual larval cohorts may be genetically differentiated. This could be due to larvae coming from different sources at different times of the year as a result of variable ocean currents. Alternatively, the large numbers of larvae released combined with high mortality common to many marine invertebrates may explain the temporal genetic variability in cohorts (Hedgecock 1994). Overall this can lead to a weak (or no) association between genetic differentiation and geographic distance.

1.4: Marker choice

Adding to the complexity in interpreting phylogeographic data, is the variability in genetic signals that come from different types of markers used for the same species. The variability in markers can originate from differences in: the size/number of loci used, levels of selective pressures, rates of mutation and mode of inheritance (Hellberg *et al* 2002). Choosing an appropriate marker requires consideration of the advantages/disadvantages in the context of the question being asked as different markers can reveal genetic signals at different spatial/temporal scales.

Allozymes: These are different forms of proteins produced by protein coding alleles. Variation in allozyme length is measured by electrophoretic separation across a gel and comparison to a reference sample ladder. The advantages of this method are its low cost

and simplicity as no DNA extraction is needed and the fact that allozymes are generally co-dominant enabling easy integration in to standard population-genetic models. However this method is limited by the lack of variability that can be detected as only changes that alter amino-acid sequence will be observed (~30% of total potential variation). Another issue is that allozymes may also be under selection as they are protein-coding by nature. Historically this method was frequently used but it is becoming less common as more effective methods such as mtDNA and microsatellite sequencing becomes cheaper.

Microsatellite DNA: These are short tandem repeating regions with a repeat motif typically 2-10 bps long. Specific primers have to be developed to amplify the repeating region and then the genetic distances can be used to measure the number of tandem repeats. They are increasingly being seen as the most appropriate marker for population genetic studies as their high polymorphism can reveal more contemporary information than other markers (Gardner *et al.* 2010). The main disadvantage is the cost and technical challenges for developing specific primers for each loci. However once primers are developed this method can be cheap to reproduce and easily automated for large numbers of samples. Additionally these markers are generally selectively neutral and co-dominantly inherited.

Nuclear DNA (nuDNA): This involves using PCR primers to amplify and sequence short sequences of DNA (typically up to 1000bps long). As all the DNA is sequenced, more variability can be detected than from allozymes while still retaining codominant inheritance. If a diploid loci are sequenced, heterozygosity can lead to an unreadable sequence. Generally nuDNA has a slower rate of mutation than mtDNA but recombination due to bi-parental inheritance can mean that genetic differentiation is homogenised quicker, potentially removing the signature of historic events faster. The main advantage of using this method is the simplicity and ease of PCR amplification and Sanger sequencing.

Mitochondrial DNA (mtDNA): mtDNA is a short (roughly 16500bp) circular form of DNA that is separate from nuDNA and is maternally inherited. Mitochondrial DNA is typically sequenced in the same way as nuDNA through the use of PCR primers. Some of the primers may be applied to many different species increasing the reproducibility and reducing the cost of this method (e.g. the "Folmer" primers, Folmer *et al.* 1994). The maternal inheritance of mtDNA means it has a smaller effective population size than comparable nuDNA markers leading to more genetic drift and faster reciprocal monophyly between two

genetically restricted populations. The lack of recombination can mean that historical effects persist longer in mtDNA than in nuDNA. A combination of using both mtDNA and nuDNA may be more appropriate if one wishes to differentiate between the historical (phylogeographic) and contemporary (population genetics) effects of gene flow without investing the time/money to develop microsatellites.

1.5: Geography and oceanography of New Zealand

New Zealand is an archipelago, containing over 700 islands that span from the Kermadec Islands (29° S) in the north to Campbell Island (52° S) in the south (Ross *et al.* 2009). It has one of the world's largest Exclusive Economic Zones (EEZs) covering roughly 480,000 Km². The geology of New Zealand has resulted in largely from tectonic activity along the plate boundary that runs roughly from North to South through the main North and South Islands. Tectonic uplift along this boundary formed the Southern Alps and much of the contemporary geography. Repeated changes in sea levels since the Pliocene (roughly 5 million years BP) have also resulted in the creation/removal of land-bridges and straits in the Cook Strait and Auckland isthmus (Lewis *et al.* 1994).

The New Zealand archipelago sits across the subtropical front where the warmer subtropical water from the north meets the colder subantarctic water from the south (Fig. 1.2). The interaction between major currents and topographic features (coastlines, headlands, straits etc.) produces a complex pattern of local currents and eddies (Fig. 1.2)(Ross *et al.* 2009). Subtropical water arriving from the Tasman front to the northwest bifurcates across the northern tip of the North Island creating the West/East Auckland currents (WAC/EAC). The EAC interacts with the East Cape headland and the Chatham Rise to produce the East Cape eddy (ECE) and Wairarapa eddy (WAE). Subantarctic water travels along the Southland current (SOC) across the east coast of the South Island until the Chatham Rise where the majority diverts eastward along the subtropical front. The complexity produced by the various currents and eddies can prevent the accurate prediction of gene flow from oceanography alone.

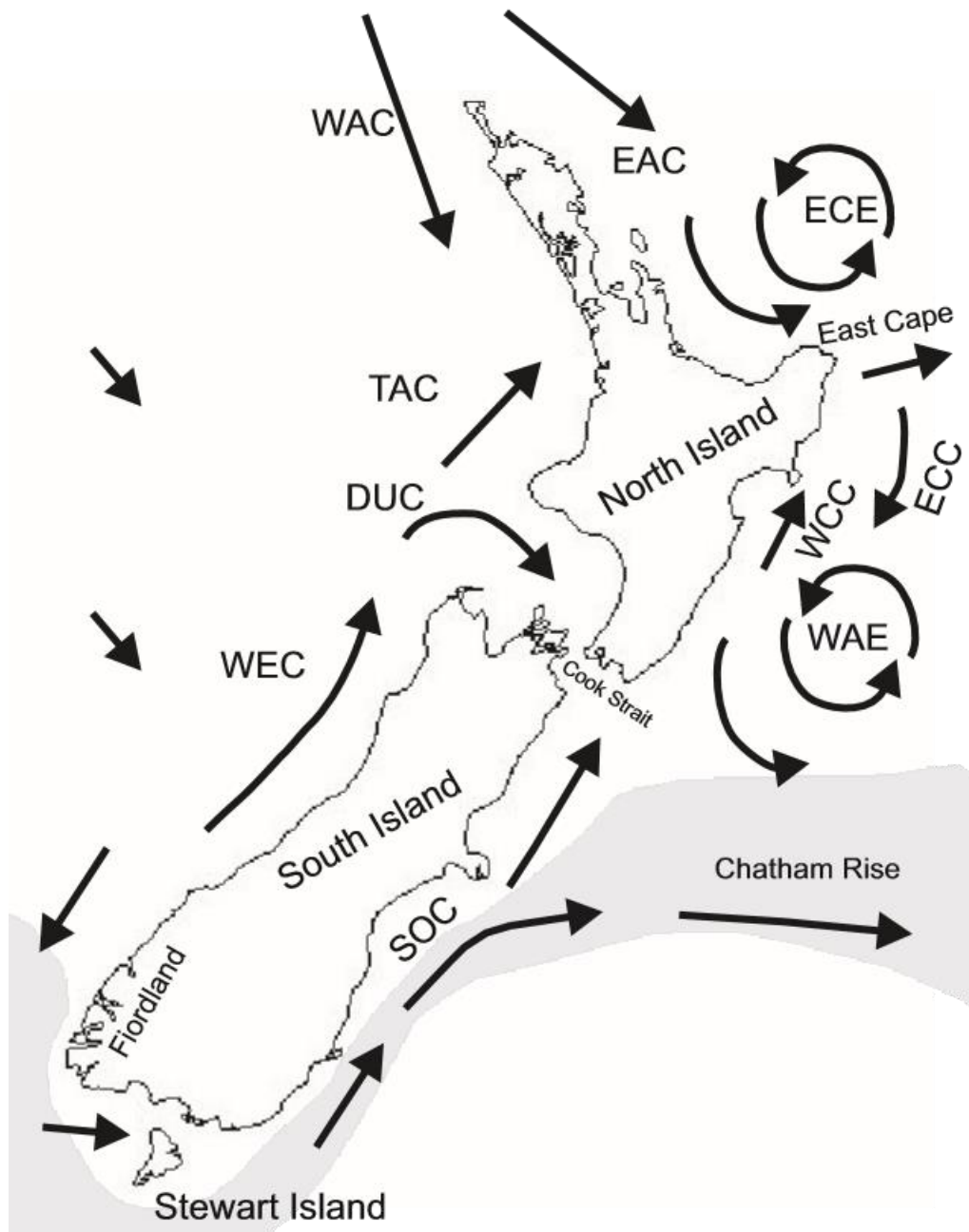


Figure 1.2: Map of New Zealand and its major coastal current systems. Grey region=Sub-tropical front DUC=D'Urville Current, EAC=East Auckland Current, ECC=East Cape Current, ECE=East Cape Eddy, SOC=Southland Current, TAC=Tasman Current, WAC=West Auckland Current, WAE=Wairarapa Eddy, WCC=Wairarapa Coastal Current, WEC=Westland Current (adapted from Ross *et al.* 2009).

1.6: New Zealand marine phylogeography and population genetics

The complex current systems, diverse geography, isolation and high degree of endemism makes New Zealand an ideal environment in which to study patterns of gene flow and population structure in the marine environment (Giribet and Boyer 2010, Ross *et al.* 2009). The recent reviews by (Ross *et al.* 2009) and (Gardner *et al.* 2010) provide an overview of over 60 phylogeographic studies of New Zealand marine taxa, revealing a number of commonly observed genetic patterns and a lack of comparability between studies. Additionally, a number of marine phylogeographic papers have been published subsequently to 2010 with some examining rafting as a potential means of long-distance dispersal.

Panmixia: 16 of the 58 papers (28%) reviewed by (Gardner *et al.* 2010) displayed this pattern. This pattern hypothetically makes logical sense for indirect-developing species with longer pelagic larval durations such as the rock lobster *Jasus verreauxi* (Brasher *et al.* 1992) which has a PLD of >8months and was observed to be panmictic. A longer PLD could enable regular long-distance dispersal and homogenise variation. Another explanation for this pattern is that the marker used may have had insufficient resolution to show the “true” genetic structure that could be observed with alternative markers. One example of this is the cockle *Austrovenus stutchburyi* where an initial study using allozymes (Lidgard 2001) showed no differentiation, but a subsequent study using mtDNA COI sequences (Ross *et al.* 2012) showed the presence of six subpopulations with a genetic pattern of IBD.

IBD and geographic clines: the majority of studies reporting this pattern were designed to test structure at finer-scales mainly in Fiordland (Ross *et al.* 2009) and it is unknown if this would be seen in studies across a wider region. In these cases local geography may act to limit gene flow to closer sites and thus produce a relationship between genetic differentiation and geographic distance. An exception to this was the NZ-wide study of the waratah anemone *Actinia tenebrosa* (Veale and Lavery 2012) using microsatellites that found genetic IBD. In this case, a latitudinal cline in reproductive strategies (sexual vs asexual) was observed and may have acted to maintain a pattern of genetic IBD. Confirming a pattern of IBD ideally requires sampling across the entire extent of the species’

distribution using a marker that is polymorphic enough to show contemporary rather than historical gene flow.

Congruent north-south phylogeographic breaks: The most consistent genetic pattern in New Zealand appears to be a north-south differentiation being reported by over half of the papers reviewed by (Ross *et al.* 2009) and over a third of the papers reviewed by (Gardner *et al.* 2010). Upwelling has repeatedly been raised as a possible cause for this disjunction (Apte and Gardner 2002; Star *et al.* 2003; Waters and Roy 2004; Ayers and Waters 2005; Veale and Lavery 2011) however assigning a causality to upwelling is inappropriate without evidence from multiple studies using a consistent sampling regime and a better understanding of larval behaviour in relation to local oceanography. The location of the break varies across studies (Gardner *et al.* 2010) possibly reflecting species-specific differences in dispersal potential or inconsistent sampling. Alternative explanations for this break include larval retention in local eddies (Ross *et al.* 2009, Chiswell and Roemich 1998), habitat preference (Jones *et al.* 2008) or southward range-expansion following the Last Glacial Maximum (LGM) (Newnham *et al.* 1999, Fleming *et al.* 2018). The combined work of Fleming *et al.* (2018) and Walton (2017) showed that populations of *Cominella maculosa* display a north-south division of high to low haplotype diversity respectively, across the species' geographic distribution. Post-glacial recolonisation appears to be the most likely explanation for this particular division and population structuring in the south may have been determined by interactions between the movements of range expanding populations and local geography.

Long-distance dispersal via rafting: Rafting on a variety of substrates or on ships has often been debated (Jackson 1986) as a potential explanation for the wide distributions of many marine invertebrates with limited dispersal potential (e.g. direct-developers or species with a short PLD). In New Zealand, recent genetic studies present strong evidence for the historic and continuing effect of rafting on marine genetic connectivity specifically implicating the buoyant Bull Kelp, *Durvillaea antarctica* as a vector. (Bussolini and Waters 2015) found that bull kelp appear to have an oceanographically-determined genetic structure specifically matching the same north-south break at Cape Campbell seen in many marine invertebrate taxa. Genetic evidence of long-distance connectivity has been observed in a number of New Zealand marine invertebrates that are strongly associated with *D. antarctica* including: the

chiton *Sypharochiton sinclairi*, (Nikula *et al.* 2011), the bluish top shell, *Diloma nigerrima* (Donald *et al.* 2011), the sub-antarctic top shell *Cantharidus antipoda* (Donald and Spencer 2016), the sea slugs of the genus *Onchidella* (Cumming *et al.* 2014) and limpets of the genus *Cellana* (Gonzalez-Wevar *et al.* 2017). While these studies support the possibility of *D. antarctica* acting as a regular vector for marine dispersal assigning causality would require direct evidence of rafting activity, such as when (Fraser *et al.* 2011) found 10 epifaunal species rafting on bull kelp over 400km away from their inferred origin site. Alternative vectors could include pumice stones or human mediated transport (e.g. on ballast rocks) however studies have not provided conclusive evidence for this occurring in New Zealand (Ross *et al.* 2009). Fleming *et al.* (2018) found that Nelson populations of *C. virgata* showed unusual similarity to more northern populations potentially indicating that this population arose via inadvertent human-mediated translocation given that this is a busy port though the actual vector is uncertain.

