The Genetic Analysis of *Lasaea hinemoa*: The Story of an Evolutionary Oddity

KATHERINE LOCKTON

A thesis submitted for the degree of
Master of Science
at the University of Otago, Dunedin, New Zealand

1 March 2019

ABSTRACT

Lasaea is a genus of molluscs that primarily consists of minute, hermaphroditic bivalves that occupy rocky shores worldwide. The majority of *Lasaea* species are asexual, polyploid, direct developers. However, two Australian species are exceptions: *Lasaea australis* is sexual, diploid and has planktotrophic development, whereas *Lasaea colmani* is sexual, diploid and direct developing.

The New Zealand species *Lasaea hinemoa* has not been phylogeographically studied. I investigated the phylogeography of *L. hinemoa* using mitochondrial and nuclear gene sequencing (COIII and ITS2, respectively). Additionally, I investigated population-level structuring around Dunedin using microsatellite markers that I developed.

It was elucidated that the individuals that underwent genetic investigation consisted of four clades (Clade I, Clade II, Clade III and Clade IV). Clade I and Clade III dominated in New Zealand and support was garnered through gene sequencing and microsatellite analysis for these clades to represent separate cryptic species, with biogeographic splitting present. Clade II consisted of individuals that had been collected from the Antipodes Island. The Antipodes Island contained individuals from two clades (Clade I and Clade II), with *Lasaea* from the Kerguelen Islands being more closely related to individuals from Clade II than Clade I was to Clade II. This genetic distinction between Clade I and Clade II seemed to indicate transoceanic dispersal via the Antarctic Circumpolar Current (ACC) between the Kerguelen Islands and Antipodes Island. Clade IV clustered very distinctly from *L. hinemoa*, appearing to represent transoceanic dispersal by another *Lasaea* species.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of some truly invaluable people throughout the course of my master's degree. First and foremost, I'd like to acknowledge the efforts of my supervisor Hamish Spencer; his support and guidance has been integral to this project, and I've appreciated his response time to emails and patience listening to my concerns.

I've appreciated the support and advice of Tania King, Martyn Kennedy and Kirsten Donald in the lab, the bioinformatic wisdom of Ludo Dutoit, and the microsatellite guidance of Graham McCulloch.

This project would have been impossible without the help of my many fieldwork assistants throughout New Zealand; particularly Luke Easton, Clare Cross, William Pearman, Tiffany Lo, Rachael Hudson and Yasmin Foster.

My friends for their endlessly appreciated support, from listening to me in times of stress to catching up for social activities. Ellie Hay, for your advice with phylogenetic analysis and friendship. Amber Helliwell, for all help throughout the past two years, at the gym, providing a long-suffering soundboard or just socialising. Philip Greenspoon, for surviving me as an 'office intruder' and reading drafts of my work.

Lastly, I'd like to thank my family for their endurance and patience these past two years. Harry and Jonathan for helping me sample throughout Taranaki, despite well expressed concerns regarding their limited distribution. Antonia, for being a bedrock of support (even if it was non-verbally, and frequently unexpressed). Cordelia, for her ability to make me laugh in stressful situations. My Mum, for her endless love and support in all aspects of my life. My Grandfather (Grandad) for nutritionally supporting throughout all of January 2019, and for ensuring that I had as many cushions to study with as I could ever want.

Everyone I've mentioned, and many more people around the department every day have enabled me to work through the past two years complete this thesis, and you have my endless gratitude.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. General Introduction

1.1 The genus Lasaea

1.1a Lasaea

Lasaea is a genus of molluscs that consists of minute, hermaphroditic bivalves that occupy rocky shores worldwide, is found on all continents (only absent from Antarctica), and several oceanic Islands (Beauchamp 1986; Ó Foighil, 1989) (Figure 1.1).

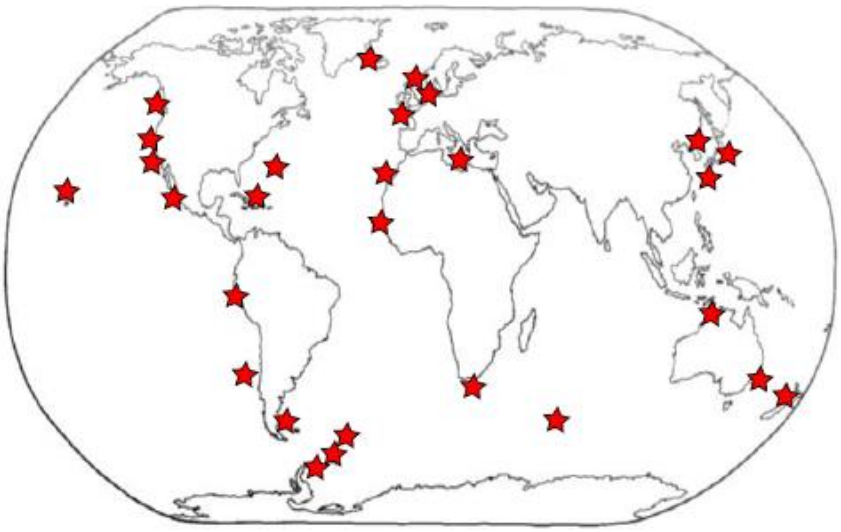


Figure 1.1: Example of the broad distribution observed by *Lasaea* species (locations for species taken primarily from Keen (1938)). Figure modified from Beauchamp (1986).

Brown (1827) first identified the genus and many more species in this genus have since been characterised and studied (Beauchamp 1986; Ó Foighil & Thiriot-Quiévreux, 1999). *Lasaea* consist almost exclusively of direct-developing (i.e. ovoviparous), polyploid species (Ó Foighil & Thiriot-Quiévreux, 1999). Members of the *Lasaea* genus are commonly known as brooding clams as adults brood embryos in the suprabranchial chamber until the juvenile stage of development (Figure 1.2) (Beauchamp, 1986; Thiriot-Quiévreux, Soyer, De Bovée, & Albert, 1988).

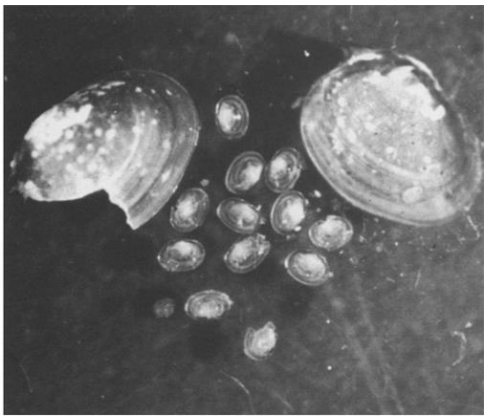


Figure 1.2: Adult and brooded offspring of *Lasaea subviridis*. Shell length of adult = 2.70mm; shell length of juveniles = 0.60mm. Figure taken from Beauchamp (1986).

The genus is particularly interesting due to a combination of unusual genetic features; including predominant asexual reproduction, direct development and polyploidy. Within *Lasaea*, there is a dichotomy between two sexual, diploid Australian species and other known *Lasaea* species, which are asexual polyploids (Ó Foighil & Thiriot-Quiévreux, 1999). One of the Australian species, *L. australis*, exhibits planktotrophic development, whilst the rest of the genus and the other sexual Australian species, *L. colmani*, exhibit direct development.

1.1b Asexual Reproduction in the genus Lasaea

The mode of reproduction in the *Lasaea* genus has been a source of contention for decades, with numerous studies having been dedicated to discovering how particular lineages reproduce (Beauchamp 1986; Ó Foighil & Eernisse, 1988; Ó Foighil & Thiriot-Quiévreux, 1991). The genus was found to be hermaphroditic through studies on *L. rubra* (Oldfield, 1961), *L. subviridis* (Ó Foighil, 1985) and *L. australis* (Roberts, 1984). Crisp and Standen (1988) suggested an apomictic reproductive system for *L. rubra* from breeding experiments, electrophoretic studies of wild populations and the lack of male gonads in histological preparations. An apomictic reproductive system is where the eggs develop without any overt meiotic process or segregation of alleles. A lack of sexual reproduction was supported by Ó Foighil (1985) who found that *L. subviridis* had a relatively small number of sperm coupled with incomplete nuclear condensation and poor sperm motility. The progeny from pair mating experiments (Ó

Foighil & Eernisse, 1988) and field brooding specimens were found to preserve maternal protein phenotypes, further supporting a lack of sexual reproduction (Ó Foighil & Eernisse, 1988; Tyler-Walters & Davenport, 1990).

Polyploid *Lasaea* have been found to consist of simultaneous hermaphrodites with minute male allocation (Ó Foighil & Eernisse, 1988). Ó Foighil and Thiriot-Quiévreux (1991) confirmed Crisp and Standen's (1988) hypothesis that *Lasaea* with non-pelagic larvae do not reproduce via self-fertilization, but instead develop parthenogenetically when triggered gynogenetically (i.e. without fertilization) by autosperm. This is sperm that is produced by the reproductive individual to trigger development but does not contribute genetically to the offspring. The incorporated sperm nucleus disintegrates in the egg cortex and does not fuse with the egg (syngamy does not materialise). Both polar bodies have a diploid chromosome number, which suggests they are the result of mitotic divisions. Pseudogamy (gynogenesis) occurs in a wide variety of taxa and has been induced in many mollusc species (e.g. *Crassostrea gigas* (Guo, Hershberger, Cooper, & Chew, 1993), *Mytilus edulis* (Fairbrother, 1994), and *Haliotis discus hannai* (Li, Osada, Kashihara, Hirohashi, & Kijima, 2000)) and pseudogamous individuals are usually the sexual parasites of a closely related cross-fertilizing species (Kiester, Nagylaki, & Shaffer, 1981; Hughes, 1989). *Lasaea* clones are exceptional in that individuals are reproductively independent; using their own sperm to trigger asexual development (Ó Foighil & Thiriot-Quiévreux, 1991).