1.7: Study species: *Haustrum scobina* (Quoy and Gaimard 1833)

While taxonomic relationships have historically been contentious (Tan 2003), according to the most recent phylogeny based on molecular evidence (Barco *et al.* 2015) the taxonomic subfamily Haustriinae contains two extant genera (*Bedevea* Iredale, 1924 and *Haustrum* Perry, 1811) and between eight and ten species (depending on the author). The genus *Bedevea* Iredale, 1924 is endemic to southern Australia with the exception of *Bedevea paivae*, being introduced via human activity to South Africa, Madeira and the Canary Isles. COI mtDNA sequences from *B. paivae* and *B. vinosa* form a monophyletic clade (Barco *et al.* 2015). The genus *Haustrum* is endemic to New Zealand with four species (*H. scobina*, *H. lacunosum*, *H. albomarginatum*, *H. haustorium*) forming a separate monophyletic clade. The placement of *H. albomarginatum* has been contentious, previously being synonymised with *H. scobina* due to similar shell morphologies (Tan 2003) however molecular evidence supports the separation of these two species (Barco *et al.* 2015). The concordance between the two monophyletic clades and geographic separation supports the possibility of a vicariant separation event leading to speciation. Given that the remaining members of Haustriinae are all endemic to southern Australia it is logical to hypothesise that molecular

evidence would also place these in the *Bedevea* genus. Additionally, six Pleistocene fossil shells discovered in southern South America showed significant morphological similarity to other members of Haustriinae specifically *Bedevea* spp. (Gordillo and Nielsen 2013). It was inferred that this putative *Bedevea* species may have arrived via rafting on bull kelp across the Antarctic Circumpolar Current (ACC) during the Quaternary period (Gordillo and Nielsen 2013).

Haustorium scobina is an endemic, carnivorous, intertidal gastropod that is usually found inhabiting rocky-reef environments in great numbers and also soft shores if sufficient hard substrates are present (Morton and Miller 1968). *H. scobina* exhibits direct-development without any pelagic larval stage, which presumably limits its potential for dispersal. The reproductive strategies (Carrasco and Phillips 2014) and resulting effects on maternal lipid provisioning (Carrasco *et al.* 2016) have been well studied for *H. scobina* in relation to two other whelk species from the genus *Cominella* (*C. virgata* and *C. maculosa*). *H. scobina* deposits egg capsules in an irregular fashion sometimes even on the shells of conspecifics (Carrasco and Phillips 2014). The egg capsules display adelphophagy with over typically 200 embryos per capsule, many of these embryos are nutritive nurse embryos and will be consumed by developing juveniles (Carrasco and Phillips 2014). Since the early 1990's tributyltin (TBT), a highly toxic marine anti-fouling paint, was shown to induce widespread imposex (female infertility due to masculinisation) in *H. scobina* populations in almost every port with more than a few boats (Smith and McVeagh 1991) and it was used as a bio-indicator for TBT pollution. Reassuringly for conservationists, following the ban on the use of TBT in 2003 severely affected populations in Waitemata and Tauranga harbours appear to have almost completely recovered (Jones and Ross 2018). This rapid recovery could imply that *H. scobina* has a far greater potential for migration/recolonisation than would be predicted from its mode of development alone.

To date only eight *H. scobina* mtDNA COI sequences have been reported (Barco *et al.* 2015). Given its widespread distribution, low potential for dispersal and endemism, *H. scobina* represents an ideal species to study realized gene flow in a direct-developing species. Furthermore, the two previously mentioned *Cominella* species have both had their phylogeographies studied across the North Island and upper South Island (Dohner 2016,

Walton 2017, Fleming *et al* 2018) providing a robust dataset of species differing in early life-history traits from which to draw comparisons.

1.8: Thesis aims

The first aim of this thesis is to investigate the population-genetic structure of *H. scobina* populations using mitochondrial DNA sequencing of samples collected across the South and East Coasts of the North Island from Gisborne to the west of Wellington. Some of the sites chosen were identical to those used in Dohner *et al.* (2018) and Fleming *et al.* (2018) and one goal is to see whether *H. scobina* populations are genetically subdivided across the southeast coast of the North Island as was found in *Cominella maculosa* populations in this previous work. A second aim is to assess whether a larger genetic dataset from a wider range of sites supports the separation of *H. scobina* and *H. albomarginatum* as separate species proposed by (Barco *et al.* 2015) and other authors. Chapter Two provides a comprehensive description and discussion of the DNA sequencing of mitochondrial cytochrome c oxidase subunit one (COI) genes. Chapter Three provides a general discussion of the results and how they could inform further research into the population-genetic structure of *H. scobina* and other direct developer species.

Chapter 2: Phylogeography of the oyster borer: *Haustorium scobina* (Quoy and Gaimard 1833)

2.1: Introduction

In the marine environment, genetic connectivity is determined by a diverse range of factors and interactions between these can produce a complex genetic patterns (Hellberg *et al* 2002). Many marine benthic invertebrates have a bi-phasic life history with the majority of migration occurring during the early stages of life as pelagic larvae (Palumbi 1994). In other species, including those lacking pelagic larvae, sporadic rafting events (Fraser *et al.* 2011) and human mediated dispersal (Crispo *et al.* 2011) may also have a significant impact on genetic connectivity. Alternate factors such as ocean currents, habitat availability, life history, behaviour and historical events among others may also affect genetic connectivity. Interactions between these different factors means that making accurate predictions of genetic connectivity is difficult when using only a single factor. In some cases an inability to directly track individual dispersers or make accurate predictions about connectivity means that for many marine species, indirect genetic studies is one of the best alternative methods for investigating migration and genetic connectivity. Comparative phylogeographic studies of heterospecifics can reveal the impact that species-specific traits have on genetic connectivity assuming that relevant biological information about the species is available. An understanding of genetic connectivity can reveal the effect of historical events in addition to ongoing ecological processes and can be essential to the sustainable management of species in the face of harvesting pressures and a changing climate. Historically marine reserves have been located in isolated regions to increase the ease of management (Bell and Okamura 2005). However the importance of maintaining genetic connectivity to sustain long-term survivability is increasingly being seen as an essential component in conservation efforts (Palumbi 2003).

The choice of marker used in a phylogeographic study can have a significant impact on any observed genetic pattern. Microsatellites are commonly used as a marker in phylogeography as their high degree of polymorphism provides sufficient resolution to observe contemporary patterns of gene flow (Hellberg *et al.* 2002). The disadvantage to using microsatellites is the high cost and complexity involved in developing the method

compared to more affordable options such as mitochondrial DNA (mtDNA) markers. The maternally inherited mtDNA has a smaller effective population size than diploid nuclear markers. This means that genetic drift acts faster on mtDNA and thus mtDNA may show genetic differentiation that may not be detected by nuclear markers. The lack of recombination in mtDNA means that maternal ancestry can be directly followed allowing historical events such as the direction of a range-expansion to be tracked. While contemporary gene-flow is generally the most important consideration for conservation, historical patterns such as bottlenecks and range-expansions can also reveal how a species could respond to future demographic changes (Drummond *et al.* 2005). The mtDNA cytochrome c oxidase subunit 1 (COI) is a commonly used marker for phylogeographic and population-genetic studies and has been proposed as a potential “DNA barcode” for identifying and delimiting species (Savolainen *et al.* 2005). The mtDNA COI gene was chosen for use in this study due to its affordability, potential for revealing historical patterns and wide usage in phylogeographic studies of marine benthic invertebrates (Ross *et al.* 2009, Fleming *et al.* 2018).

The complex geography and oceanography of New Zealand makes it an ideal environment to study the ability for species to disperse across a range of different environments (Giribet and Boyer 2010, Ross *et al.* 2009). Studies of New Zealand marine taxa have reported various different genetic patterns with the most common being a North-South divide occurring at or near the Cook Strait (Ross *et al.* 2009, Gardner *et al.* 2010). While upwelling has often been invoked to explain this North-South pattern, a lack of consistent methodology across studies has hindered phylogeographic comparisons that could add support to the suggestion that this divide is a result of environmental factors (Ross *et al.* 2009). To overcome this, future studies will need to promote unbiased comparisons by sampling across the same sites, using appropriate (large) sample sizes and identical markers where possible.

Haustrum scobina is an intertidal marine mollusc, endemic to New Zealand that is typically found in large numbers on rocky reefs and other hard substrates (Morton and Miller 1968). *H. scobina* exhibits direct-development where juveniles hatch from egg capsules attached to a substrate without a pelagic larval phase (Carrasco and Phillips 2014) and this may limit dispersal relative to other species with a larval phase. While direct development is generally

assumed to result in limited dispersal (Palumbi 1994) sporadic rafting events may explain the genetic patterns seen in similar, endemic intertidal marine molluscs such as *Cominella maculosa* (Dohner *et al.* 2018, Fleming *et al.* 2018, Walton 2017). The phylogeography of *C. maculosa* has been well-studied across New Zealand (Dohner *et al.* 2018, Fleming *et al.* 2018, Walton 2017) and the differences in reproductive strategies between *C. maculosa* and *H. scobina* have also been characterised (Carrasco and Phillips 2014, Carrasco *et al.* 2016). This presents an opportunity to perform a comparative phylogeographic study between *H. scobina* and *C. maculosa* that could reveal whether species-specific traits or environmental factors are the primary influences on genetic connectivity for these two species. To date the only *H. scobina* mtDNA COI sequences were taken from eight samples from Hawera (Barco *et al.* 2015). This study showed *H. scobina* is a separate species from the closely related congeneric *Haustrum albomarginatum* which was previously considered to be one species (Tan 2003). The lack of a larger population-level study and the similarity in shell morphologies between *H. scobina* and *H. albomarginatum* raises the possibility that other cryptic species may also be present within *H. scobina*.

The aim of this thesis research was to investigate the population-genetic structure of *H. scobina* populations across the South and East Coasts of the North Island by determining the DNA sequence of the mtDNA COI genes sampled from *H. scobina* populations across this region. The sites chosen were identical to some of those used in Fleming *et al.* (2018) and Dohner *et al.* (2018) to enable fair comparisons to be made with the phylogeography of *C. maculosa*/*C. virgata*. Another goal was to determine whether there is evidence of long-distance dispersal events such as the “northern” *C. virgata* haplotype that was found in a Nelson site by Fleming *et al.* (2018). Finally, this study also further investigated the species’ delimitation of *H. scobina* that was proposed by Barco *et al.* (2015).

2.2: Methods

2.2.1: Sampling plan/collection/storage

In this study 376 *Haustrum scobina* samples were collected from a total of 12 sites across the North Island. *Haustrum scobina* were typically found in rocky intertidal areas along the exposed coastline. They were visible around low tide attached to solid substrates and usually found in groups of >5 individuals. The collection sites correspond to areas sampled in

previous studies (Fleming *et al.* 2018). Additionally, 16 *Haustrum albomarginatum* samples were collected from Kawau Island (site code=KIS) and were included in the phylogeny.

Samples were collected by hand from the intertidal zone at low tide and immediately frozen to -20°C during a series of trips between February and July 2017. Samples were then kept frozen until they could be dissected in the lab where the foot muscle tissue and operculums were excised and stored in 95% EtOH at 4°C. The emptied shells were cleaned and saved for long-term storage at The Museum of New Zealand Te Papa Tongarewa under the reference number (TBA).

2.2.2: DNA extraction, purification, amplification and sequencing

DNA extraction was performed using the phenol-chloroform extraction method (Sambrook *et al.* 1989). A small (2 mm³) piece of each tissue sample was placed in 580 µl of extraction buffer (10 mM TrisHCl pH 8.0, 50 mM NaCl, 10 mM EDTA, and 0.2% Sodium Dodecyl Sulphate) and 20 µl of Proteinase K (20 mg/µl). These were gently rotated and incubated at 40 °C overnight (>8 hours) until the tissue was completely digested.

Proteins were precipitated by adding 600 µl of phenol-chloroform-isoamyl alcohol (25:24:1), the tubes were then placed on a rocker for 15mins, centrifuged at 13000RPM for 5mins to clarify the DNA-containing supernatant. The aqueous layer was then removed and the process repeated with an identical second wash of phenol-chloroform-isoamyl alcohol (25:24:1). The process was then repeated a final, third time with chloroform-isoamyl alcohol (24:1).

As a final step the DNA was precipitated by adding 2.5x volume of chilled 98% EtOH and 35 µl of 3M sodium acetate. Tubes were placed on the rocker for 10 minutes and cooled to 4°C for >1hour. The precipitated DNA was pelleted by centrifugation at 13000 RPM for 25 mins at 4°C. The solution was completely removed and the pellet was washed with 1mL of 70% EtOH and centrifuged again at 13000 RPM for 15 mins. The remaining solution was removed and the pelleted DNA was dried in a vacuum concentrator. The DNA was then resuspended in 30-50µl TE buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA) and stored at 4°C. The DNA was

quantified using a Nanodrop ND1000 spectrophotometer (Thermofisher Scientific).

Following quantification, the DNA samples were diluted to 25 ng/μl.

A 710 bp portion of the COI gene was amplified using the Polymerase Chain Reaction (PCR) with the two universal “Folmer primers” for invertebrates: LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer *et al.*, 1994). The PCR mix was a total volume of 25μl and contained 67 mM Tris-HCl, 16 mM (NH₄)₂SO₄, 0.1% stabilizer, 3mM MgCl, 0.6mg/ml BSA, 0.1μM of each primer, 0.4mM dNTPs, 0.05 U/μl of Taq polymerase and 25 ng of template DNA. The thermocycler conditions were an initial denaturing on 95°C for 180s, then subjected to 40 cycles of 95°C for 35s, 50°C for 35s, 72°C for 45s, followed by a final 72°C extension for 10 mins.

The resultant PCR products were electrophoresed at a 1.5% agarose gel, stained with ethidium-bromide and visualised using UV light. Amplified products were purified by adding 0.5μL of EXO-SAP-IT (Amersham Pharmacia Biotech) and incubated at 37°C for 30min and 80°C for 15 min. The DNA sequence of the PCR products were determined using a 3730xl Genetic Analyser (Applied Biosystems) at Macrogen Inc. (Seoul, South Korea).

Five DNA sequences were discarded due to poor quality. Eight, previously published (Barco *et al.*, 2015), COI sequences were added from GenBank giving a final dataset of all 379 *H. scobina* COI sequences. All sequences were aligned in Geneious (version 8.1.9) using the geneious alignment tool (Kearse *et al.*, 2012). COI sequences were then trimmed to 610bp to match the COI dataset reported in the previous study of Fleming *et al.* (2018).

2.2.3: Phylogenetic tree

The most appropriate model of sequence evolution for the complete dataset was the Tamura 1992 model (Tamura 1992) with no invariant sites and identical substitution rates for all bases (T92). This was due to this model having the lowest Bayesian/Akaike information criterion (BIC/AIC) when the dataset was tested in JModelTest Version 2.0 (Darriba *et al.* 2012, Guindon and Gascuel 2003). A Maximum-Likelihood (ML) tree was produced in MEGA Version 7.0 (Kumar, Stecher, and Tamura 2015) using the seventeen *H. scobina* COI haplotypes identified here, the sixteen *H. albomarginatum* COI sequences as

well as other Muricid COI sequences acquired from Barco *et al.* (2015) via GenBank and bootstrapped 500 times.

2.2.4: Statistical analyses

Haplotypes were initially identified using DnaSP version 5.10 (Librado and Rozas 2009). Summary statistics for the complete dataset, including the number of segregating sites, haplotype diversity and nucleotide diversity were also recorded. Haplotype distributions and the median-joining network (Bandelt *et al.* 1999) were produced in PopART (version 1.7)(Leigh and Bryant, 2015). To estimate how much of the total haplotype diversity was sampled in this study a rarefaction analysis was performed using Analytic Rarefaction (version 1.3, <https://strata.uga.edu/software/index.html>).

Population structure was initially tested via mixture analysis using the linked-loci clustering model (Corander and Tang 2007) that is implemented in Bayesian Analysis of Population Structure (BAPS)(version 6). Under this model, sequences are assigned to one of (k) populations and then plotted against the sampling sites. To determine the most likely (k) value multiple upper bounds were tested iteratively from k=2 to k=12 and repeated 5 times. A further admixture analysis (Corander and Marttinen 2006) as implemented in BAPS was then used on the mixture analysis results. For the admixture analysis, a k-value of 5 was used (this is the minimum number of sequences within a population needed for it to be considered) for 10,000 iterations, 100 reference individuals were specified with 100 iterations of each reference individual.

The pairwise genetic comparisons between sites to estimate inter-site gene flow were calculated using ϕ_{ST} and bootstrapped with 10,000 permutations in Arlequin (Version 3.5.2.2) (Excoffier and Lischer 2010). Phi statistics such as ϕ_{ST} measure the fixation of alleles ranging from 0(=no fixation) to 1(=complete fixation of alleles between sites/groups/populations) and are analogous to Wright's F statistics. Populations were then assigned to a number of sub-groups for later analysis of molecular variance (AMOVA) testing. AMOVA tests using both pairwise genetic differences and haplotype frequencies were then performed in Arlequin(Version 3.5.2.2) (Excoffier and Lischer 2010) using 10,000 permutations to calculate the variance components and the fixation values ($\phi_{ST}/\phi_{SC}/\phi_{CT}$) within/between regions and across the total population. AMOVA enables comparisons between different

predefined population groups. The most accurate population grouping is determined by the highest significant ϕ_{CT} value as this measures the between-group variation.

Four population structures were tested via AMOVA. The first structure was assigned *a priori* after visually inspecting the haplotype distributions. The second structure tested was two groups split into North and South divided along the 15C⁰ isotherm of the last glacial maximum (LGM)(Barrow and Juggins 2005). The third grouping was produced via spatial analysis of molecular variance (SAMOVA) in SAMOVA 2.0 (Dupanloup *et al.* 2002) (see below). The fourth was derived from the BAPS admixture analysis by assigning each sampling site to its most highly represented population, though some sites had sequences from multiple populations.

SAMOVA is similar to AMOVA as it clusters sample sites in to a predefined number of groups and calculates fixation indices but different in that it attempts to maintain geographic continuity. It can be used to find a structure without the *a priori* bias of standard AMOVA by iteratively testing (k) groups from k=2 to k=N-1 (N=13 sampled sites in this study) and then finding the highest ϕ_{CT} value which shows the optimum group number and structure. Groups of k=2 to k=12 were tested and the optimal number of groups was 7 (ϕ_{CT} =0.7326 p<0.0001). The accuracy of this seven-group structure was then assessed via AMOVA testing.

2.2.5: Demographic history

The demographic history of *Haustrum scobina* was also examined by a number of methods. Arlequin (Version 3.5.2.2) (Excoffier and Lischer 2010) was used to calculate the neutrality test statistics: Tajima's D (Tajima, 1989) and Fu's F (Fu, 1997) for all sites including those sampled by (Barco *et al.*, 2015). Significant results were those with a p value<0.05 for Tajima's D and a p value<0.02 for Fu's F. Non-significant, neutrality test results mean that the null hypothesis of neutral evolution cannot be rejected. Significant negative/positive values can show whether the population has undergone a recent population expansion/contraction or if the gene is under purifying/balancing selection respectively. However, the use of these values to measure historical population size changes can be confounded by the presence of a few highly divergent haplotypes or incomplete sampling of a site.

Mismatch distributions were produced in DnaSP version 5.10 (Librado and Rozas 2009) for all individual sites and all sites combined. These show the distribution for the number of genetic differences for all pairwise comparisons within a given group. A population that has undergone a recent population expansion will have a smooth unimodal distribution. A population at constant size can accumulate many pairwise differences overtime as new mutations arise and will therefore have a bi-modal/multi-modal graph. Populations that had only one haplotype were not mapped this way as the graphs would be uninformative. Additionally, Harpending's raggedness statistic (Harpending *et al.* 1993) was also calculated for every mismatch distribution and the significance was determined by using 1000 coalescent simulations in DnaSP version 5.10 (Librado and Rozas 2009). If this value is significant and sufficiently large (>0.03) it shows that the observed pairwise differences vary significantly from the expected differences under a model of recent population expansion and this null hypothesis can be rejected. While the raggedness statistic may show evidence for historical population size changes, it has less power to detect these changes than other statistics such as Tajima's D or Fu's F which use more information from the dataset than just the pairwise differences (Fu, 1997).

2.3: Results

In this study the DNA sequence of a portion of the COI gene was determined from 376 *Haustrum scobina* samples from twelve sites. No Insertions or deletions were need to align the sequences, but five poor quality sequences were removed. Eight *H. scobina* COI sequences that were deposited into GenBank by Barco *et al.* (2015), which were sampled from Hawera, were added to the data set to give a final aligned set of 379 *H. scobina* COI sequences. Overall the dataset contained seventeen (possibly sixteen, see discussion of haplotype 5) unique haplotypes of which six were private (only found in one location)(Figs. 2.1 and 2.2). There were 51 segregating sites (polymorphic nucleotides in the sequence alignment)(Table 2.1) with 44 being parsimony informative and seven singletons. The levels of haplotype diversity (H_d) varied greatly from five sites, which had a single haplotype ($H_d = 0$) (MAS, PBS, CAS, CPS, MBS in Table 2.1), to Titahi Bay (TBS) that had $H_d = 0.719$. The average nucleotide diversity (π) was 0.00734. All sites had a similar pattern of nucleotide diversity with values ranging from 0 to 0.00775. The rarefaction curve (Fig. 2.3) of sampling sizes vs haplotypes show a trend to asymptote as the sample size increased.

Table 2.1: Table of sample-site locations and respective summary statistics for the 610bp *Haustrum scobina* COI sequences. Key: n=sample size, s= number of segregating sites, N_h= number of haplotypes, N_p= number of private haplotypes, H_d=haplotype diversity, π=nucleotide diversity. The row in bold is the site sampled by (Barco *et al.*, 2015) *=statistically significant p-value

Site code	Site	Latitude/ longitude	n	s	N _h	N _p	H _d	π	Tajima's D (p-value)	Fu's F _s (p-value)	Harpending's raggedness statistic (p-value)
TBS	Titahi Bay	-41.10/ 174.83	23	12	4	0	0.719	0.00775	1.5734 (0.9535)	5.8229 (0.9818)	0.2386 (0.961)
MBS	Makara Bay	-41.22/ 174.71	32	0	1	0	0	0	0.000 (1.000)	0.000 (N/A)	-
KPS	Kau Point	-41.29/ 174.83	32	32	3	0	0.331	0.00652	-1.7935 (0.0192)*	7.9817 (0.9915)	0.2744 (0.970)
CPS	Cape Palliser	-41.61/ 175.29	32	0	1	0	0	0	0.000 (1.000)	0.000 (N/A)	-
FPS	Flat Point	-41.24/ 175.96	31	5	4	1	0.297	0.00092	-1.4875 (0.0545)	0.09123 (0.2119)	0.4935 (0.770)
CAS	Castlepoint	-40.87/ 176.23	31	0	1	0	0	0	0.000 (1.000)	0.000 (N/A)	-
WHS	Whangaehu Beach	-40.40/ 176.63	32	8	3	1	0.179	0.00212	-1.0489 (0.1550)	2.4202 (0.8869)	0.6873 (0.970)
PBS	Pourerere Beach	-40.11/ 176.87	32	0	1	0	0	0	0.000 (1.000)	0.000 (N/A)	-
PAS	Port Ahuriri	-39.48/ 176.90	32	10	4	1	0.464	0.00431	0.1810 (0.6205)	3.6744 (0.9386)	0.2461 (0.902)
MAS	Mahia Peninsula	-39.09/ 177.96	31	0	1	0	0	0	0.000 (1.000)	0.000 (N/A)	-
TAS	Tatapouri Point	-38.65/ 178.15	32	9	4	2	0.333	0.00333	-0.2804 (0.4385)	2.6053 (0.8938)	0.4416 (0.948)
KIS	Kawau Island	-36.40/ 174.85	31	33	4	0	0.501	0.00520	-2.2273 (0.0026)*	4.5289 (0.9596)	0.1975 (0.850)
HBS	Hawera	-39.61/ 174.24	8	3	2	1	0.250	0.00123	-1.4475(0.0634)	1.415 (0.7027)	0.6875 (0.910)
All sites combined			379	51	17	6	0.760	0.00734	-1.0836 (0.1142)	1.451 (0.7275)	0.2601 (0.997)

Table 2.2: Table of AMOVA results from the combined *Haustrum scobina* COI dataset. Key: df=degrees of freedom, Var.C.=variance components,% of Var.=percentage of variance, $\phi_{ST}/\phi_{SC}/\phi_{CT}$ =fixation indices, p=p-value for the fixation index, statistically significant fixation indices are in bold

Groups tested by AMOVA	Measured by using:	Between Populations					Between populations, within predefined groups					Between predefined groups				
		df	Var.C.	% of Var.	ϕ_{ST}	p	df	Var.C.	% of Var.	ϕ_{SC}	p	df	Var.C.	% of Var.	ϕ_{CT}	p
1) <i>a priori</i> groups: (KIS), (TAS,MAS, PAS,PBS,WHS,CAS, FPS) and (CPS,KPS, MBS,TBS,HBS)	Haplotype frequency	366	0.1120	23.01	0.7699	<0.0001	10	0.1706	41.96	0.6036	<0.0001	2	0.2043	41.96	0.4196	0.0017
	Pairwise distances	366	0.7227	27.61	0.7239	<0.0001	10	1.2903	49.29	0.6410	<0.0001	2	0.6049	23.11	0.2311	0.0428
2) 15°C isotherm: (KIS,TAS,MAS,) and (PAS,PBS,WHS,CAS, FPS,CPS,KPS,MBS, TBS,HBS)	Haplotype frequency	366	0.1120	26.65	0.7335	<0.0001	11	0.2788	66.32	0.7134	<0.0001	1	0.0296	7.03	0.0703	0.2629
	Pairwise distances	366	0.7227	26.71	0.7330	<0.0001	11	1.4160	52.33	0.6621	<0.0001	1	0.5675	20.97	0.2097	0.0416
3)SAMOVA groups: (PAS), (KIS), (MAS), (CPS,KPS,MBS), (HBS), (TBS) and (TAS,PBS,WHS, CAS,FPS)	Haplotype frequency	366	0.1120	23.66	0.7634	<0.0001	6	0.0040	0.83	0.0340	<0.0001	6	0.3576	75.51	0.7551	<0.0001
	Pairwise distances	366	0.7227	26.09	0.7391	<0.0001	6	0.0182	0.66	0.0246	0.0016	6	2.0294	73.26	0.7326	<0.0001
4)BAPS groups: (KIS), (MAS,PAS), (TAS,PBS,WHS,CAS, FPS), (CPS,KPS, MBS) and (TBS,HBS)	Haplotype frequency	366	0.1120	24.02	0.7598	<0.0001	8	0.0597	12.79	0.3475	<0.0001	4	0.2947	63.18	0.6319	<0.0001
	Pairwise distances	366	0.7227	25.97	0.7403	<0.0001	8	0.1426	5.13	0.1649	<0.0001	4	1.9170	68.90	0.6890	<0.0001

Table 2.3: Pairwise ϕ_{ST} sample site comparison for the *Haustrum scobina* 610bp COI sequences. ϕ_{ST} values are below the diagonal and respective p-values are above the diagonal, statistically significant ϕ_{ST} values are in bold. Site codes along the top and left correspond to those in table 2.1.

	KIS	MAS	TBS	PAS	CAS	FPS	WHS	PBS	TAS	MBS	KPS	CPS	HBS
KIS		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
MAS	0.385		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
TBS	0.322	0.522		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.018
PAS	0.379	0.513	0.391		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CAS	0.801	1.000	0.621	0.830		0.246	0.237	0.999	0.052	<0.001	<0.001	<0.001	<0.001
FPS	0.771	0.961	0.572	0.799	0.030		0.170	0.026	0.047	<0.001	<0.001	<0.001	<0.001
WHS	0.705	0.897	0.484	0.735	0.060	0.036		0.244	0.710	<0.001	<0.001	<0.001	<0.001
PBS	0.804	1.000	0.625	0.833	0.000	0.031	0.062		0.053	<0.001	<0.001	<0.001	<0.001
TAS	0.640	0.828	0.404	0.672	0.120	0.090	0.0113	0.122		<0.001	<0.001	<0.001	<0.001
MBS	0.804	1.000	0.567	0.826	1.000	0.863	0.742	1.000	0.643		0.056	0.999	<0.001
KPS	0.640	0.767	0.352	0.642	0.491	0.426	0.402	0.495	0.365	0.036		0.057	<0.001
CPS	0.804	1.000	0.568	0.826	1.000	0.863	0.742	1.000	0.643	0.000	0.036		<0.001
HBS	0.555	0.966	0.229	0.623	0.980	0.920	0.836	0.981	0.754	0.981	0.657	0.981	

Table 2.4: Table of the sequences used in the maximum-likelihood phylogeny (Fig. 2.6) including those derived from Barco et al. (2015) via GenBank. Sampling locations, sample sizes for sites used in this study, GenBank accession numbers and museum registration numbers are provided where applicable. Lines in bold show samples from this study, all others were acquired from (Barco *et al.* 2015)