Intriguingly, this asexual mode of reproduction has persisted alongside sexual reproduction in two species in the genus (Ó Foighil & Smith, 1995). *L. australis* and *L. colmani* (Ó Foighil & Thiriot-Quiévreux, 1999), unlike other species of *Lasaea* have been revealed to be sexual. *L. australis* also has a smaller egg size, higher fecundity and releases straight-hinged planktotrophic veligers rather than juveniles that crawl away (Ó Foighil, 1988).

Lynch and Gabriel (1990) have shown through modelling that asexual lineages don't have long term viability. Sexual reproduction allows populations to generate genetic diversity through recombination and thus be able to more effectively manage change in the environment (Crow, 1992). The majority of studies have found that extant asexual lineages have arisen recently, suggesting that asexual lineages generally have high

extinction rates, and are therefore at an evolutionary disadvantage (Nunney, 1989; Schwander & Crespi, 2009). The high level of mitochondrial genetic divergence observed between sexual and asexual *Lasaea* is evidence of the significant length of evolutionary time between the two lineages, which in turn implies that the asexual *Lasaea* lineages have persisted for a significant period (Taylor & Ó Foighil, 2000).

Asexual lineages of *Lasaea* exhibit a wide distribution globally, whilst sexual *Lasaea* are apparently restricted to Australia (Ó Foighil, 1989). The global distribution of asexual *Lasaea* lineages also indicates that they may be of relatively ancient origin compared to previously studied asexual taxa. The restricted distribution of the sexual *Lasaea* species remains an interesting phenomenon in evolutionary biology, as ecological theories of sexual reproduction widely argue that sexuality is advantageous, particularly in heterogenous environments (Toman & Flegr, 2018). For example, the Red Queen hypothesis (Moritz, McCallum, Donnellan, & Roberts, 1991), the evolutionary arm-races hypothesis (Dawkins & Krebs, 1979) and the fast-sexual-response hypothesis (Becerra, Brichette, & Garcia, 1999) all stress the importance of sexual reproduction providing the means to co-evolutionarily respond to biotic pressures. Additional ecological theories, focussing on abiotic pressures, further favour the maintenance of sexual reproduction in heterogeneous environments (Toman & Flegr, 2018).

Toman and Flegr (2018) found that when they examined the maintenance of ancient asexual clades (>1 million years), the majority of sexual species that were closely related to asexual species were found in relatively more heterogeneous environments than the asexual species. However, *Lasaea* was found to be an exception to this trend, as the heterogeneity of the environment experienced was not noticeably different between the sexual and asexual species. Asexual *Lasaea* remain an intriguing and unusual example of ancient asexuals that globally dominate over their sexual sister taxa. One possible explanation of this pattern is that there is a lack of specialist predators or parasites that would drive the maintenance of asexual reproduction (in accordance with the Red Queen Hypothesis). Apart from occasional opportunistic predation by gastropods, there do not appear to be many biotic threats, let alone a strongly specialised predator (Ponder & Taylor, 1992; Toman & Flegr, 2018). Without

specialised enemies to drive the maintenance of sexual reproduction, asexual lineages may continue to dominate.

1.1c Direct Development in the genus Lasaea

Marine benthic invertebrates primarily display three developmental modes: planktotrophic, lecithotrophic and direct development (Thorson, 1950; Mileikovsky, 1971). A planktotrophic developmental mode consists of the production of pelagic larvae that feed in the plankton until the onset of metamorphosis. A lecithotrophic mode is characterised by pelagic but non-feeding larvae that largely depend on yolk material for nutrition. Direct development involves the release of non-pelagic 'crawl-away' juveniles that do not spend time in the plankton.

Lasaea almost exclusively consists of direct developing species (Ó Foighil & Thiriot-Quiévreux, 1999), which are ovoviviparous and brood embryos up to a juvenile stage of development in the suprabranchial chamber (Beauchamp, 1986). However, in contrast to all other known species *L. australis* exhibits planktotrophic development (Ó Foighil & Thiriot-Quiévreux, 1999). It is thought that the planktotrophic mode is the ancestral state, because complex modes of development are more likely to be lost in a species than gained (Strathmann, 1978; Li & Ó Foighil, 2016). In effect, this interpretation means that evolutionary transitions from planktotrophic development to lecithotrophic or direct development are more common, than transitions to planktotrophic development from lecithotrophic or direct development.

Direct development, however, is linked to the asexual reproductive mode, and has evolved several times in the genus independently (Ó Foighil, 1989; Li & Ó Foighil, 2016). However, an exception is *L. colmani* which is direct developing and sexual (Ó Foighil & Thiriot-Quiévreux, 1999).

The developmental mode has implications for species dispersal, which in turn can directly influence geographic range, population structure, speciation and extinction events (Jablonski, 1986). Planktotrophic species generally have the highest rate of dispersal whilst direct developers have the lowest. Oddly for this genus, however, it has

been found that those with planktotrophic development have the smallest geographic range, whilst the majority of the species of this genus lacking pelagic larvae have the largest range, including all continents (excluding Antarctica), and a number of oceanic Islands (Ó Foighil, 1989). It is thought that direct developing *Lasaea* have achieved such a broad range through rafting i.e., the passive transport of sedentary or sessile life history stages on drifting objects (Jackson, 1986). Species with non-pelagic larvae also have better survival as adults protect the larvae (Booth, 1979).

The practicality of rafting for direct developing *Lasaea* was examined by Ó Foighil and Jozefowicz (1999) who tested colonization hypotheses for North Atlantic oceanic islands. Thirty individuals each were sampled from two continental putative source populations (Florida, Iberia) and two oceanic island populations (Bermuda, Azores). They underwent sequencing for the mitochondrial large ribosomal subunit (16S) gene. No Amphi-Atlantic genotypes were detected; Bermudan lineages co-clustered exclusively with Floridian congeners, and Azorean samples formed an exclusive clade with Iberian haplotypes. Geographic proximity to continental source populations was found to be a better predictor of phylogenetic relationships in North Atlantic *Lasaea* than modern-day oceanic surface circulation patterns. The phylogenetic trees they produced were not consistent with trans-oceanic rafting events (Figure 1.3), but they were found to be consistent with limited (≤2000km) colonization through rafting (against present-day circulation patterns in the case of the Azores), and through anthropogenic introduction.

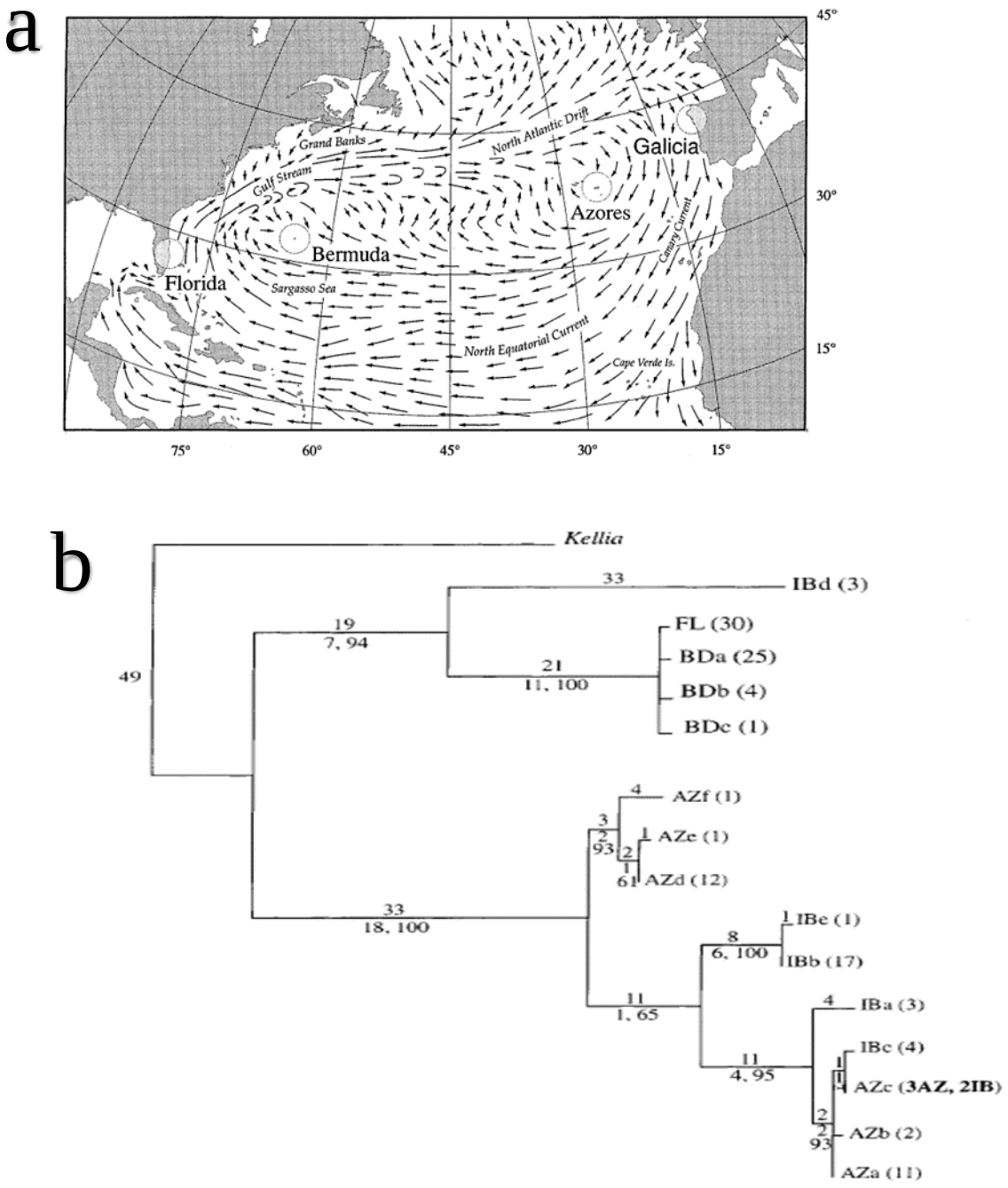


Figure 1.3: Figure taken from Ó Foighil & Jozefowicz, 1999.

a) Sampling locations used for continental (Florida; Galicia, Spain) and oceanic islands (Bermuda; Sao Miguel, Azores) *Lasaea* populations. The predominant surface currents of the North Atlantic Gyre are indicated. b) *Lasaea* spp. Strict consensus of the five most parsimonious trees obtained by a heuristic search for optimal trees of the mitochondrial 16S rDNA dataset, utilizing *Kellia laperousi* as an outgroup. The respective numbers of steps are indicated above each branch, and the decay index and bootstrap values (500 branch and bound iterations) supporting each node are respectively presented below the branches. Prefixes indicate the location from which a mitochondrial genotype was first obtained (AZ Azores; IB Galicia, Spain; BD Bermuda; FL Florida) and numbers in parentheses after haplotype labels indicate the number of individuals detected per mitochondrial lineage.