Species	Museum registration number	Sampling location (site code) [sample size]	GenBank accession number
<i>Bedevea paivae</i>	AMS C458273	Australia, Tasmania, Bruny I., Isthmus Bay	FN677412
<i>Bedevea vinosa</i>	AMS C458268	Australia. Tasmania, Diana's Beach, NE of Scamander	FN677413
<i>Poirieria syrinx</i>	M.301837	New Zealand, ENE of Cape Palliser	KP845003
<i>Poirieria zelandica</i>	M.303377/1	New Zealand, Tasman Bay, off Kaiteriteri	KP845005
<i>Haustrum haustorium</i>	UO HH-PB	New Zealand, Pukerua Bay	FN677410
<i>Haustrum lacunosum</i>	UO LL2-SC	New Zealand, St Clair, Dunedin	FN677411
<i>Haustrum albomarginatum</i>	M.301488/1	New Zealand, SW of Hawera	KP844980
	M.301488/2	New Zealand, SW of Hawera	KP844979
	M.301488/3	New Zealand, SW of Hawera	KP844978
	M.301488/4	New Zealand, SW of Hawera	KP844977
	M.301488/5	New Zealand, SW of Hawera	KP844976
	M.301488/6	New Zealand, SW of Hawera	KP844975
	M.301488/7	New Zealand, SW of Hawera	KP844981
	M.301488/8	New Zealand, SW of Hawera	KP844974
	TBA	New Zealand, Kawau Island (KIS) [n=16]	TBA
<i>Haustrum scobina</i>	TBA	New Zealand, Titahi Bay (TBS) [n=23]	TBA
	TBA	New Zealand, Makara Bay (MBS) [n=32]	TBA
	TBA	New Zealand, Kau Point (KPS) [n=32]	TBA
	TBA	New Zealand, Cape Palliser (CPS) [n=32]	TBA
	TBA	New Zealand, Flat Point (FPS) [n=31]	TBA
	TBA	New Zealand, Castlepoint (CAS) [n=31]	TBA
	TBA	New Zealand, Whangaehu Beach (WHS) [n=32]	TBA
	TBA	New Zealand, Pourerere Beach (PBS) [n=32]	TBA
	TBA	New Zealand, Port Ahuriri (PAS) [n=32]	TBA

Table 2.4: continued

Species	Museum registration number	Sampling location (site code) [sample size]	GenBank accession number
<i>Haustrum scobina</i>	TBA	New Zealand, Mahia Peninsula (MAS) [n=31]	TBA
	TBA	New Zealand, Tatapouri Point (TAS) [n=32]	TBA
	TBA	New Zealand, Kawau Island (KIS) [n=31]	TBA
	M.301489/1	New Zealand, SW of Hawera	KP844991
	M.301489/2	New Zealand, SW of Hawera	KP844990
	M.301489/3	New Zealand, SW of Hawera	KP844989
	M.301489/4	New Zealand, SW of Hawera	KP844988
	M.301489/5	New Zealand, SW of Hawera	KP844985
	M.301489/6	New Zealand, SW of Hawera	KP844992
	M.301489/7	New Zealand, SW of Hawera	KP844986
	M.301489/8	New Zealand, SW of Hawera	KP844987

The network analysis showed that the majority of the sequences (>65%) belonged to two large haplogroups that were closely related (Fig. 2.2). The first of these large haplogroups (haplotypes 1, 3, 4, 13) formed a starburst-pattern centred on haplotype 1 (n=144) in Fig 2.2. The second large haplogroup contained haplotypes 2 (n=105) and 10 (n=10). Haplotype 2 was two mutational steps away from haplotype 1. Both haplotypes 1 and 2 had strong regional associations with the East and South coast of the North Island respectively. Four sites in this region were completely fixed for one of the two common haplotypes with sites PBS and CAS only having haplotype 1 and CPS and MBS only having haplotype 2 present. Only one site (FPS) had both haplotypes 1 and 2 present, which could be a contact zone between the east coast/south coast regions.

Most of the other sequences comprised a more broadly distributed network of nine haplotypes (haplotypes 6, 7, 8, 9, 12, 14 and 16) separated from the two previously mentioned haplogroups by a minimum of six mutational steps. Some of these haplotypes did have strong associations to specific sites. For example, Mahia Peninsula was fixed for haplotype 9 and all of the haplotype 12 sequences were found in Port Ahuriri (PAS) (Figs. 2.1 and 2.2). However these sites were not always contiguous and the distributed appearance of this part of the median-joining network may reflect the limited sampling range, which could mean intermediary haplotypes in northern populations were missed in the sampling.

One highly divergent haplotype (haplotype 5) was sequenced from both Kau Point (KPS) in Wellington (n=2) and at Kawau Island (KIS) in the Hauraki Gulf (n=1)(Fig. 2.2). Initially this sequence was thought to be from the closely related species *Haustorium albomarginatum* as these two species can be easily misidentified because of their similar shell structures. After closer examination of the original shells and by comparing haplotype 5 to previously sequenced *H. albomarginatum* COI haplotypes, it was determined that the divergent haplotype was most likely from *Haustorium scobina* but this is not a certainty (see discussion 2.4.3). The minimum sequence difference between any *H. scobina* COI sequence and any *H. albomarginatum* COI sequence was 28 mutational steps (4.59% sequence difference). Haplotype 5 was 30 mutational steps away from the most similar *H. scobina* COI sequence (4.92% sequence difference) and 40 mutational steps away from the most similar *H. albomarginatum* COI sequence (6.56% sequence difference).

The BAPS mixture analysis divided the 379 DNA sequences into six groups and the second admixture analysis reduced this to five groups (Fig. 2.4) represented by different colours. All sampled sites had at least 50% of their sequences fall into a single group with many sites having all of their sequences fall into a single group. While there was evidence of admixture between sites, each site could still be assigned to the group that it was most highly represented in. This would be used as a set of groups for use in the AMOVA analysis. The population structure this analysis produced divided the sites along the East/South coast in to two regional groups dominated by either Haplotype 1 or 2 across the grouped sites.

The SAMOVA test found that the optimal group number was seven ($\phi_{CT}=0.7326$ $p<0.0001$). Five of the sites made up five individual groups and the other two groups were identical to the two East/South coast groups found by BAPS: (KPS, CPS, MBS) and (FPS, CAS, WHS, PBS, TAS).

AMOVA tests were performed on four different population structures using both haplotype frequencies and genetic distances to determine the most appropriate (Table 2.2). The first was a predefined, *a priori* structure based on the distribution map that had a significant $\phi_{CT}=0.2311/0.4196$. The second structured all sequences in to two groups split by the 15°C isotherm of the LGM and returned the lowest ϕ_{CT} values (<0.21) showing this was not an appropriate structure to explain the dataset. The third structure produced both the highest significant ϕ_{CT} value ($\phi_{CT}=0.7326/0.7551$) for variance between groups and the lowest within group variation ($\phi_{SC}=0.0246/0.0340$) for this structure around 75% of the variation occurred between groups and around 1% of the variation occurred within groups. This result provides strong evidence that the SAMOVA grouping is the most appropriate structure for this dataset. The fourth group derived from the BAPS analysis had the second highest ϕ_{CT} ($\phi_{CT}=0.6319/0.6890$) and second lowest ϕ_{SC} ($\phi_{SC}=0.3475/0.1649$) making this an acceptable structure for this dataset but not as appropriate as that found via SAMOVA.

The pairwise ϕ_{ST} comparisons (Table. 2.3) showed many instances of statistically significant genetic differentiation (the majority of ϕ_{ST} indices being >0.5). For example, there were highly divergent sites such as Mahia Peninsula (MAS) which ranged from $\phi_{ST}=0.385$ to $\phi_{ST}=1.00$. Within the two southern and eastern groupings identified by the BAPS ϕ_{ST} values were considerably lower and sometimes $\phi_{ST}=0$ such as between Pourerere Beach (PBS) and Castlepoint (CAS).

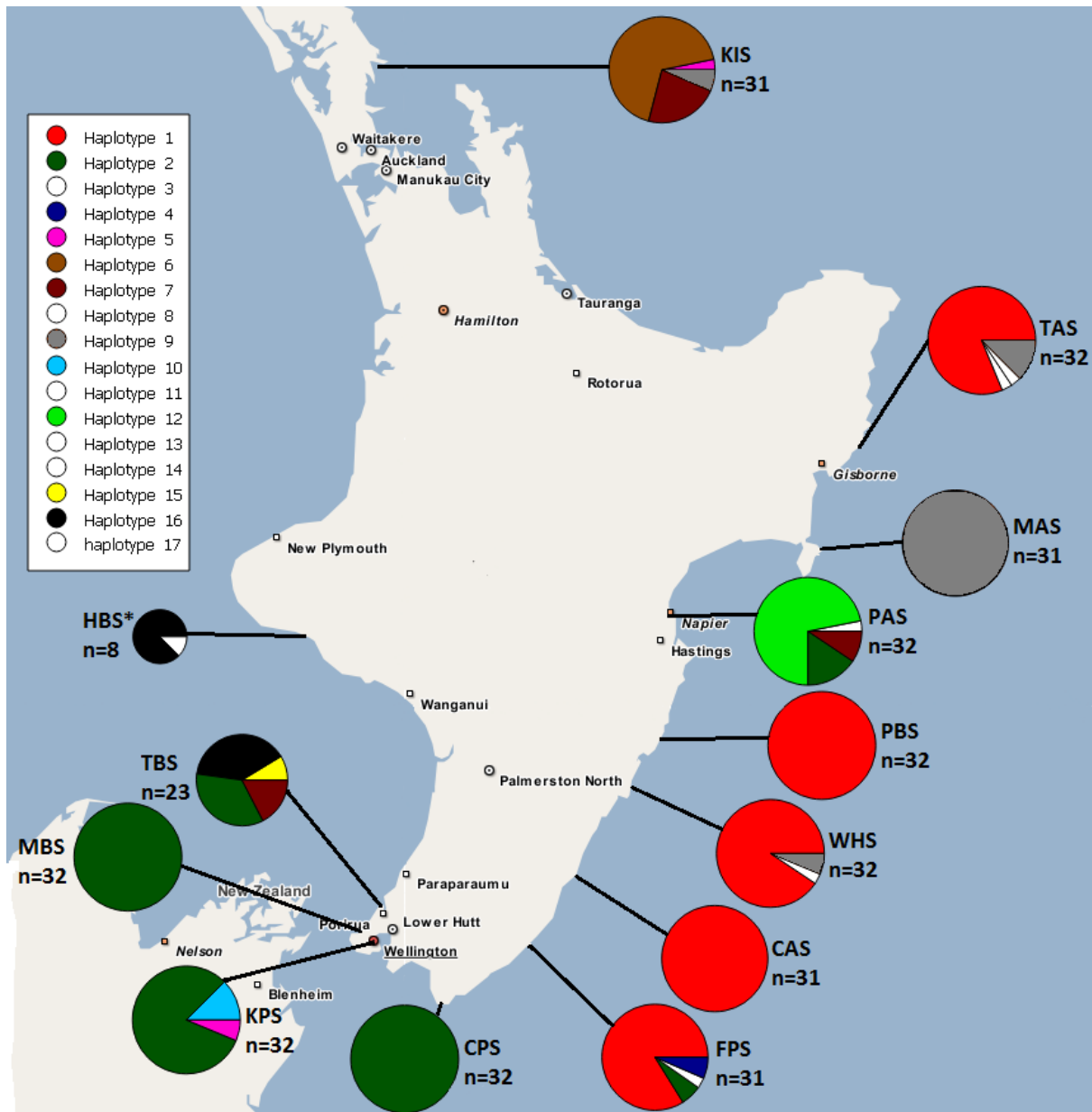


Figure 2.1: Map showing the geographic distribution of *Haustorium scobina* COI haplotypes. Pie chart diameters correspond to sample sizes (n) which are also provided. Private haplotypes are coloured white. Site codes correspond to those in Table 2.1. *=This site was sampled by (Barco *et al.*, 2015)

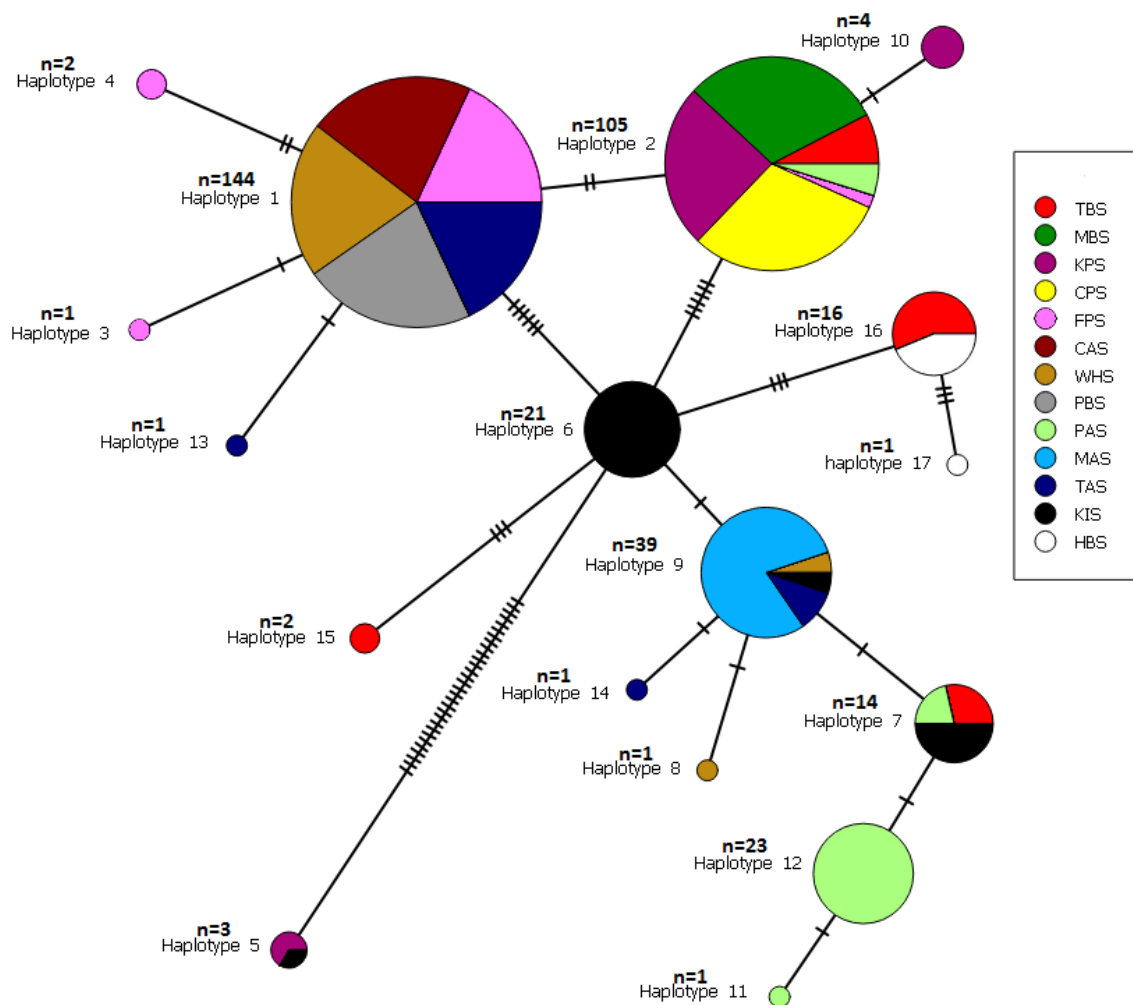


Figure 2.2: COI median-joining network for *Haustorium scobina*. Hatch marks signify the number of mutational steps between haplotypes. The areas of the pie charts correspond to numbers of haplotypes (n) which are also provided. Site codes in the key correspond to those in Table 2.1

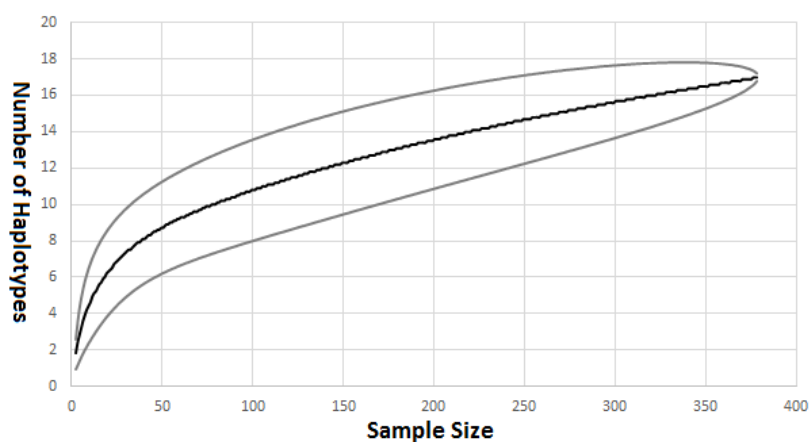
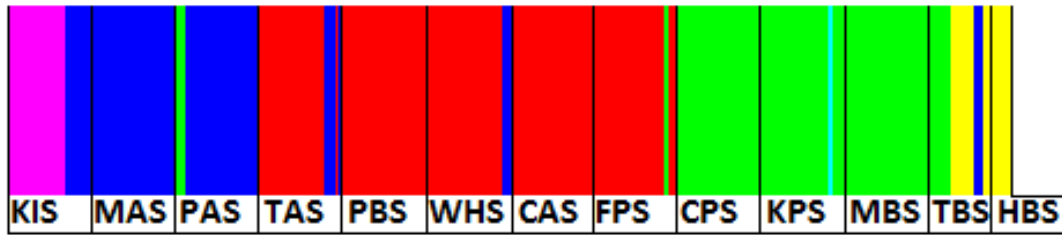


Figure 2.3: Rarefaction curve for all 610bp *Haustorium scobina* COI haplotypes sampled. The black line shows the expected samples at each sample size while the grey lines show the upper/lower 95% confidence intervals

A)



B)

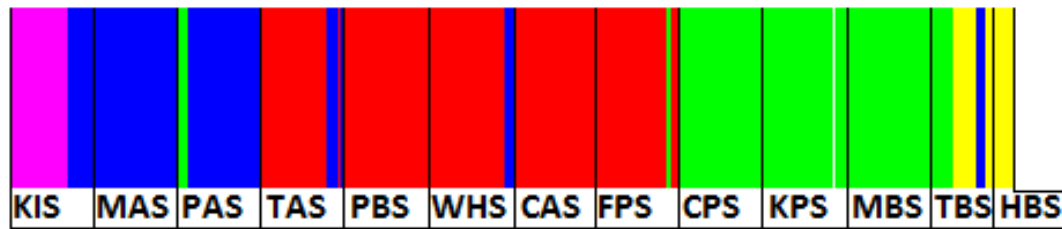


Figure 2.4: A) A linked-loci clustering of the 379 *Haustrum scobina* COI sequences produced using Bayesian Analysis of Population Structure (BAPS) showing six groups B) Admixture analysis of the linked loci-clustering showing five groups. Sequences were coloured to show which group they are assigned to by each analysis. Site codes listed at the bottom correspond to those used in Table 2.1

Table 2.5: Haplotype composition for the complete *Haustrum scobina* COI dataset. Total values are provided in grey along the bottom/right edges. *=site sampled by (Barco *et al.* 2015)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	total
KIS	-	-	-	-	1	21	7	-	2	-	-	-	-	-	-	-	-	31
TAS	26	-	-	-	-	-	-	-	4	-	-	-	1	1	-	-	-	32
MAS	-	-	-	-	-	-	-	-	31	-	-	-	-	-	-	-	-	31
PAS	-	5	-	-	-	-	3	-	-	-	1	23	-	-	-	-	-	32
PBS	32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32
WHS	29	-	-	-	-	-	-	1	2	-	-	-	-	-	-	-	-	32
CAS	31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	31
FPS	26	2	1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	31
CPS	-	32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32
KPS	-	26	-	-	2	-	-	-	-	4	-	-	-	-	-	-	-	32
MBS	-	32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32
TBS	-	8	-	-	-	-	4	-	-	-	-	-	-	-	2	9	-	23
HBS*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	1	8
total	144	105	1	2	3	21	14	1	39	4	1	23	1	1	2	16	1	379

The neutrality tests were statistically insignificant for eleven of the thirteen sites (Table 2.1). Insignificant neutrality test results indicate that the null hypothesis of a neutrally evolving COI gene cannot be rejected. However this does not reject the possibility of non-neutral evolution. The two significant, negative Tajima's D values were for the KPS ($D=-1.7935$ $p=0.0192$) and KIS ($D=-2.2273$ $p=0.0026$) sites. The Tajima's D statistic is sensitive to demographic processes in a population and the negative values could suggest that the sites have undergone a recent population expansion. However it is more likely that these results were significant due to confounding caused by the presence of the highly divergent, haplotype 5 at both these sites. This possibility is supported by the respective Fu's F values being insignificant.

The mismatch analysis was performed on all sites combined and all individual sites with a haplotype diversity (H_d)>0 (Fig. 2.5). All individual mismatch analyses showed similar expected and observed values aside from Titahi Bay which had a multi-modal distribution. The corresponding Harpending's raggedness statistics were all insignificant (all p -values>0.77) meaning the null hypothesis of recent population expansion cannot be rejected (Table 2.1) yet the lack of significance makes these values relatively uninformative.

The COI maximum likelihood phylogeny (Fig. 2.6) showed that the *H. albomarginatum* sequences formed two clades each with strong bootstrap support, one containing the eight samples from Hawera collected by (Barco *et al.* 2015) (bootstrap=1.00), the other containing the sixteen *H. albomarginatum* samples from Kawau island collected in this study (bootstrap=0.998)(see Table 2.4 for sequences used). Sixteen of the seventeen *H. scobina* COI haplotypes formed a separate well supported clade (bootstrap=0.998) with the highly divergent, putative *H. scobina* COI haplotype 5 being placed outside of the other three previously mentioned clades. The bootstrap support values for the relationships between the two *H. albomarginatum* clades, the *H. scobina* clade and the putative *H. scobina* COI haplotype 5 were all weak ranging from 0.344 to 0.568.

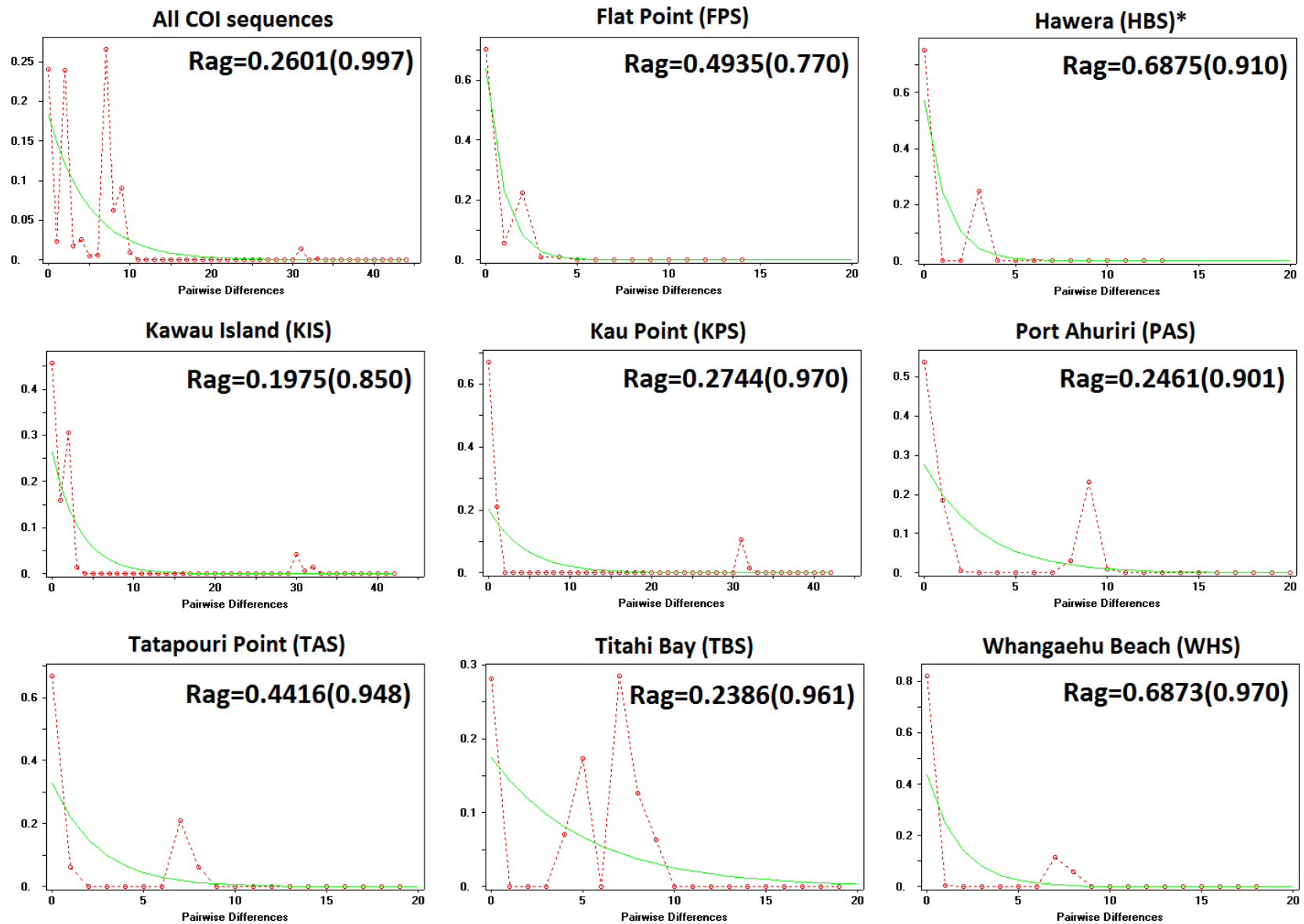


Figure 2.5: Mismatch distributions for the individual sites and all sites combined. Green lines show the expected pairwise differences while red lines show observed pairwise differences. Harpending's raggedness statistics are provided with respective p-values in brackets. NB: the sites MBS, CPS, CAS, PBS and MAS are not shown individually as they only had a Haplotype diversity (H_d)=0 and were uninformative.

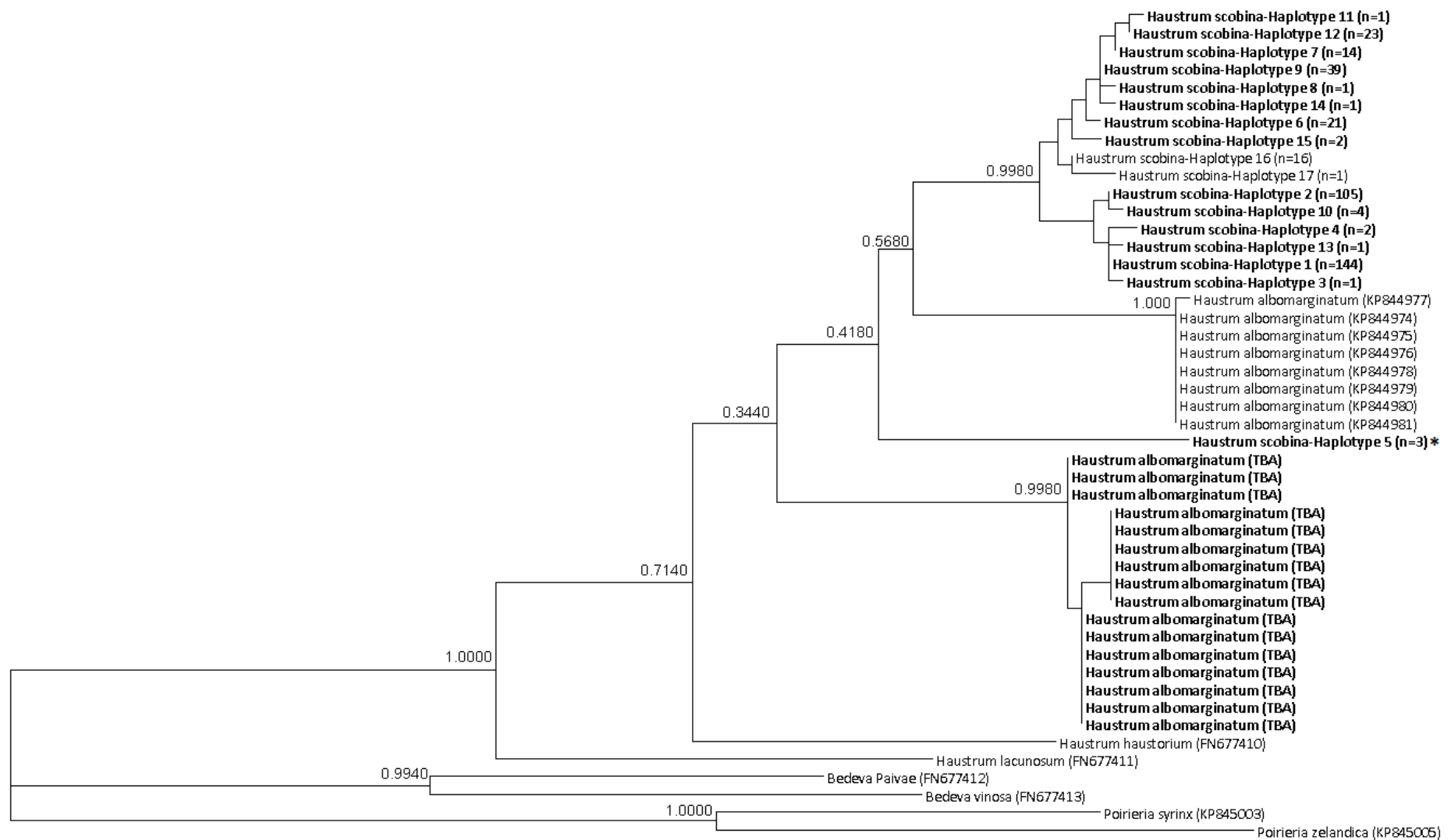


Figure 2.6: Maximum Likelihood tree for (610bp) COI sequences showing the 17 *Haustum scobina* haplotypes and other Muricid sequences acquired from Barco *et al.* (2015) via GenBank. Node labels show bootstrap support values. Bold labels denote sequences during this study all others are derived from Barco *et al.* (2015). The number of samples with each *H. scobina* haplotype is provided in brackets or for samples from Barco *et al.* (2015) GenBank accession numbers are provided in brackets. *= Haplotype 5 was putatively assigned to *H. scobina* (see discussion 2.4.3 for the reasons why this may not be the case). NB: Haplotypes 16 and 17 aren't in bold as they were initially identified amongst *H. scobina* samples from Barco *et al.* (2015).