1.1d Polyploidy in the genus Lasaea

Polyploidy is rare in bivalve molluscs and has been recorded in only one other species, the freshwater clam *Corbicula leana* (Okamoto & Arimoto, 1986). Few taxa have comparable levels of pronounced genome duplication found in asexual, polyploid *Lasaea* species (Ó Foighil & Thiriot-Quiévreux, 1991). These direct-developing populations of *Lasaea* are uncommon amongst marine bivalves in that they comprise many polyploid, clonal lineages, with high levels of fixed heterozygosity. Analysis has failed to find evidence of meiotic metaphases in polyploid *Lasaea*, due to ploidy levels and supernumary chromosomes (Thiriot-Quiévreux et al., 1988; Ó Foighil & Thiriot-Quiévreux, 1991).

Ploidy levels in *Lasaea* can range from 3N to 6N, even amongst sympatric clones and typically include supernumary chromosomes (Ó Foighil & Thiriot-Quiévreux, 1991; 1999). The latter are chromosomes that are additional to the standard chromosomes of a normal diploid complement, rather than being members of homologous pairs (Evans, 1960). As polyploid *Lasaea* are parthenogenetic and rely on sperm activation, leakage of chromosomes from sperm may have contributed to the origin of supernumary chromosomes (Ó Foighil & Thiriot-Quiévreux, 1991).

A substantive comparison of karyotypes between different *Lasaea* populations is difficult because homologous chromosome sets cannot be easily distinguished from supernumary chromosomes. Some conclusions can be reached through comparing the first 17 chromosome pairs of three different populations of *Lasaea*, from Kerguelen Island (Thiriot-Quiévreux et al., 1988), Europe (Thiriot-Quiévreux, Insua Pombo, & Albert, 1989) and the North-eastern Pacific (Ó Foighil & Thiriot-Quiévreux, 1991). These lineages differed in their relative numbers of metacentric, submetacentric, subtelocentric, and telocentric chromosome sets. These distinct karyotypes indicated that these three populations are not only reproductively incompatible but also that they have experienced different evolutionary mechanisms producing polyploidy. North-eastern Pacific *Lasaea* clones are triploid and are thought to have arisen from rare hybridization events between ancestral lineages (Ó Foighil & Thiriot-Quiévreux, 1991).

There are two known species of diploid *Lasaea*: *L. australis* and *L. colmani* which are restricted to Australia, and exhibit morphological dissimilarity from their polyploid congeners (primarily hinge characters, and prodissoconch structure) (Ó Foighil & Smith, 1995; Ó Foighil & Thiriot-Quiévreux, 1999). Interestingly while both *L. australis* and *L. colmani* are diploid and sexual, *L. australis* has pelagic larvae whereas *L. colmani* instead undergoes direct development. Australia also has an asexual, polyploid species of *Lasaea* consisting of many clonal lineages. Similarities found in shell type morphology and mitochondrial gene sequence between *L. australis* and the polyploid clones suggest a common ancestor. However, karyological studies failed to find obvious karyological similarities between the two. It is thought that the polyploid clones were generated by hybridisation followed by a radical karyological rearrangement (Ó Foighil & Thiriot-Quiévreux, 1999).

1.2 Lasaea *species relationships*

Ó Foighil and Smith (1995; 1996) reconstructed the phylogenetic relationships of reproductively characterized *Lasaea* populations using Cytochrome Oxidase III (COIII) mitochondrial gene sequences. Ó Foighil and Smith (1995) found that asexual *Lasaea* lineages are polyphyletic with mitochondrial contributions from at least two unknown parental species and could be the result of multiple hybridization events. A gynogenetic reproductive mode and the presence of polyploid, highly heterozygous nuclear genomes is also consistent with a hybrid origin for *Lasaea* asexual clones (Ó Foighil & Thiriot-Quiévreux, 1991).

The high degree of genetic divergence of asexual lineages from co-clustering sexual congeners (16%-22%) and between geographically restricted monophyletic clones (9%-11%) suggests that asexual *Lasaea* lineages may be evolutionarily long-lived (Ó Foighil & Smith, 1995). This genetic divergence suggests an estimated divergence time of six to seven million years. Because of the large genetic divergences among sexual and asexual lineages, neither sexual lineage could be identified as a parental species to the asexual lineages. Taylor and Ó Foighil (2000) also concluded that Australian sexual species could not be identified as convincing parental species to any of the clonal

lineages due to pronounced mitochondrial genetic divergence levels and developmental differences. Australian sexual species formed sister taxa to a minority of the clonal lineages. Monophyly is supported for the diploid Australian direct-developing lineage together with the remaining polyploid asexual lineages from the North-eastern Pacific, North-eastern Atlantic, Mediterranean, and Southern Indian Ocean. It appears that the majority of asexual lineages have originated in areas where no sexual congeners are presently known.

1.3 Rationale for study

Lasaea is an intriguing genus of bivalves with marked developmental and reproductive differences between *L. australis* and all other polyploid, asexual, direct-developing lineages (Ó Foighil & Thiriot-Quitvreux, 1991; 1999). Genetic analysis has provided insight into the presence of these defining characteristics and into the different evolutionary mechanisms that have operated in various lineages (Booth, 1979; Ó Foighil & Thiriot-Quiévreux, 1991; Taylor & Ó Foighil, 2000).

There are thought to be two New Zealand *Lasaea* species; *L. hinemoa* and *L. maoria* (Booth, 1979). The two species appear to occupy different habitats on beach (Booth, 1979). Batham (1956) found that *L. hinemoa* tends to live on the red alga *Bostrychia arbuscula* at the intertidal zone. Powell (1933) found that *L. maoria* (previously identified as *Kellia maoria*) lives at the intertidal zone on the underside of stones. The strongest trend appears to be that *L. hinemoa* is found throughout New Zealand but is more abundant in the South Island, whilst *L. maoria* is has been found in the North Island but not south of Cook Strait (Ponder, 1971). *L. hinemoa* is being utilized for the course of study due to being more widespread.

Past studies involving the phylogeny of New Zealand *Lasaea* have focused on *L. hinemoa*, sampled from intertidal sites in Wellington (Taylor & Ó Foighil, 2000). *L. hinemoa* appear to consist of direct developing polyploids (Booth, 1979, Taylor & Ó Foighil, 2000). The aforementioned traits and genetic sequencing with the COIII gene suggests that *L. hinemoa* is not closely related to the Australian species of *Lasaea*

(Taylor & Ó Foighil, 2000). It is known that *L. hinemoa* exhibits a moderate amount of genotypic diversity which could contribute to an interesting population structure (Taylor & Ó Foighil, 2000).

Through phylogenetic and population level analysis questions regarding the phylogenetic place of *L. hinemoa* across New Zealand will be answered, allowing us to know how they relate to each other and to other *Lasaea* globally. As *L. hinemoa* appears to exhibit a broad range across New Zealand (Ponder, 1971), sampling sites should be reflective of this range. This broad sampling range will capture more genetic variation and will allow it to be observed if they exhibit cryptic genetic structuring with fidelity to biogeographic province (as was the case with *L. australis*) (Li, Ó Foighil & Park, 2013).

1.4 Objectives

- Through sequencing mitochondrial Cytochrome Oxidase III (COIII) and the nuclear Internal Transcribed Spacer 2 (ITS2), I plan to resolve the phylogeny of New Zealand *L. hinemoa* across New Zealand. By doing so, I will be able to infer the phylogenetic relationship of *L. hinemoa* to *Lasaea* species globally and within the currently defined species.
- I will use microsatellite markers to conduct a population-genetic analysis of Dunedin populations of *L. hinemoa*. I will study their demographic histories, patterns of gene flow and effective population sizes.

1.5 References

Batham, E. (1956). Ecology of southern New Zealand sheltered rocky shore. *Transactions of the Royal Society of New Zealand*, *84*(2), 447-465.

Becerra, M., Brichette, I., & Garcia, C. (1999). Short-term evolution of competition between genetically homogeneous and heterogeneous populations of *Drosophila melanogaster*. *Evolutionary Ecology Research*, *1*(5), 567-579.

Beauchamp, K. A. (1986). Reproductive ecology of the brooding, hermaphroditic clam *Lasaea subviridis*. *Marine Biology*, *93*(2), 225-235.

Booth, J. D. (1979). Common bivalve larvae from New Zealand: Leptonacea. *New Zealand Journal of Marine and Freshwater Research*, *13*(2), 241-254.

Brown, T. (1827). *Illustrations of the Conchology of Great Britain and Ireland: Drawn from Nature*. Edinburgh, Scotland, U.K. and London, England, U.K.: D and W.H Lizars, and S. Highley.

Crisp, D. J., & Standen, A. (1988). *Lasaea rubra* (Montagu) (Bivalvia: Erycinacea), an apomictic crevice-living bivalve with clones separated by tidal level preference. *Journal of Experimental Marine Biology and Ecology*, *117*(1), 27-45.

Crow, J. F. (1992). An advantage of sexual reproduction in a rapidly changing environment. *Journal of Heredity*, *83*(3), 169-173.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 489-511.

Evans, H. J. (1960). Supernumerary chromosomes in wild populations of the snail *Helix pomatia* L. *Heredity*, *15*(1), 129-138.

Fairbrother, J. E. (1994). Viable gynogenetic diploid *Mytilus edulis* (L.) larvae produced by ultraviolet light irradiation and cytochalasin B shock. *Aquaculture*, *126*(1-2), 25-34.