2.4: Discussion

2.4.1: Distributions and sampling of *Haustum scobina* and *Haustum albomarginatum*

H. scobina was found in high abundance at all sites sampled, including rocky reef environments such as Cape Palliser (CPS) to artificial breakwaters such as at Port Ahuriri (PAS) from mid to low tide. Previous studies have reported that *H. scobina* is distributed across the entirety of the North, South and Chatham Islands of New Zealand (see Fig. 29 in Tan 2003). The rarefaction curve (Fig. 2.3) indicated that sampling >30 *H. scobina* samples per site was enough to capture most of the diversity, however a larger sample size would most likely lead to the discovery of some additional rare haplotypes. Despite sampling across apparently suitable habitats, *H. albomarginatum* was only found at Titahi Bay (TBS) and at Kawau Island (KIS) occurring sympatrically with *H. scobina* though further up the shore, in high abundance. It is unknown whether this apparent distribution gap from Gisborne to Wellington is a sampling artefact because these two *Haustum* species have frequently been synonymised (Tan 2003, Barco *et al.* 2015) and *H. albomarginatum* may have mistakenly been identified as *H. scobina* in previous samples taken from this region. If there a gap in the distribution of *H. albomarginatum* it is unlikely that this is due to competitive exclusion by *H. scobina* as both species occur sympatrically in high abundance.

2.4.2: Genetic diversity and population structure

Sites across the south-eastern coast of the North Island of New Zealand between Pourerere Beach (PBS) and Makara Beach (MBS) had very low haplotype diversities ranging from $H_d=0$ to $H_d=0.331$ compared to other sites such as Titahi Bay ($H_d=0.719$). Cold sea temperatures during the last glacial maximum (LGM) could have restricted mollusc species including *H. scobina* to warmer northern climates (Newnham *et al.* 1999). This low haplotype diversity could indicate that these south-eastern populations originated via recolonisation from more northern sites as sea temperatures rose following the LGM (Fraser *et al.* 2012). However the AMOVA analysis (Fig. 2.2) showed that a grouping divided by the 15°C isotherm was less appropriate than the SAMOVA grouping, indicating the lack of diversity in south eastern sites may have resulted from localised extinctions/recolonisation that is unrelated to the

LGM. Despite the apparently low dispersal potential of direct-developers, *H. scobina* has been shown to rapidly recolonise suitable habitats significantly affected by tributyltin (TBT) poisoning and the concurrent induction of imposex (development of male gonads in females) that leads to widespread infertility (Jones and Ross 2018). This finding could potentially indicate that any recolonisation may have occurred rapidly as new, suitable habitats became available following either the LGM or a local extinction contributing to the lack of haplotype diversity via successive founder events (Mayr 1942).

The populations between Pourerere Beach and Makara Beach were geographically structured into two contiguous regions. The southern sites of (MBS), (KPS) and (CPS) were dominated by haplotype 2 and the eastern sites of (FPS), (CAS), (WHS) and (PBS) were dominated by haplotype 1 (Fig. 2.1). The BAPS (Fig. 2.4), SAMOVA and AMOVA results (Table 2.2) all supported this geographic structuring. Interestingly, the SAMOVA analysis included the Tatapouri Point (TAS) site, dominated by haplotype 1, within the eastern group but not the intermediary sites of (PAS) and (MBS) despite the fact that SAMOVA analyses attempt to maintain geographic homogeneity (Dupanloup *et al.* 2002). This unexpected grouping could be explained by either leapfrog dispersal via rafting (Thiel and Haye 2006), genetic isolation of intermediary sites or events subsequent to post-glacial recolonisation such as environmental disturbances leading to local extirpation. It is highly unlikely that genetic homogeneity within these two regions is maintained by regular gene flow as direct-development likely limits migration between distant sites (Palumbi 1994) and there is an absence of contiguous suitable habitats (i.e. solid substrates such as rocky reefs) between the sites sampled. The only site in this region where haplotypes 1 and 2 were both present was Flat Point (FPS) indicating that this may be contact zone between southern and eastern populations that have undergone a range expansion. It is also worth noting that at Flat Point (FPS) the north and south bound currents along the southeast coast of the North Island meet and then move out to sea, potentially limiting migration and contributing to a phylogeographic break.

The pattern of genetic diversity and specifically the location of the phylogeographic break found across this region was very similar to that found in *C. maculosa* populations (Fleming *et al.* 2018, Dohner *et al.* 2018). Interestingly, the fine-scale sampling of Dohner *et al.* (2018) revealed that *C. maculosa* populations also formed a contact zone at Flat Point between two

geographic regions dominated by single haplotypes. If the genetic structures of populations of *C. maculosa* and *H. scobina* are determined primarily by the environmental conditions in this region then it would be logical to hypothesise that the same phylogeographic structure would be seen in future studies of similar New Zealand direct-developers using the same sampling regime. One potential explanation for the similarity in phylogeographic break locations could be that both species disperse between distant sites primarily via sporadic rafting events mediated by the same prevailing ocean currents (Fig 1.2). However this possibility is speculative and proving this would require direct evidence of rafting in both species and a more detailed understanding of local oceanography.

A combination of successive founder events via sporadic migrations (Mayr 1942), local genetic drift and density-dependent blocking processes limiting successful colonisation by later migrants (Waters *et al.* 2013) could have led to large stretches of coastline being dominated by a single haplotype as was seen in the sampled sites. Low diversity due to isotherm tracking (colonisation following temperature changes) could be maintained over time if the rates of migration were low enough to limit gene flow from more genetically diverse source populations. This hypothesis is further supported by the presence of two rare haplotypes (haplotypes 3 and 4 Fig 2.2) that were only found at Flat Point (FPS) which were closely related to the dominant haplotype at Flat Point (Haplotype 1) indicating that this population has existed long enough for new mutations to arise in COI haplotypes. Other private haplotypes may exist at such low frequencies that the sampling may not have been sufficient to detect them (Fig. 2.3).

The most common haplotype at Hawera (haplotype 16), Mahia Beach (haplotype 9), Port Ahuriri (haplotype 12) and Kawau Island (haplotype 6) were part of a loosely distributed cluster of haplotypes that were at least six mutational steps away from haplotypes 1 and 2. With the exception of Port Ahuriri in Napier this loosely distributed cluster appears to be regionally associated with more northern sites. However further sampling across the northern half of the North Island is needed to prove this regional association. Mahia Beach was completely fixed for haplotype 9 which was also found at low frequencies at Tatapouri Point and Whangaeahu Beach. Mahia Peninsula used to be an island until wave diffraction over tens of thousands of years led to the formation of a tombolo (Sauvage *et al.* 1954) connecting it to the mainland as the peninsula that is there today. If *H. scobina* was present

on Mahia when it was an island, the isolation of the site could have allowed local genetic drift to act independently and thus produce the haplotype fixation observed today. Another possibility is that the sampled Mahia population arose via a recent founder event from another site possibly following the tombolo formation. In this case the absence of diversity could have been maintained via limited ongoing migration to Mahia, local genetic drift and density-dependent blocking processes (Waters *et al.* 2013). Given that mitochondrial DNA markers reach reciprocal monophyly quicker and historical effects tend to persist for a longer time than in nuDNA markers (Hellberg *et al.* 2002), nuDNA markers may show a different pattern across this region that better reflects contemporary patterns of gene flow.

2.4.3: Placement of haplotype five

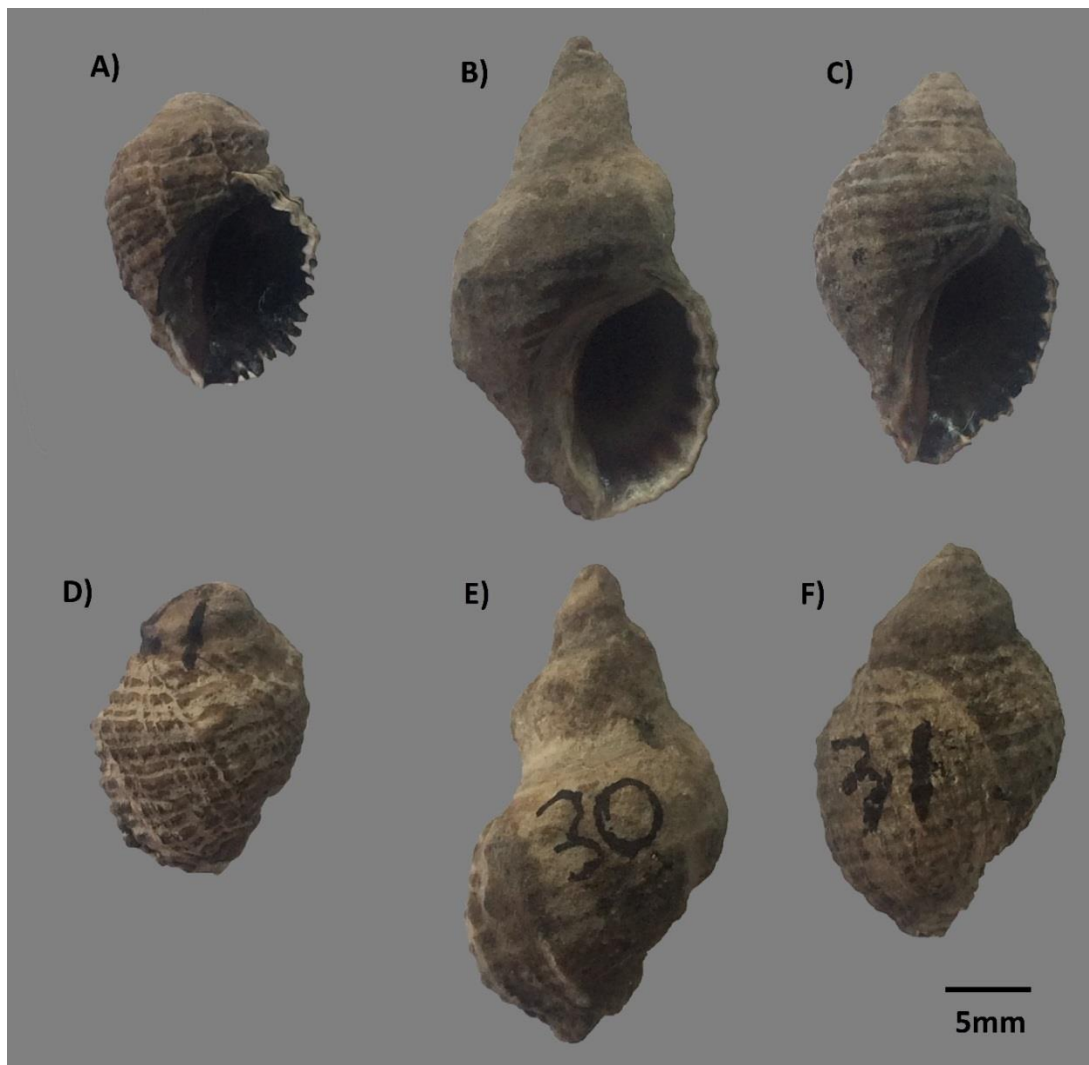


Figure 2.7: Front and rear views of the shells from the three samples with the highly divergent haplotype 5. (A,D)= sample KIS 21 from Kawau Island, (B,E)= sample KPS 30 from Kau Point, (C,F)= sample KPS 31 from Kau Point

The three samples with haplotype 5 (see Fig 2.7 for images of the shells) were putatively assigned as *H. scobina* samples on the basis of genetic similarity as haplotype 5 had a sequence difference of 4.92% with the most similar *H. scobina* haplotype and a sequence difference of 6.56% with most similar *H. albomarginatum* COI sequence. When initially identifying samples the character trait used to differentiate between *H. scobina* and *H. albomarginatum* was the presence of nodules on *H. scobina* shells (see Fig. 2.8A) that was absent in the smooth-surfaced *H. albomarginatum* (Tan 2003). In many cases weathering due to wave action led to the nodules on *H. scobina* being worn down (see Fig. 2.8B for an example of this). This led to a lack of consistent trait differences and made accurate identification of *H. scobina* and *H. albomarginatum* difficult. During the present study, sixteen *H. albomarginatum* from Kawau Island were unintentionally sequenced, because it was difficult to distinguish them from *H. scobina*. The lack of clear differences between species could have contributed to the historical synonymisation of these two species (Tan 2003). Weathering on the shells of the three samples with haplotype 5 (Fig. 2.7) made the shells relatively uninformative in species identification. The assignment of these samples to *H. scobina* due to slight differences in genetic similarity is contentious and it is entirely possible that these were in fact *H. albomarginatum*. *H. albomarginatum* were found at Kawau Island and while none were sampled from Kau Point this may reflect sampling effort rather than absence of this species at this site. Given the wide distributions of both species and the limited sampling range of this thesis it is possible that further sampling of either species could reveal intermediate haplotypes that are more similar to haplotype 5 potentially clarifying the assignment to a species. A third possibility is that these three highly divergent samples represent a cryptic species. However this is highly unlikely as a cryptic species would be indistinguishable from *H. scobina* and extensive sampling for *H. scobina* across the south of North Island to the east and West of the Kau Point site found no other haplotypes that were as similar to haplotype 5. Therefore the most likely explanation is that haplotype 5 is either a *H. scobina* or *H. albomarginatum* haplotype potentially originating from an un-sampled location.

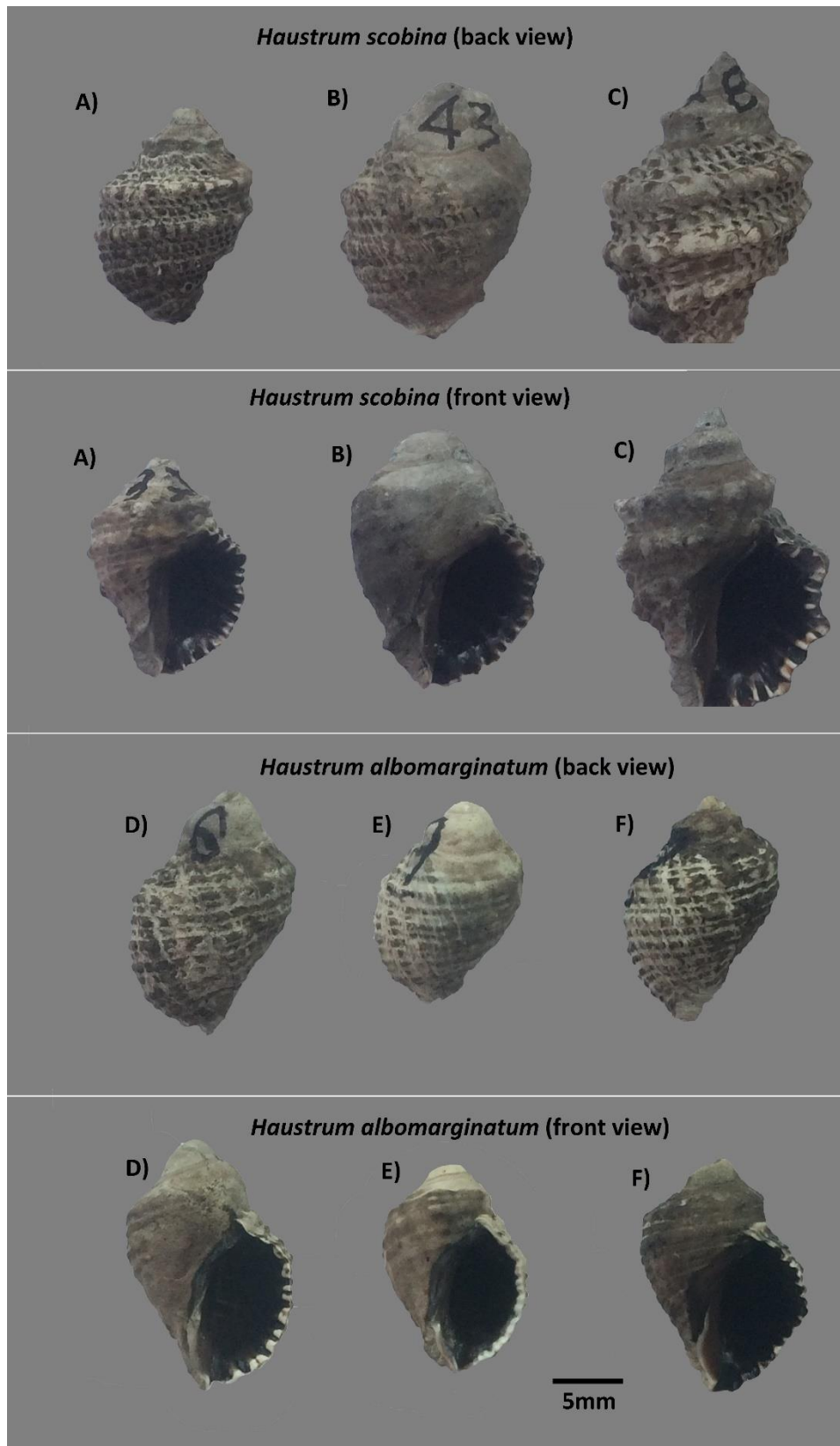


Figure 2.8: Front and back views of a reference selection of *H. scobina* and *H. albomarginatum* shells collected from Kawau Island that had their assignment to a species confirmed via COI sequencing. (A-C)= three *H. scobina* shells from Kawau Island, (D-F)= three *H. albomarginatum* shells from Kawau Island

2.4.4: Long distance dispersal events

Possible evidence of multiple long-distance dispersal/translocation events was found in the Kau point, Whangaehu Beach and Port Ahuriri sites (Figures 2.1 and 2.2). Assuming that haplotype 5 is a *H. scobina* haplotype the presence of this at Kau point in two sampled individuals may indicate that these arrived via human-mediated translocation. The absence of any similar haplotypes in other sampled site aside from Kawau Island indicates that long-distance dispersal is an unlikely explanation for its presence here. Given that this is a busy port region, human-mediated translocation is a more logical explanation for its presence here. Evidence of inadvertent human-mediated translocation has been found in a phylogeographic study of another marine intertidal gastropod *Littorina saxatilis* (Panova *et al.* 2011) indicating that the occurrence of a similar event may at least be possible in *H. scobina*.

Port Ahuriri in Napier also showed evidence of similar translocations where haplotype 2 was found in five sampled individuals (Table 2.5). As was previously mentioned, haplotype 2 was the most frequently found haplotype in the southern sites (MBS, KPS, CPS) and was virtually absent from the intermediate eastern sites between Flat Point and Port Ahuriri (CAS, WHS, PBS). The presence of a haplotype with a different regional association in a busy port makes human-mediated translocation the most likely explanation. Environmental disturbances due to human activity such as TBT-poisoning may also help to explain why the genetic diversity observed at this site was markedly different from all nearby sites. Smith and McVeagh (1991) found that TBT-poisoning had widely induced imposex (the development of male gonads in females) in *H. scobina* females at nearly every site with boat activity that they sampled significantly reducing the survivability of these populations. If imposex induction had repeatedly occurred at Port Ahuriri the resulting genetic bottlenecking could have significantly altered the distribution of haplotypes relative to other nearby sites with less boat activity.

H. scobina generally inhabits areas with solid substrates. While *H. scobina* do not attach strongly to surfaces (personal observation) it is possible that they could become fouled to the underside of ships or fishing equipment such as crayfish pots. Another possible vector could be as dry ballast (Hewitt *et al.* 2009), rocks were used as ballast for ships in New

Zealand as late as the 1900s (Moore and Kenny 1986) and this could explain why an apparently southern haplotype was found in a more northern port and not found in other nearby sites. Similar evidence of a human-mediated translocation was found in *C. virgata* at the busy port of Napier (Fleming *et al* 2018). If human mediated translocation events are common amongst direct-developers it is likely that further studies will find evidence of this primarily at port regions.

As previously mentioned, the samples from the Mahia population (MAS) all had the COI haplotype 9 and haplotype 9 was also present in the Tatapouri Point and Whangaehu beach sites. While dispersal via crawling could explain the presence of haplotype 9 at Tatapouri and Mahia the absence of haplotype 9 at sites between Mahia and Whangaehu beach (PAS and PBS sites, Fig. 2.1) indicates that long-distance dispersal may have occurred between Mahia and Whangaehu beach. *H. scobina* generally lay eggs in irregular groups on hard substrates (Carrasco and Phillips 2014) though it is possible that eggs may become detached due to wave action or be laid on a potential vector such as the bull kelp *D. antarctica* (Bussolini and Waters 2015). Detached eggs or those laid on a vector could then be dispersed along a current such as the East Cape current or the Wairarapa coastal current (Fig. 1.2). Additionally, adult *H. scobina* may move on to buoyant natural vectors such bell kelp, floating driftwood or pumice stones, though this has not been directly observed. Possible evidence for long-distance dispersal has been found in a number of other New Zealand species (Ross *et al.* 2009) including the New Zealand mud snail *Zeacumantus subcarinatus* (Keeney *et al.* 2013). Long distance dispersal may also help to explain the observed genetic homogeneity between distant *H. scobina* populations where dispersal via crawling may be insufficient to maintain homogeneity and where intermediate sites may have markedly different haplotype distributions (e.g. Port Ahuriri). As *H. scobina* generally hatch an average of ten juveniles per egg capsule (Carrasco and Phillips 2014) the movement of only a few egg capsules may be sufficient to provide enough migrants for a distant genetic lineage to become established at a new site. Furthermore, direct developers may have an advantage in establishing themselves following a long distance dispersal/translocation event as other migrants and thus potential mates are likely to remain in close proximity to each other following such an event.

Chapter 3: Conclusions and future research

3.1: Conclusions

H. scobina populations in the southeast of North Island showed significantly reduced haplotype diversity relative to other sites such as Titahi Bay and Kawau Island. This north-south contrast in haplotype diversity indicates that these populations resulted from recolonisation following local extirpations or possibly following the LGM. *H. scobina* populations in this region were geographically structured in to two regional groups that was concordant with the geographic structuring seen in *C. maculosa* populations (Fleming *et al.* 2018, Dohner *et al.* 2018). This result was unexpected as the geographic structuring in *C. maculosa* initially appeared to result from the confluence of two southward range expansions following the LGM rather than from environmental factors. One potential explanation for the concordance between these two species could be that recolonisation following local extinctions or a southward range expansion following the LGM occurred primarily via sporadic rafting between sites (Fraser *et al.* 2012). The prevailing currents in this region could support this explanation as a southwards current from the circular Wairarapa eddy meets the northwards Southland current (Fig. 1.2) at the roughly same location as the confluence of the two regional *H. scobina* and *C. maculosa* groups.

The strong bootstrap support values in the maximum likelihood tree (Fig. 2.6) for separate clades reinforces the division of *H. scobina* and *H. albomarginatum* as separate species proposed by (Barco *et al.* 2015) and other authors. Considering the significant morphological similarity between the two species (Tan 2003) this result shows the value in using genetic studies to provide an independent line of evidence in situations where the delineation of species remains contentious. The placement of three samples with a highly divergent haplotype within *H. scobina* remains uncertain and further sampling of *H. scobina* populations across a wider range would be required to determine if this placement is correct.

Evidence of multiple long distance dispersal/translocation events was found though the exact means of dispersal remains uncertain. The observation that most of these events appear to have occurred in busy port regions indicates that inadvertent human mediated translocations may have occurred. Long distance dispersal may also explain the significant

lack of haplotype diversity observed across the south and east of the North Island. This could occur either through enabling rapid recolonisation following range restriction/ local extinction or by maintaining gene flow via sporadic migration as the migration of only a few individuals per generation may be enough to maintain genetic homogeneity between sites such as those found on the south-eastern North Island (Ovenden 2013).

3.2: Future research

A number of unexpected results found during this thesis research indicate that further research into *H. scobina* and other direct developing marine molluscs will produce some more interesting findings. To further investigate whether the observed lack of haplotype diversity in *H. scobina* populations across the southeast of the North Island is a result of post-glacial recolonisation, further sampling across the entirety of its distribution would be required. The potential location of a low-high haplotype diversity transition is presently unknown. If a significant transition across a short distance was found this would provide strong evidence for post-glacial recolonisation occurring in *H. scobina*. Sampling for such a study would likely be straightforward as it appears that *H. scobina* occurs in high abundances at rocky reef environments and could be thus be undertaken at the other sites used in Fleming *et al.* (2018) that were not sampled in this thesis. Sampling across sites used in studies of other direct developing marine molluscs is recommended in order to develop a more effective comparative phylogeographic dataset. Further evidence for post glacial recolonisation in *H. scobina* populations could be provided by using genetic markers with higher levels of variability, such as microsatellite DNA loci. High-diversity markers generally reveal more about contemporary patterns of gene flow as signatures of historical events such as range expansions tend to dissipate quicker (Hellberg *et al.* 2002). If a high diversity marker showed that *H. scobina* populations were genetically subdivided across the southeast of the North Island this would indicate that the observed lack of haplotype diversity was a result of historical events rather than due to regular, contemporary gene flow maintaining genetic homogeneity.

A phylogeographic study of *H. albomarginatum* in addition to a wider ranging study of *H. scobina* may also be useful in clarifying the correct species placement for the samples with the highly divergent haplotype 5 that was found in this thesis. While genetic similarity

indicated that these samples were most likely *H. scobina* this placement remains uncertain. Sampling across the entirety of the distribution of both species would provide a better characterisation of the haplotype diversity present within both species and the potential location of source populations for long distance dispersal/translocation events. Given that the strong morphological similarities between *H. scobina* and *H. albomarginatum* (Tan 2003) may make accurate species identification during sampling difficult, it may be more practicable to perform a phylogeographic study of both species in parallel and to use the genetic differences among samples to identify each species.

While this thesis extends the inferences that can be made from the results of other studies, the number of phylogeographic studies of direct developing marine molluscs in New Zealand (Keeney *et al.* 2009, Keeney *et al.* 2013, Fleming *et al.* 2018, Dohner *et al.* 2018, Walton 2017) remains small. The finding that *H. scobina* and *C. maculosa* populations (Fleming *et al.* 2018) around the southeast of the North Island both display low haplotype diversity may indicate both species have undergone similar post-glacial recolonisations. However the true extent of isotherm tracking amongst direct developers in New Zealand will not be understood until a wider range of comparable studies can provide a more comprehensive picture of the general phylogeographic trends. Further studies of direct developers could also reveal more instances of long distance dispersal/translocation. It is logical to assume that the limited dispersal abilities of direct developers would usually produce a pattern of genetic isolation by distance (Palumbi 1994) that would contrast greatly with the genetic signatures of sporadic long distance dispersal/translocation events.

Even though detection of long distance dispersal/translocation events may be possible in studies of other direct developers, assigning causality of these events to either natural or human mediated mechanisms remains difficult. One method to overcome this may be to specifically study species that show strong associations with possible natural vectors.

Members of the gastropod subfamily Cantharidinae in New Zealand frequently associate with the buoyant bull kelp *D. antarctica* (Donald and Spencer 2016) and live *Cantharidus antipodum* specimens were found engaging in long distance rafting on bull kelp by Fraser *et al.* (2011). Therefore, phylogeographic studies of Cantharidine molluscs in New Zealand where bull kelp exists in abundance could be compared with studies of similar Cantharidine molluscs from temperate Australia where bull kelp is absent to assess whether the

association with *D. antarctica* promotes more long distance dispersal in New Zealand species as opposed those in Australia. Another approach may be to ensure that future phylogeographic studies of New Zealand direct developers include the sampling of sites within busy port regions as well as areas outside of ports. If human mediated translocations have occurred frequently it is likely that evidence of long distance translocations will primarily be found within port regions.

While phylogeographic studies of individual marine invertebrate species across limited ranges may reveal general trends of gene flow, accurately assigning directionality of gene flow requires sampling across the entirety of a species distribution. By combining evidence from genetic markers that evolve at different rates, differences in historical vs. contemporary gene flow may be detected. Phylogeographic studies need to promote unbiased comparisons in order to effectively investigate the underlying ecological processes that determine genetic connectivity. It is hoped that, in future, phylogeographic studies of New Zealand's coastal benthos will use sampling regimes designed to answer specific research questions rather than simply relying on *post hoc* speculation.