Guo, X., Hershberger, W. K., Cooper, K., & Chew, K. K. (1993). Artificial gynogenesis with ultraviolet light-irradiated sperm in the Pacific oyster, *Crassostrea gigas*. I. Induction and survival. *Aquaculture*, *113*(3), 201-214.

Hughes, R. N. (1989). *A Functional Biology of Clonal Animals*. London, U.K. and New York, NY: Chapman and Hall.

Jablonski, D. (1986). Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science, 39*(2), 565-587.

Jackson, J. B. C. (1986). Modes of dispersal of clonal benthic invertebrates: consequences for species' distributions and genetic structure of local populations. *Bulletin of Marine Science, 39*(2), 588-606.

Keen, A. M. (1938). New pelecypod species of the genera *Lasaea* and *Crassinella*. *Journal of Molluscan Studies*, *23*(1), 18-32.

Kiester, A. R., Nagylaki, T., & Shaffer, B. (1981). Population dynamics of species with gynogenetic sibling species. *Theoretical Population Biology*, *19*(3), 358-369.

Li, J., Ó Foighil, D., & Park, J. K. (2013). Triton's trident: cryptic Neogene divergences in a marine clam (*Lasaea australis*) correspond to Australia's three temperate biogeographic provinces. *Molecular Ecology*, *22*(7), 1933-1946.

Li, J., & Ó Foighil, D. (2016). Multiple losses of planktotrophic development in the cosmopolitan bivalve genus *Lasaea*. *American Malacological Bulletin*, *33*(2), 302-307.

Li, Q., Osada, M., Kashihara, M., Hirohashi, K., & Kijima, A. (2000). Cytological studies on artificially induced gynogenesis in the Pacific abalone. *Fisheries Science*, *66*(4), 701-707.

Lynch, M., & Gabriel, W. (1990). Mutation load and the survival of small populations. *Evolution*, *44*(7), 1725-1737.

Mileikovsky, S. A. (1971). Types of larval development in marine bottom invertebrates, their distribution and ecological significance: a re-evaluation. *Marine Biology*, *10*(3), 193-213.

Moritz, C., McCallum, H., Donnellan, S., & Roberts, J. D. (1991). Parasite loads in parthenogenetic and sexual lizards (*Heteronotia binoei*): support for the Red Queen hypothesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *244*(1310), 145-149.

Nunney, L. (1989). The maintenance of sex by group selection. *Evolution*, *43*(2), 245-257.

Ó Foighil, D. (1985). Fine structure of *Lasaea subviridis* and *Mysella tumida* sperm (Bivalvia, Galeommatacea). *Zoomorphology*, *105*(2), 125-132.

Ó Foighil, D. (1988). Random mating and planktotrophic larval development in the brooding hermaphroditic clam *Lasaea australis* (Lamarck, 1818). *The Veliger*, *31*(3-4), 214-221.

Ó Foighil, D., & Eernisse, D. J. (1988). Geographically widespread, non-hybridizing, sympatric strains of the hermaphroditic, brooding clam *Lasaea* in the Northeastern Pacific Ocean. *The Biological Bulletin*, *175*(2), 218-229.

Ó Foighil, D. (1989). Planktotrophic larval development is associated with a restricted geographic range in *Lasaea*, a genus of brooding, hermaphroditic bivalves. *Marine Biology*, *103*(3), 349-358.

Ó Foighil, D., & Thiriot-Quiévreux, C. (1991). Ploidy and pronuclear interaction in Northeastern Pacific *Lasaea* clones (Mollusca: Bivalvia). *The Biological Bulletin*, *181*(2), 222-231.

Ó Foighil, D., & Smith, M. J. (1995). Evolution of asexuality in the cosmopolitan marine clam *Lasaea*. *Evolution*, *49*(1), 140-150.

Ó Foighil, D., & Smith, M. J. (1996). Phylogeography of an asexual marine clam complex, *Lasaea*, in the northeastern Pacific based on cytochrome oxidase III sequence variation. *Molecular Phylogenetics and Evolution*, *6*(1), 134-142.

Ó Foighil, D., & Jozefowicz, C. J. (1999). Amphi-Atlantic phylogeography of direct-developing lineages of *Lasaea*, a genus of brooding bivalves. *Marine Biology*, *135*(1), 115-122.

Ó Foighil, D., & Thiriot-Quiévreux, C. (1999). Sympatric Australian *Lasaea* species (Mollusca: Bivalvia) differ in their ploidy levels, reproductive modes and developmental modes. *Zoological Journal of the Linnean Society*, *127*(4), 477-494.

Okamoto, A., & Arimoto, B. (1986). Chromosomes of *Corbicula japonica*, *C. sandai* and *C.* (*Corbiculina) leana*. *Japanese Journal of Malacology*, *45*(3), 194-202.

Oldfield, E. (1961). The functional morphology of *Kellia suborbicularis* (Montagu), *Montacuta ferruginosa* (Montagu) and *M. substriata* (Montagu), (Mollusca, Lamellibranchiata). *Journal of Molluscan Studies*, *34*(5), 255-295.

Ponder, W. F. (1971). Some New Zealand and subantarctic bivalves of the Cyamiacea and Leptonacea with descriptions of new taxa. *Records of the Dominium Museum*, *7*(13), 119-141.

Ponder, W. F., & Taylor, J. D. (1992). Predatory shell drilling by two species of *Austroginella* (Gastropoda: Marginellidae). *Journal of Zoology*, *228*(2), 317-328.

Powell, A. W. B. (1933). The high-tidal Mollusca of Rangitoto Island, Auckland; with descriptions of a new genus and two new species. *Transactions of the New Zealand Institute, 63*(1), 144-153.

Roberts, D. (1984). A comparative study of *Lasaea australis*, *Vulsella spongiarum*, *Pinna bicolor* and *Donacilla cuneata* (Mollusca; Bivalvia) from Princess Royal Harbour, Western Australia. *Journal of Molluscan Studies*, *50*(3), 129-136.

Schwander, T., & Crespi, B. J. (2009). Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Molecular Ecology*, *18*(1), 28-42.

Strathmann, R. R. (1978). The evolution and loss of feeding larval stages of marine invertebrates. *Evolution*, *32*(4) 894-906.

Taylor, D. J., & Ó Foighil, D. (2000). Transglobal comparisons of nuclear and mitochondrial genetic structure in a marine polyploid clam (*Lasaea*, Lasaeidae). *Heredity*, *84*(3), 321-330.

Thiriot-Quiévreux, C., Soyer, F., De Bovee, F., & Albert, P. (1988). Unusual chromosome complement in the brooding bivalve *Lasaea consanguinea, Genetica, 76*(2), 143-151.

Thiriot-Quiévreux, C., Insua Pombo, A. M., & Albert, P. (1989). Polyploidy in an incubating bivalve, *Lasaea rubra* (Montagu). *Proceedings of the Academy of Sciences. Series 3, Life Sciences*, *308*(5), 115-120.

Thorson, G. (1950). Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews*, *25*(1), 1-45.

Toman, J., & Flegr, J. (2018). General environmental heterogeneity as the explanation of sexuality? Comparative study shows that ancient asexual taxa are associated with both biotically and abiotically homogeneous environments. *Ecology and Evolution*, *8*(2), 973-991.

Tyler-Walters, H., & Davenport, J. (1990). The relationship between the distribution of genetically distinct inbred lines and upper lethal temperature in *Lasaea rubra*. *Journal of the Marine Biological Association of the United Kingdom*, *70*(3), 557-570.

2. Nuclear and Mitochondrial Gene Sequencing of *Lasaea hinemoa*

2.1 Abstract

Previous gene-sequencing studies of New Zealand *Lasaea hinemoa* have been restricted to a small population in Wellington. In this chapter populations across New Zealand and the Antipodes Island were sampled and sequenced for the mitochondrial gene, Cytochrome Oxidase III (COIII) (n = 77), and the nuclear locus, Internal Transcribed Spacer 2 (ITS2) (n=69). The analysis of the gene sequences revealed that two distinct clades within what is currently recognized as *L. hinemoa* are present throughout much of the two main islands of New Zealand. Although there are no obvious morphological differences between members of these two clades, there is evidence of some biogeographic structure, with North Island and northern South Island populations dominated by one clade, which is less common further south.

In addition, two individuals sequenced for COIII and one individual for ITS2 fell well outside of the *L. hinemoa* ingroup. These individuals proved to be novel in New Zealand. The unusual phylogenetic placement for these individuals gives support for long distance dispersal or anthropogenic introduction as an explanation for their presence in New Zealand.

2.2 Introduction

Elucidating the phylogeny of *Lasaea* species has been a focus of many studies (Ó Foighil & Smith, 1996; Taylor & Ó Foighil, 2000; Li, Ó Foighil, & Park, 2013). The mitochondrial gene, Cytochrome Oxidase III (COIII) has been the marker of choice for many studies of *Lasaea*, and thus remains the best way to put *Lasaea* into a global context (Ó Foighil, & Jozefowicz, 1999; Taylor & Ó Foighil, 2000) (Figure 2.1).

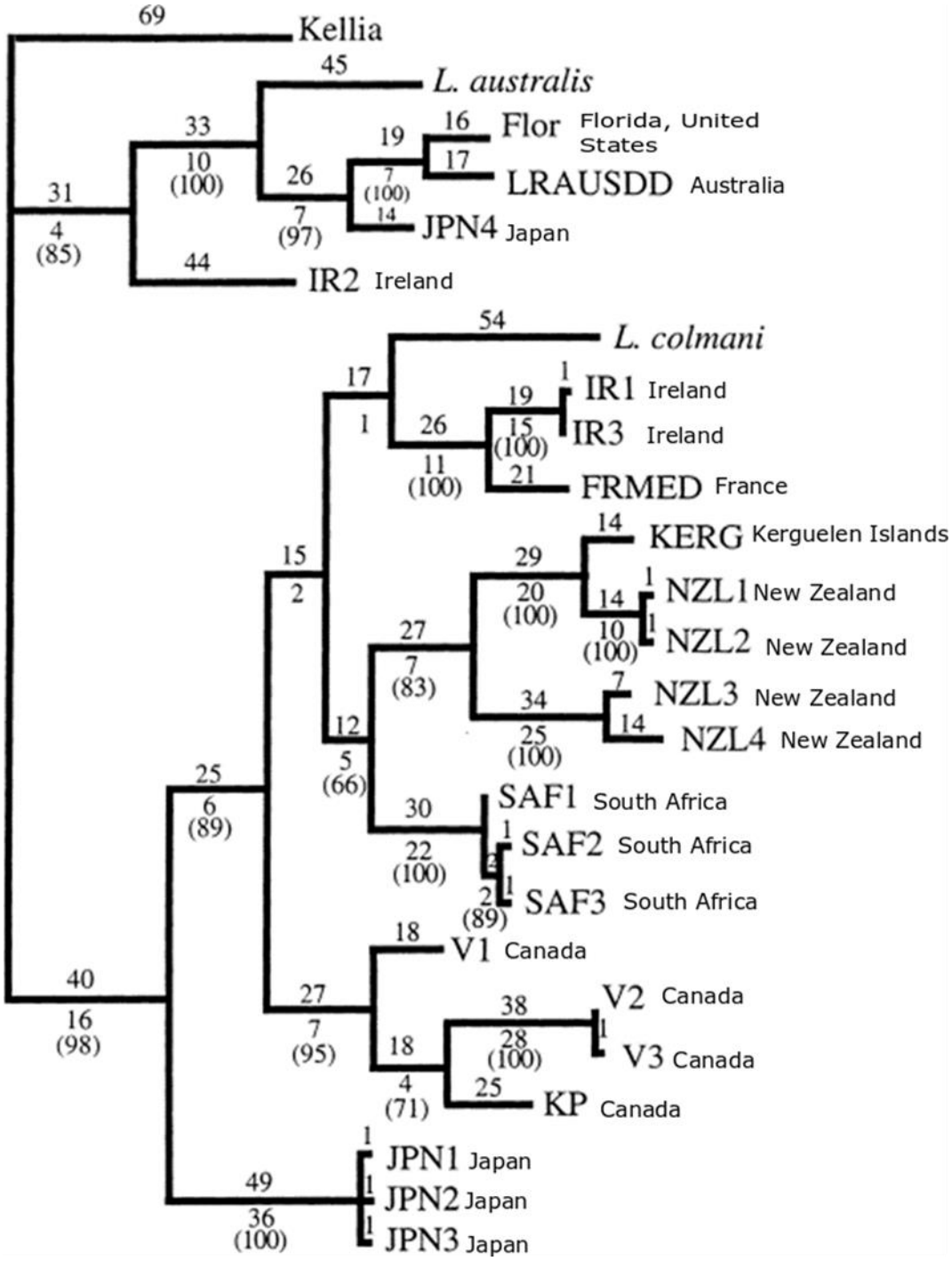


Figure 2.1: Figure adapted from Taylor and Ó Foighil, (2000). The strict consensus of the two most parsimonious trees (898 steps) obtained by a heuristic search for optimal trees using the 12 *Lasaea* COIII genotypes (Taylor, & Ó Foighil, 2000), along with homologous data from other studied populations (Ó Foighil & Smith, 1995). *Kellia laperousi* is used as an outgroup and the two sexual diploid Australian species, *L. australis* and *L. colmani*, are in italics. The respective number of steps is indicated above each branch, and decay index values supporting each node are presented below the branches. Bootstrap values are shown in parentheses.

The use of nuclear markers to analyse the *Lasaea* genus is less common. There has been progress made using Internal Transcribed Spacer 2 (ITS2) alongside mitochondrial markers to demonstrate the presence of cryptic species present within the nominal taxon *L. australis* (Li et al., 2013). *L. australis* has a widespread distribution across Australia, similar to the dominance *L. hinemoa* exhibits across New Zealand (Ponder 1971; Li et al., 2013). Three distinct biogeographic provinces distinguished primarily by intertidal community composition form the southern coast of Australia. *L. australis* is present in all three provinces, and a test was made to see if the species exhibits cryptic genetic structuring complementary to the provinces and, if so, what mechanisms were responsible. Variation in two mitochondrial genes (16S and COIII) and one nuclear gene (ITS2) was assayed to test for genetic structuring and to reconstruct the clam's phylogeny. The results showed that *L. australis* is composed of three cryptic mitochondrial clades corresponding to the biogeographic provinces. This cladogenesis is thought to have occurred in the Neogene. The support from both form of markers gives clear indication of genetic distinction between the clades of *L. australis* (Figure 2.2).

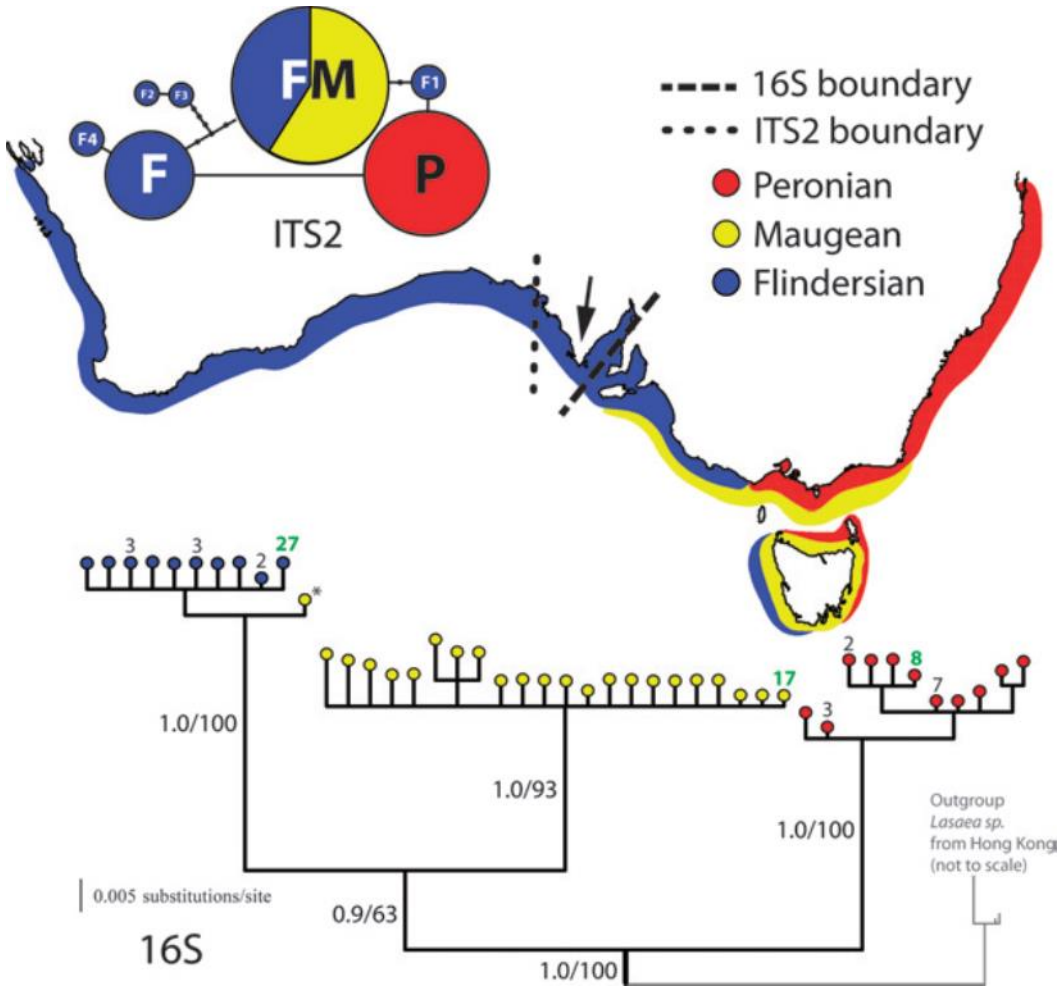


Figure 2.2: Figure taken from Li et al. (2013). a) ITS2 haplotype network. Each circle represents one unique haplotype. The size of each circle is proportional to numbers of that unique haplotype in the population and haplotypes are colour-coded according biogeographic province. Each black dot represents one inferred base pair change. The arrow on the map points out Port Lincoln on the Eyre Peninsula, where most clams shared the same haplotype as Maugean individuals. The inferred geographic boundaries between the Maugean and Flindersian lineages based on 16S (heavy dashed line) and ITS2 (light dashed line) are shown on the map, respectively.
b) Bayesian mitochondrial 16S phylogeny of *L. australis*. Clams haplotypes are colour-coded according to their biogeographic provinces of origin. Haplotype frequencies (N > 1) were indicated by the accompanying terminal number. Branch labels represent Bayesian posterior probabilities and maximum-likelihood bootstrap values, respectively. An ectopic Flindersian clade haplotype, recovered from a Maugean clam, is indicated by an asterisk.

Nuclear genes tend to evolve more slowly than mitochondrial genes (Brown, George & Wilson, 1979). In many groups of animals, the rate of nucleotide substitution among mitochondrial protein-coding genes is accelerated compared to the rate of nucleotide substitution among protein-coding regions of nuclear genes (Vawter & Brown, 1986).

In one study on marine bivalves, it was estimated that the substitution rate of ITS2 can be ten times slower than COI the mitochondrial gene (Faure, Jollivet, Tanguy, Bonhomme, & Bierne, 2009).

Past studies involving the phylogeny of New Zealand *Lasaea* appear to have focused on *L. hinemoa*. Taylor and Ó Foighil (2000) investigated populations of *Lasaea* from Japan, New Zealand, South Africa, Florida and Bermuda. Twelve mitochondrial genotypes were detected among the twenty *Lasaea* specimens sequenced from four global populations. Four haplotypes occurred in both the New Zealand and Japanese subsamples, three in the South African subsample and a single mitochondrial genotype was discovered in the Floridian subsample. COIII data indicated that this species of New Zealand *Lasaea* diverged into two distinct clades, and that one of these clades was more closely related to Kerguelen Island *Lasaea* than to the other New Zealand clade (Figure 2.1). An allozyme study of *L. hinemoa* showed genetic structure with a marked deviation from expected random mating patterns (within and among loci), frequent fixed heterozygosity, and reduced genotypic diversity (Taylor & Ó Foighil, 2000). This pattern and the finding of multiple asymmetric allozymic heterozygotes, indicated a clonal structure consistent with allopolyploid origins.