References:

- Apte, S. & Gardner, J.P.A. (2002) Population genetic subdivision in the New Zealand greenshell mussel (*Perna canaliculus*) inferred from single-strand conformation polymorphism analysis of mitochondrial DNA. *Molecular Ecology* **11**, p1617–1628.
- Avise, J.C. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.
- Ayers, K.L. & Waters, J.M. (2005) Marine biogeographic disjunction in central New Zealand. *Marine Biology*, **147**(4), p1045–1052.
- Bandelt, H., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**(1), p37–48.
- Barco, A., Marshall, B., Houart, R. & Oliverio, M. (2015) Molecular phylogenetics of Haustriinae and Pagodulinae (Neogastropoda: Muricidae) with a focus on New Zealand species *Journal of Molluscan Studies* **81**, p476–488.
- Barrows, T. T. & Juggins, S. (2005) Sea-surface temperatures around the Australian margin and Indian Ocean during the Last Glacial Maximum. *Quaternary Science Reviews* **24**, p1017–1047.
- Barton, N.H., Etheridge, A.M., Kelleher, J. & Veber, A. (2013). "Genetic hitchhiking in spatially extended populations". *Theoretical Population Biology*. **87**, p75–89
- Bell, J.J. & Okamura, B. (2005) Low Genetic Diversity in a Marine Nature Reserve: Re-Evaluating Diversity Criteria in Reserve Design. *Proceedings: Biological Sciences* **272**(1567), p1067–1074
- Benzie, J.A.H. & Williams, S.T. (1997) Genetic structure of giant clam (*Tridacna maxima*) populations in the West pacific is not consistent with dispersal by present-day ocean currents. *Evolution* **51**, p768–783.
- Birky, C.W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**, p613–627.
- Botsford, L.W., Hastings, A. & Gaines, S.D. (2001) Dependence of sustainability on the configuration of marine reserves and larval dispersal distance. *Ecology Letters* **4**, p144–150.
- Brasher, D.J., Ovenden, J.R., Booth, J.D. & White, R.W.G. (1992) Genetic subdivision of Australian and New Zealand populations of *Jasus verreauxi* (Decapoda, Palinuridae)—preliminary evidence from the mitochondrial genome. *New Zealand Journal of Marine and Freshwater Research* **26**, p53–58.
- Brown, J.H. & Kodric-Brown, A. (1977) Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* **58**, p445–449.

- Buroker, N.E. (1983). Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. *Mar. Biol.* **75**, p99–112.
- Bussolini, L.T. & Waters, J.M. (2015) Genetic analyses of rafted macroalgae reveal regional oceanographic connectivity patterns *Journal of Biogeography* **42**, p1319–1326
- Carrasco, S.A. & Phillips, N.E. (2014) Encapsulation and development of three New Zealand neogastropods with contrasting embryo packaging and maternal provisioning, *New Zealand Journal of Zoology*, **41**(3), p171-186
- Carrasco, S.A., Phillips, N.E. & Sewell M.A. (2016) Maternal lipid provisioning mirrors reproductive strategies in direct-developing whelks. *Biol. Bull.* **230**, p188-196
- 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**, p385–397.
- Colson, I. & Hughes, R.N. (2004) Rapid recovery of genetic diversity of dogwhelk (*Nucella lapillus* L.) populations after local extinction and recolonization contradicts predictions from life-history characteristics *Molecular Ecology* **13**, p2223–2233
- Corander, J. & Marttinen, P. (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, **15**, p2833-2843.
- Corander, J. & Tang, J. (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences*, **205**(1), p19-31.
- Cumming, R.A., Nikula, R., Spencer, H.G. & Waters, J.M. (2014) Transoceanic genetic similarities of kelp-associated sea slug populations: long-distance dispersal via rafting? *Journal of Biogeography* **41**, p2357–2370
- Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**(8), p772.
- Dohner, M. (2016) Crawling to connectivity? The direct-developing journey of the spotted whelk (*Cominella maculosa*). Published MSc thesis. Victoria University of Wellington, New Zealand.
- Dohner, M., Phillips, N.E. & Ritchie, P. (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**, p593–602.
- Donald, K.M. & Spencer, H.G. (2016) Phylogeographic patterns in New Zealand and temperate Australian cantharidines (Mollusca: Gastropoda: Trochidae: Cantharidinae): Trans-Tasman divergences are ancient. *Molecular Phylogenetics and Evolution* **100**, p333–344
- Donald, K.M., Keeney, D.B. & Spencer, H.G. (2011) Contrasting population makeup of two intertidal gastropod species that differ in dispersal opportunities *Journal of Experimental Marine Biology and Ecology* **396**, p224–232

- Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**(12), p2571-81.
- Excoffier, L. & Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**(3), p564–567.
- Fleming, A.M., Dohner, M.M., Phillips, N.E. & Ritchie, P.A. (2018) Genetic connectivity among populations of two congeneric direct-developing whelks varies across spatial scales *New Zealand Journal of Marine and Freshwater Research* **52**(1), p100-117
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**(5), p294–299.
- Fraser, C.I., Nikula, R. & Waters, J.M (2011) Oceanic rafting by a coastal community *Proc. R. Soc. B* **278**, p649–655
- Fraser, C.I., Nikula, R., Ruzzante, D.E. & Waters, J.M. (2012) Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends in Ecology and Evolution*, **27**(8), p462–471.
- Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**(2), p915-925.
- Gardner, J.P.A, Bell, J., Constable, H., Hannan, D., Ritchie, P.A. & Zuccarello, G. (2010) Multi-species coastal marine connectivity: a literature review with recommendations for further research. *New Zealand Aquatic Environment and Biodiversity Report*, **58**.
- Giribet, G. & Boyer, S. (2010) ‘Moa’s Ark’ or ‘Goodbye Gondwana’: is the origin of New Zealand’s terrestrial invertebrate fauna ancient, recent, or both? *Invertebrate Systematics*, **24**, p1–8.
- Gonzalez-Wevar, C.A., Nakano, T., Palma, A. & Poulin E. (2017) Biogeography in *Cellana* (Patellogastropoda, Nacellidae) with Special Emphasis on the Relationships of Southern Hemisphere Oceanic Island Species. *PLoS ONE* **12**(1): e0170103. doi:10.1371/journal.pone.0170103
- Gordillo, S. & Nielsen, S.N. (2013) The Australasian muricid gastropod *Lepsiella* as Pleistocene visitor to southernmost South America *Acta Palaeontol. Pol.* **58**(4), p777–783
- Guindon, S. & Gascuel, O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood”. *Systematic Biology* **52**, p696-704.
- Harpending, H.C., Sherry, S.T., Rogers, A.R. & Stoneking, M. (1993) Genetic structure of ancient human populations *Curr. Anthr* **34**, p483-496

- Hedgecock, D. (1994). Does variance in reproductive success limit effective population size of marine organisms? p122–134 in A. Beaumont, ed. *Genetics and evolution of aquatic organisms*. Chapman & Hall, London.
- Hellberg, M.E., Burton, R.S., Neigel, J.E. & Palumbi S.R. (2002) Genetic assessment of connectivity in marine populations. *Bulletin of Marine Science*, **70**(1), p273–290
- Henry, P.Y. & Jarne, P. (2007) Marking hard-shelled gastropods: tag loss, impact on life-history traits, and perspectives in biology. *Invertebrate Biology* **126**, p138–153.
- Hewitt, C.L., Gollasch, S. & Minchin, D. (2009) The vessel as a vector–biofouling, ballast water and sediments. p117–131. In: Rilov G., Crooks J.A. (eds) *Biological Invasions in Marine Ecosystems. Ecological Studies (Analysis and Synthesis)*, **204**. Springer, Berlin, Heidelberg
- Jackson, J.B.C. (1986) Modes of dispersal of clonal benthic invertebrates: consequences for species' distribution and genetic structure of local populations. *Bull. Mar. Sci.* **39**, p588–606.
- Johnson, M.S. & Black, R. (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria sp.* *Marine Biology* **70**, p157–164.
- Jones, M.R.L. & Ross, P.M. (2018) Recovery of the New Zealand muricid dogwhelk *Haustorium scobina* from TBT-induced imposex *Mar Pollut Bull.* **126**, p396–401
- Jones, T.C., Gemmill, C.E.C. & Pilditch, C.A. (2008) Genetic variability of New Zealand seagrass (*Zostera muelleri*) assessed at multiple spatial scales. *Aquatic Botany* **88**, p39–46
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**(12), p1647–1649.
- Keeney, D.B., King, T.M., Rowe, D.L. & Poulin, R. (2009) Contrasting mtDNA diversity and population structure in a direct-developing marine gastropod and its trematode parasites. *Molecular Ecology*, **18**(22), p4591–4603.
- 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**, p1477–1488
- Kumar, S., Stecher, G. & Tamura, K. (2015) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* (submitted).
- Leigh, J.W. & Bryant, D. (2015) PopART: Full-feature software for haplotype network construction. *Methods Ecol Evol* **6**(9), p1110–1116.

- 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**(3), p293–312.
- Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**(11), p1451–1452.
- Lidgard, C.W. (2001) Evaluating the population genetic structure of the New Zealand cockle *Austrovenus stutchburyi* using allozyme electrophoresis, MSc thesis, University of Waikato
- Lowe, W.H., & Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**(15), p3038–3051.
- Marko, P.B. (2004) ‘What’s larvae got to do with it?’ Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential *Molecular Ecology* **13**, p597–611
- Mayr, E. (1942). Systematics and the origin of species from the viewpoint of a zoologist. New York: Columbia University Press.
- Mayr, E. (1970) ‘Populations, Species and Evolution.’ (Harvard University Press: Cambridge, MA.)
- Moore, P.R. & Kenny, J.A. (1986) Exotic rocks on the northern New Zealand coast: a probable association with early shipwrecks. *Tane*, **31**, p125–131.
- Morton, J. & Miller, M. (1968) The New Zealand Sea Shore. Liverpool, C. Tinling and Co.
- Newnham, R.M., Lowe, D.J. & Williams, P.W. (1999) Quaternary environmental change in New Zealand: a review. *Progress in Physical Geography*. **23**, p567–610.
- Nikula, R., Spencer, H.G. & Waters, J.M. (2011) Evolutionary consequences of microhabitat: population-genetic structuring in kelp- vs. rock-associated chitons *Molecular Ecology* **20**, p4915–4924
- 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**, p201–207.
- Palumbi, S.R. (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Syst.* **25**, p547–572.
- Palumbi, S.R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, p146–158.
- Panova, M., Blakeslee, A.M., Miller, A.W., Makinen, T., Ruiz, G.M., Johannesson, K., Andre, C. & Steinke, D. (2011) Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. *PLoS One*. **6**(3): e17511. <https://doi.org/10.1371/journal.pone.0017511>

- Quoy, J.C.R. & Gaimard, P. (1833) Voyage de découvertes 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.
- Reeb, C.A. & Avise, J.C. (1990) A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**, p397–406.
- Rieseberg, L.H. & Burke, J.M. (2001) A genic view of species integration. *Journal of Evolutionary Biology* **14**, p883–886.
- 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**, p1009–1027.
- Ross, P.M., Hogg, I.D., Pilditch, C.A., Lundquist, C.J. & Wilkins, R.J. (2012) Population Genetic Structure of the New Zealand Estuarine Clam *Austrovenus stutchburyi* (Bivalvia: Veneridae) Reveals Population Subdivision and Partial Congruence with Biogeographic Boundaries *Estuaries and Coasts* **35**, p143-154
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning* **2**, Cold spring harbor laboratory press, New York.
- Sauvage, S., Marc, M.G., & Vincent, M.G. (1954) Transport littoral, formation de fleches et de tombolo. *Proceedings of 5th Conference on Coastal Engineering, Council on wave research*, p296-327
- Scheltema, R.S. (1986). On dispersal and planktonic larvae of benthic invertebrates: an eclectic overview and summary of problems. *Bull. Mar. Sci.* **39**, p290-322.
- Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, p787–793.
- Smith, P.J. & McVeagh, M. (1991). Widespread organotin pollution in New Zealand coastal waters as indicated by imposex in dogwhelks. *Mar. Pollut. Bull.* **22**, p409–413.
- Star, B., Apte, S. & Gardner, J.P.A. (2003) Genetic structuring among populations of the greenshell mussel *Perna canaliculus* revealed by analysis of randomly amplified polymorphic DNA. *Marine Ecology Progress Series* **249**, p171–182.
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3), p585-595.
- Tamura, K. (1992) "Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases". *Molecular Biology and Evolution*. **9**(4), p678–687
- Tan, K.S. (2003) Phylogenetic analysis and taxonomy of some southern Australian and New Zealand Muricidae (Mollusca: Neogastropoda), *Journal of Natural History*, **37**(8), p911-1028

- Thiel, M. & Haye, P.A. (2006) The ecology of rafting in the marine environment. III. Biogeographical and evolutionary consequences. *Oceanography and Marine Biology* **44**, p323–429.
- Thorrold, S.R., Jones, G., Hellberg, M.E., Burton, R.S., Swearer, S.E., Neigel, J.E., Morgan, S.G. & Warner, R.R. (2002) Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science* **70**, p291–308.
- Turgeon, K. & Kramer, D.L. (2012) Compensatory immigration depends on adjacent population size and habitat quality but not on landscape connectivity. *Journal of Animal Ecology* **81**, p1161–1170.
- Veale, A.J. & Lavery, S.D. (2011) Phylogeography of the snakeskin chiton *Sypharochiton pelliserpentis* (Mollusca: Polyplacophora) around New Zealand: are seasonal near-shore upwelling events a dynamic barrier to gene flow? *Biological Journal of the Linnean Society*, **104**, p552–563.
- Veale, A.J. & Lavery, S.D. (2012) The population genetic structure of the waratah anemone (*Actinia tenebrosa*) around New Zealand, *New Zealand Journal of Marine and Freshwater Research*, **46**(4), p523–536
- Walton, K. (2017) Phylogeography of the New Zealand whelks *Cominella maculosa* and *C. virgata*. MSc thesis. Victoria University of Wellington, New Zealand.
- Waples, R.S. & Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**(6), p1419–1439.
- Waples, R.S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, p438–450.
- Wares, J.P., Gaines, S.D. & Cunningham, C.W. (2001) A comparative study of asymmetric migration events across a marine biogeographic boundary. *Evolution*, **55**(2), p295–306
- Waters, J.M. & Roy, M.S. (2004) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*, **13**(9), p2797–2806.
- Waters, J.M., Fraser, C.I. & Hewitt, G.M. (2013) Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology and Evolution*, **28**(2), p78–85.
- Weersing, K. & Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**, p1–12.
- West, J.B., Bowen, G.J., Cerling, T.E. & Ehleringer, J.R. (2006) Stable isotopes as one of nature's ecological recorders. *Trends in Ecology & Evolution* **21**, p408–414

Wilke, T. & Davis, G.M. (2000) Intraspecific mitochondrial sequence diversity in *Hydrobia ulvae* and *Hydrobia ventrosa* (Hydrobiidae: Rissooidea: Gastropoda): Do their different life histories affect biogeographic patterns and gene flow? *Biological Journal of the Linnean Society* **70**, p89-105