In this chapter populations across New Zealand and the Antipodes Island were examined and sequenced for the mitochondrial gene, Cytochrome Oxidase III (COIII) (n = 77), and the nuclear gene, Internal Transcribed Spacer 2 (ITS2) (n=69).

2.3 Methods

2.3a Sampling

Samples of *L. hinemoa* were collected from New Zealand and from Antipodes Island for the purposes of gene sequencing (Figure 2.3).

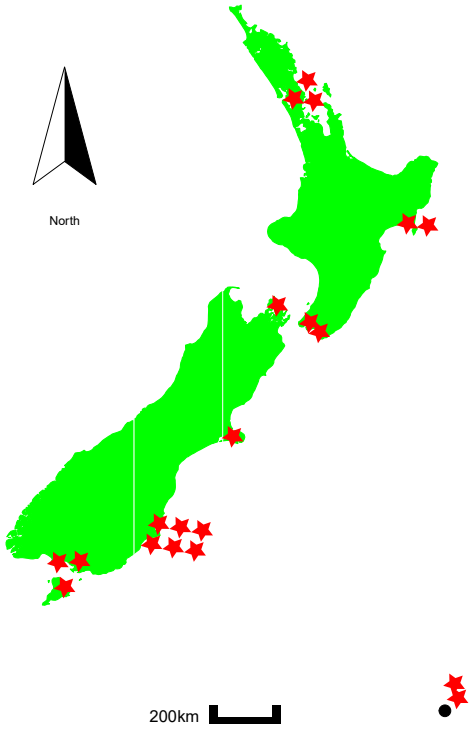


Figure 2.3: Distribution and Number of sites of *L. hinemoa* that were sequenced for Cytochrome Oxidase III (COIII) (n=20) and/or Internal Transcribed Tracer 2 (ITS2) (n=19) across New Zealand and Antipodes Island. Torbay, Auckland (36°41'44.8"S 174°45'13.0"E), Milford, Auckland (36°46'20.3"S 174°45'52.1"E), Takapuna, Auckland (36°47'16.7"S 174°46'12.7"E), Coronation Reserve, Mahia (39°05'38.0"S 177°57'00.4"E), Auroa Point, Mahia (39°05'17.6"S 177°57'30.7"E), Lyall Bay, Wellington (41°19'36.5"S 174°47'40.9"E), Moa Point, Wellington (41°20'15.1"S 174°48'38.7"E), Karaka Point, Picton (41°15'16.0"S 174°03'56.1"E), Children's Bay, Akaroa (43°48'04.6"S 172°57'45.9"E), Warrington, Dunedin (45°42'33.4"S 170°35'24.4"E), Broad Bay, Dunedin (45°50'48.6"S 170°37'35.5"E), Portobello Marine Lab, Dunedin (45°49'39.7"S 170°38'37.3"E), Quarantine Island, Dunedin (45°49'41.3"S 170°38'00.5"E), Portobello South Beach, Dunedin (45°49'47.8"S 170°38'37.0"E), Weller's Rock, Dunedin (45°47'51.6"S 170°42'55.5"E), Riverton Highway, Riverton (46°22'30.2"S 168°01'51.2"E) Riverton Rocks, Riverton (46°22'51.1"S 168°02'03.5"E), Port Pegasus, Stewart Island (47°12'47.2"S 167°41'05.2"E), Anchorage, Antipodes Island (49°39'53.9"S 178°47'59.1"E), Hutt Cove, Antipodes Island (49°40'00.6"S 178°48'21.2"E).

Numbers of *L. hinemoa* were not evenly distributed among sites, and numbers collected were approximate. Consultation with the Ngāi Tahu Research Consultation Committee was undertaken, and the committee responded favourably to the research being undertaken. Animals were stored in 99 % ethanol prior to DNA extraction. Full descriptions of sampling data can be found in Appendix A.

2.3b DNA Extraction and PCR Conditions

The Genomic DNA extraction was conducted using either a modified Chelex method (Casquet, Thebaud, & Gillespie, 2012) or a modified CTAB method (Doyle & Dickson, 1987) as described below.

The modified Chelex method featured suspending the whole animal in 300 µl of 5% chelex buffer, adding 1 µl of Proteinase K to digest the protein. Vortexing, then centrifuging and leaving in the heater (at 55°C) overnight. The next day leaving at the heat block for 10 minutes at 90°C, followed by centrifuging at 14000 rpm for 10 minutes, and storing at -4°C.

The modified CTAB method started with combining 300 µl of CTAB buffer, the whole animal and 5 µl of Proteinase K (20 mg/ml), and storing in the heater overnight at between 60-65°C. The next day 300 µl chloroform (1500 mg) was added under the fume hood and then inverted and centrifuged the mixture for 10 min. Next, the fume hood was used whilst taking off the supernatant into a new tube. 600 µl of cold ethanol was added and then 25 µl of NaOAc (3M), and followed by centrifuging for 10 minutes, pipetting off supernatant, and centrifuging for 30 seconds. Finally, the sample was air-dried/ exposed to the heat block (50°C) and resuspended in 25-50 µl MQH_2O (milli-Q H_2O).

After conducting initial DNA extractions, it was determined that DNA extractions conducted using CTAB buffer were generally higher, and this extraction method was used predominantly afterwards.

The primers chosen allowed the enzymatic amplification and direct sequencing of a 624-nt fragment (nucleotides 88-711) of the COIII gene (Ó Foighil & Smith, 1995). The amino terminal primer is 5'CATTTAGTTGATCCTAGGCC TTGACC-3' and the carboxy terminal primer is 5'CAAACCACATCTACAAAATGCCAATATC-3'. ITS2 was sequenced through a collection of different sets of primers, due to difficulties arising through the non-specificity of the initial primer set, and resulting contamination, and subsequently the poor amplification of other primer sets. The primer sets consisted of the universal primers Forward (5'-GGGTCGATGAAGAACGCAG-3') and Reverse (5'-GCTCTTCCCGCTTCACTCG-3') (Xu, Guo, Gaffney, & Pierce, 2001). LSU1 (5'-CTAGCTGCGAGAATTAATGTGA-3') and LSU-3 (5'-ACTTTCCCTCACGGTACTTG-3'), and LSU-5 5'-GTTAGACTCCTTGGTCCGTG-3 (Wade, Mordan & Naggs, 2006). A typical maximum of five individuals per location was amplified to ascertain a subset of variability. The sequencing of mitochondrial Cytochrome Oxidase III (COIII) allows phylogenetic analysis for these New Zealand

samples to be put into a global framework (Ó Foighil & Smith, 1995). The same individuals were sequenced for both genes, to allow more direct comparison of phylogenies, and to ascertain support in both COIII, which would show more evolutionary change, and in ITS2, which would be more conserved (Brown et al., 1979).

All PCRs (Polymerase Chain Reactions) made were conducted with a final volume of 15 µl with an Eppendorf Mastercycler Pro thermal cycler. PCRs were made with 5 µl of water, 1 µl of DNA extraction, 7.5 µl MyFi Mix, 0.25 µl of bovine serum albumin (BSA) and 0.75 µl of forward and reverse primer (0.5 µM).

Gradients were performed to ascertain the best annealing temperature; too low a temperature would give non-specific amplification and possible contamination, whilst too high would stop amplification of the target DNA.

For COIII the PCR cycler conditions consisted of denaturation at 95℃ for 3 minutes, followed by 40 cycles of 94℃ for 30 seconds, 40℃ annealing temperature (as in Ó Foighil & Smith, 1995), 72℃ for 1 minute for amplification, and finally an extension of 72℃ for 4 minutes.

For ITS2 the PCR cycler conditions consisted of denaturation at 95℃ for 3 minutes, followed by 40 cycles of 94℃ for 30 seconds, 55℃ annealing temperature, 72℃ for 1 minute for amplification, and finally an extension of 72℃ for 4 minutes.

All PCR products were visualised by running 2 µl of the reaction on a 1% agarose gel and imaging with a UVITEC HD5 gel imager (Cambridge, United Kingdom). After PCR purification, sequences were sent to Genetics Analysis Services, Anatomy Department, University of Otago where they were sequenced via ABI sequencing (ABI 3730, Applied Biosystems).

2.3c Sequence Processing

Prior to analysis, sequences were uploaded to BLAST (Basic Local Alignment Search Tool) (Altschul, Gish, Miller, Myers, & Lipman, 1990) to confirm their species identity and ensure that no contamination occurred. Sequences were examined for incorrect base calls, and ambiguities and were initially trimmed using Sequencher v.5.1 (Genecodes, Ann Arbor, MI) to ensure quality across the sequence length. When sequence quality was poor the forward primer was also sequenced so a consensus sequence could be made. COIII sequences were then uploaded to Geneious v.11.1.5 (https://www.geneious.com, Kearse et al., 2012) and aligned using the MAFFT algorithm (Katoh & Standley, 2013). Sequences were trimmed to the same length at this stage (598bp).

To determine the reading frame for COIII the sequences were translated by specifying the appropriate code (invmtDNA), and through starting at different positions in the alignment and examining the resulting data for stop codons. Starting the sequence on the first base was concluded to be the most accurate.

ITS2 sequences were also uploaded to Geneious after being examined in Sequencher and aligned. All sequences were trimmed to the same length (475bp).

The combined tree data had a sequence length of 1073bp.

The sequences for *L. australis* used in the combined tree were constructed by reviewing supplementary data from Li et al. (2013) and identifying individuals that had been entered into Genbank (https://www.ncbi.nlm.nih.gov/genbank/) for both COIII and ITS2 sequences. This discovery led to two individuals being constructed as acceptable outgroups to the *L. hinemoa* sequenced.

2.3d Phylogenetic Analysis

The phylogenetic relationship between *L. hinemoa* and other *Lasaea* was examined by creating single gene trees (COIII and ITS2) using Bayesian and maximum likelihood

methods, and by creating a combined gene tree. These analyses were conducted utilising the online phylogenetic tree tool CIPRES Science Gateway v.3.3 (Miller, Pfeiffer, & Schwartz, 2010). Outgroups were taken from Genbank (https://www.ncbi.nlm.nih.gov/genbank/) (Table 2.1). The species outgroup was *Kellia laperousi*.

Table 2.1: Species identification and GenBank accession numbers for all additional sequences used in the phylogenetic analyses in this study.

Taxon label	**COIII**	**ITS2**	**Reference**
Kellia laperousii	X78289		Ó Foighil & Smith, 1995
Lasaea (Portugal)	AF112280		Ó Foighil & Jozefowicz, 1999
Lasaea (Portugal)	AF112281		Ó Foighil & Jozefowicz, 1999
Lasaea (Spain)	AF112282		Ó Foighil & Jozefowicz, 1999
Lasaea (Spain)	AF112283		Ó Foighil & Jozefowicz, 1999
Lasaea (Florida, USA)	AF112284		Ó Foighil & Jozefowicz, 1999; Taylor & Ó Foighil, 2000
Lasaea (Japan)	AF276029		Taylor & Ó Foighil, 2000
Lasaea (South Africa)	AF276031		Taylor & Ó Foighil, 2000
Lasaea (Hong Kong)	JX910453		Li et al., 2013
Lasaea (Hong Kong)	JX910454		Li et al., 2013
L. australis	JX910455		Li et al., 2013
L. australis	JX910456		Li et al., 2013
L. australis	JX910457		Li et al., 2013
L. australis	JX910458	JX910468	Li et al., 2013

L. australis	JX910459	JX910469	Li et al., 2013
L. australis	JX910460		Li et al., 2013
L. australis	JX910461		Li et al., 2013
L. australis	JX910462		Li et al., 2013
L. australis	JX910463		Li et al., 2013
LundSdy02 (Australia)	JX910464		Li et al., 2013
LundSdy03 (Australia)	JX910465		Li et al., 2013
L. colmani	JX910466		Li et al., 2013
L. australis	X78290		Ó Foighil & Smith, 1995
Lasaea (Australia)	X78291		Ó Foighil & Smith, 1995
Lasaea (Ireland)	X78292		Ó Foighil & Smith, 1995
Lasaea (Australia) (*colmani*?)	X78293		Ó Foighil & Smith, 1995
Lasaea (Ireland)	X78294		Ó Foighil & Smith, 1995
Lasaea (Ireland)	X78295		Ó Foighil & Smith, 1995
Lasaea (France)	X78296		Ó Foighil & Smith, 1995
Lasaea (Canada)	X78297		Ó Foighil & Smith, 1995
Lasaea (Canada)	X78298		Ó Foighil & Smith, 1995
Lasaea (Canada)	X78299		Ó Foighil & Smith, 1995
Lasaea (Canada)	X78300		Ó Foighil & Smith, 1995

Lasaea (Kerguelen Island)	X78301		Ó Foighil & Smith, 1995
Lasaea (New Zealand)	AF276032		Taylor & Ó Foighil, 2000
Lasaea (New Zealand)	AF276033		Taylor & Ó Foighil, 2000
L. australis		JX910467	Li et al., 2013
L. australis		JX910470	Li et al., 2013
L. australis		JX910471	Li et al., 2013
L. australis		JX910472	Li et al., 2013

Model selection and optimal partitioning schemes were determined using PartitionFinder v.2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016), which implements the greedy algorithm (Lanfear, Calcott, Ho, & Guindon, 2012) and utilizes the software PhyML v.3.0 (Guindon et al., 2010).

For Bayesian analysis, model testing was restricted to implementation in MrBayes and the Bayesian information criterion (BIC) was used for evaluating the likelihood of the proposed models and partitioning schemes. Bayesian analyses were performed in MrBayes v.3.2.6 (Ronquist et al., 2012). For COIII 2 optimal models were identified for different codon positions. For codons 1 and 2 HKY+I+G was identified, whilst for codon 3 GTR+G was identified. Two runs, each consisting of four Markov chains, were run for 10,000,000 generations and were sampled every 1,000 steps. The first 2,500 trees (25%) were discarded as burn-in and posterior probabilities were obtained from a majority-rule consensus. Mixing and convergence of each run was monitored by the standard deviation of statistics provided in MrBayes and Tracer v.1.7 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) was used to assess the convergence and mixing of the two Bayesian runs and effective sample sizes (ESS).

For maximum likelihood methods, the RAxML v.8 (Stamatakis, 2006) default model of GTR can be used for all markers. Bootstrap support values were obtained using the rapid bootstrap method with 1,000 replicates.

Mega X (Molecular evolutionary genetics analysis) (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). was used to calculate the genetic distance between different COIII clades (Clade I, Clade II, Clade III and Clade IV). Between group average was calculated using the Tamura-Nei model (Tamura & Nei, 1993) (as this was closest to the evolutionary model identified by PartionFinder, for the COIII sequences).

2.3e Population Genetics

The number of haplotypes, number of polymorphic sites, haplotypic diversity and nucleotide diversity was calculated using Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010), to demonstrate population differences.

PopART (Population Analysis with Reticulate Trees) (Leigh, & Bryant, 2015) was utilized in order to create a haplotype map for COIII and ITS2 to demonstrate the dominance of the clades across New Zealand. TCS networks were also developed for COIII and ITS2 to demonstrate population genetics (Clement, Posada, & Crandall, 2000).

An AMOVA (Analysis of Molecular Variance) (Excoffier, Smouse, & Quattro, 1992) was constructed in Arlequin v.3.5.2.2, to investigate genetic distinction of geographically separated COIII. COIII was used as it has a higher evolutionary rate than the conserved ITS2 (Brown et al., 1979). The genetic distance was determined using the Tamura-Nei model, as it most closely aligned with the model determined by model selection in PartionFinder. The results of the AMOVA analysis were tested for significance by 10,000 random permutations. Three groups of sequences were made, consisting of North (Auckland, Mahia, Wellington and Picton), South (Akaroa, Dunedin, Riverton and Stewart Island) and Antipodes Island. This grouping allowed insight into the North-South split present in other species (Sponer & Roy, 2002), and enabled comparison between New Zealand and Antipodes Island.

DAPC (Discriminant Analysis of Population Components) is a multivariate method that identifies and describe clusters of genetically related individuals (Jombart, Devillard, & Balloux, 2010). It was implemented in R v.3.5.2 (https://www.r-project.org/). The code for this analysis can be found in Appendix C. *DAPC* does not require group priors to be present, and instead uses sequential K-means and model selection to infer genetic clusters. Before analysis commenced, aligned ingroup *L. hinemoa* COIII sequences were imported into R using the *Ape* package (Paradis, Claude, & Strimmer, 2004). The *Adegenet* package (Jombart, 2008) was utilized to extract the single nucleotide polymorphism (SNP) sites. A total of 190 SNPs from 77 specimens was obtained. *DAPC* was implemented with the function *dapc*, which transforms the data using Principal Component Analysis (*PCA*), and then performs a Discriminant Analysis on the retained Principal Components. *DAPC* benefits from not utilizing too many principal components, providing it does not compromise the loss of genetic information (Jombart et al., 2010). Three principal components of the *PCA* were retained in the preliminary data transformation, which contained 87% of the conserved genetic variation. Subsequently, two discriminant functions were retained, capturing the most amount of information in the eigenvalues. Basic scatterplots were then obtained using the function *scatterplot.*

Population pairwise F_{ST} values indicating the amount of genetic differentiation between different populations were calculated in Arlequin using the Tamura and Nei model.

2.4 Results

2.4a Sampling

DNA extractions were completed for roughly 30 individuals per location. It was attempted to have individuals that were sequenced for COIII to be also sequenced for ITS2. The success was higher for COIII PCRs than for ITS2 PCRs but success was still low.

2.4b DNA Sequencing and Phylogenetics

The COIII ingroup dataset consists of 598bp and 77 sequences, with 190 variable sites. The GC% is 37.8%. The average ratio of transitions/transversions equated to R = 1.9. Nucleotide composition consisted of T(U) = 41.6% C = 13.5% A = 20.6% and G = 24.3%. In both phylogenetic analyses conducted, Bayesian and maximum likelihood methods yielded very similar topologies for both single-gene datasets, and so only Bayesian 50% majority-rule inference trees are shown here (Figures 2.4 & 2.5) (maximum likelihood trees are shown in Appendix B).

In the Bayesian analyses conducted for both single-gene trees separate runs converged to an average deviation split below 0.01, demonstrating good mixing, with ESS (Effective Sample Size) values above 200. It is considered good practice for the Bayesian MCMC results to be accompanied by a critical assessment of convergence (Ronquist, et al., 2012). By comparing samples obtained from independent MCMC analyses this can be accomplished. The average standard deviation of split frequencies (ASDSF) in MrBayes allows a quantitative appraisal of the similarity among such samples. ASDSF should approach 0.0 as runs converge to the same distribution (Ronquist, et al., 2012). The trees presented have Bayesian posterior probability values indicated below the nodes and maximum likelihood values above.

The COIII tree is very important as it allows *L. hinemoa* to be put into a global framework as more other *Lasaea* have been sequenced for this gene than any other (Taylor & Ó Foighil, 2000). *L. hinemoa* forms four distinct groups with strong support (Clades I, II, III and IV) (Figure 2.4). As can be seen there is some clear geographic splitting as Clade I of *L. hinemoa* appears to have few individuals from the top of the North Island. Clade IV is a very divergent clade comprising of two individuals that cluster with *Lasaea* found from Australia, Hong Kong and Japan (Table 2.2). These two individuals originate from Picton and Mahia. Genetic distance data supports Clade IV being the most different from the other three clades, whilst Clade I and Clade II are the most similar to each other (Table 2.2).

The ITS2 tree shows less variation as it evolves more slowly than the mitochondrial gene (Brown et al., 1979). This lack of strucutre allows very clear differences between

groups to be seen. This variation that we do see provides further support for the disjunction between the four clades observed (with slight variation seen in Clade II) (Figure 2.5). What is also note worthy is the distinction between Clade I and Clade II for individuals from Antipodes Island. Managing to have the nuclear gene amplify for an individual from the Clade IV (APM_01) was extremely useful, as it confirms that these individuals fall outside of the two major *L. hinemoa* clades.

The combined-gene tree allows inference into the faster rate of evolutionary changes reported by COIII and the more constrained differences enforced by ITS2 (Brown et al., 1979) (Figure 2.6). There is a clear separation into three main clades (Clades I, II and III). The clear distinction between the individual (APM_01) and the other New Zealand *L. hinemoa*, suggests it is more closely related to the outgroup *L. australis*, that it clusters with. The support values for this distinction are high.

Table 2.2: Genetic Distance data for the COIII gene measured by between group mean distance using the Tamura-Nei model.

	Clade I	**Clade II**	**Clade III**
Clade I			
Clade II	0.0582		
Clade III	0.1386	0.1319	
Clade IV	0.2977	0.2917	0.2916

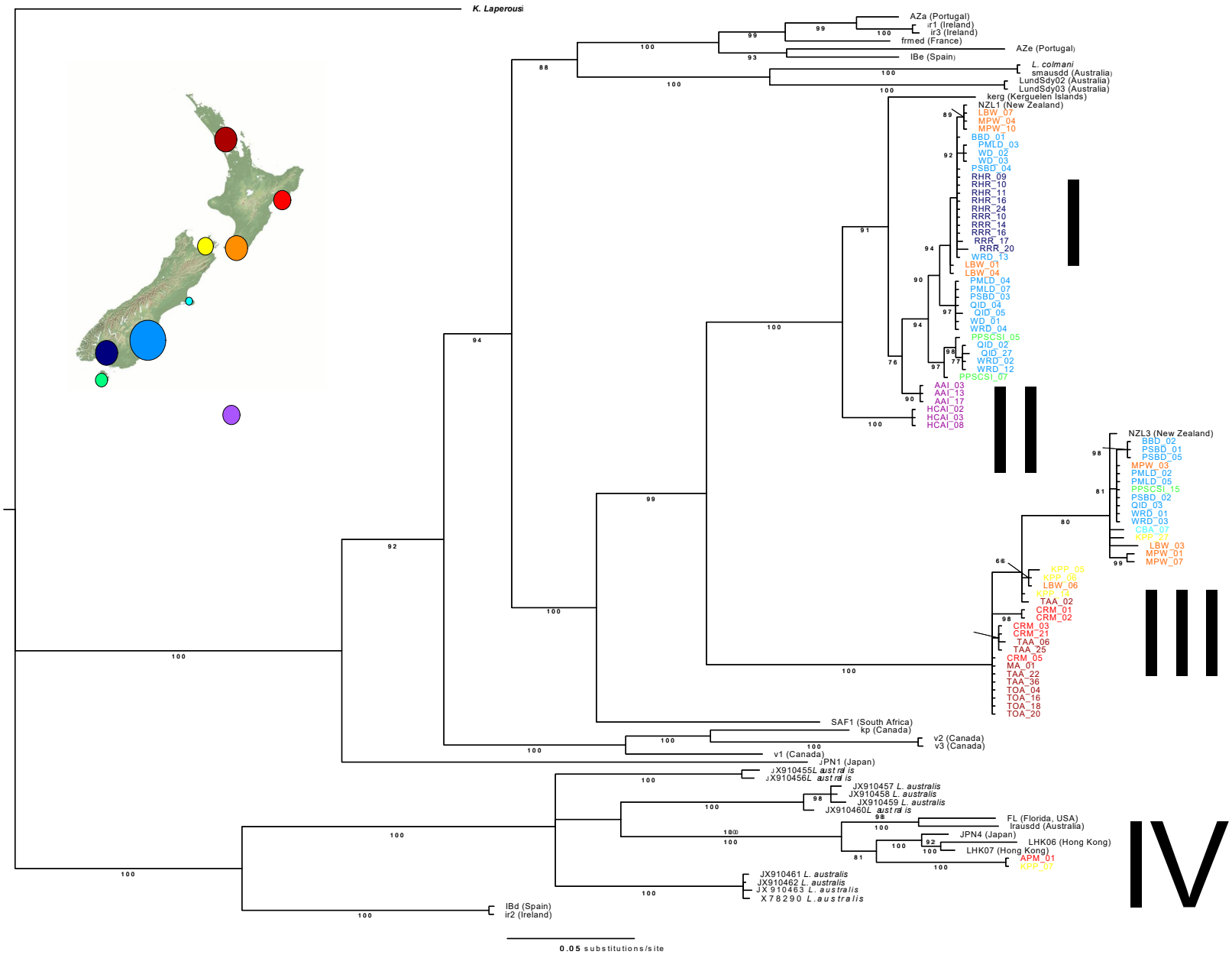
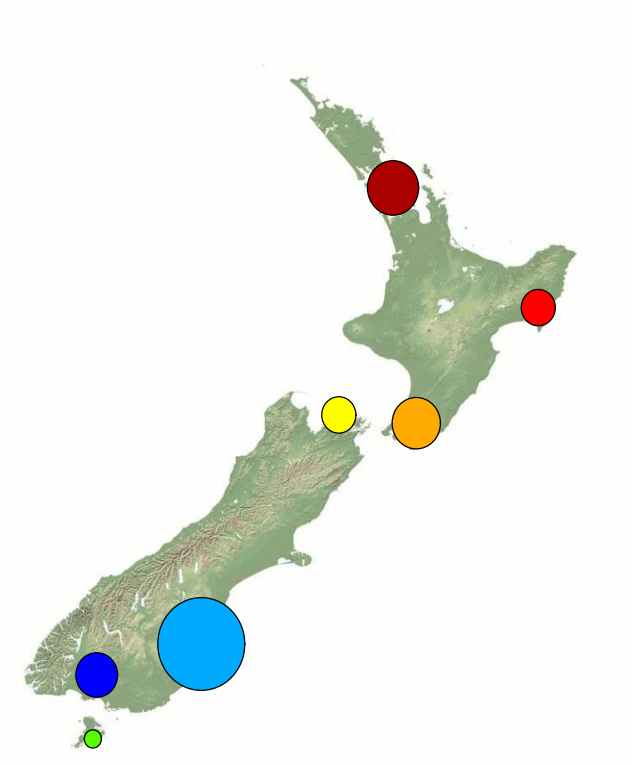

K. Laperousi
AZa (Portugal)
ir1 (Ireland)
ir3 (Ireland)
frmed (France)
AZe (Portugal)
IBe (Spain)
L. colmani
smausdd (Australia)
LundSdy02 (Australia)
LundSdy03 (Australia)
kerg (Kerguelen Islands)
NZL1 (New Zealand)
LBW_07
MPW_04
MPW_10
BBD_01
PMLD_03
WD_02
WD_03
PSBD_04
RHR_09
RHR_10
RHR_11
RHR_16
RHR_24
RRR_10
RRR_14
RRR_16
RRR_17
RRR_20
WRD_13
LBW_01
LBW_04
PMLD_04
PMLD_07
PSBD_03
QID_04
QID_05
WD_01
WRD_04
PPSCSI_05
QID_02
QID_27
WRD_02
WRD_12
PPSCSI_07
AAI_03
AAI_13
AAI_17
HCAI_02
HCAI_03
HCAI_08
NZL3 (New Zealand)
BBD_02
PSBD_01
PSBD_05
MPW_03
PMLD_02
PMLD_05
PPSCSI_15
PSBD_02
QID_03
WRD_01
WRD_03
CBA_07
KPP_27
LBW_03
MPW_01
MPW_07
KPP_05
KPP_08
LBW_06
KPP_14
TAA_02
CRM_01
CRM_02
CRM_03
CRM_21
TAA_06
TAA_25
CRM_05
MA_01
TAA_22
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
SAF1 (South Africa)
kp (Canada)
v2 (Canada)
v3 (Canada)
v1 (Canada)
JPN1 (Japan)
JX910455 L. australis
JX910456 L. australis
JX910457 L. australis
JX910458 L. australis
JX910459 L. australis
JX910460 L. australis
FL (Florida, USA)
Irausdd (Australia)
JPN4 (Japan)
LHK06 (Hong Kong)
LHK07 (Hong Kong)
APM_01
KPP_07
JX910461 L. australis
JX910462 L. australis
JX 910463 L. australis
X 78290 L. australis
IBd (Spain)
ir2 (Ireland)
0.05 substitutions/site
IV

Figure 2.4: Bayesian tree for COIII sequences. maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst Bayesian posterior probabilities (values below 75% are not shown) are indicated below nodes. Locations for *L. hinemoa* are grouped by main localities (Auckland (TOA = Torbay, Auckland, TAA = Takapuna, Auckland), Mahia (APM = Auroa Point, Mahia, CRM = Coronation Reserve, Mahia), Wellington (MPW = Moa Point, Wellington, LBW = Lyall Bay, Wellington), Picton (KPP = Karaka Point, Picton), Akaroa (CBA = Children's Bay, Akaroa), Dunedin (WD = Warrington, Dunedin, BBD = Broad Bay, Dunedin, PSBD = Portobello South Beach, Dunedin, PMLD = Portobello Marine Lab, Dunedin, QID = Quarantine Island, Dunedin, WRD = Weller's Rock, Dunedin), Riverton (RHR = Riverton Highway, Riverton, RRR = Riverton Rocks, Riverton), Stewart Island (Port Pegasus, Shipbuilder's Cove, Stewart Island) and Antipodes Island (HCAI = Hutt Cove, Antipodes Island, AAI = Anchorage, Antipodes Island)), and corresponding colours for locality are indicated on the tree. Explanation for the individual labels can be found in Table 2.2 and Appendix A. Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_02-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01-KPP_07).



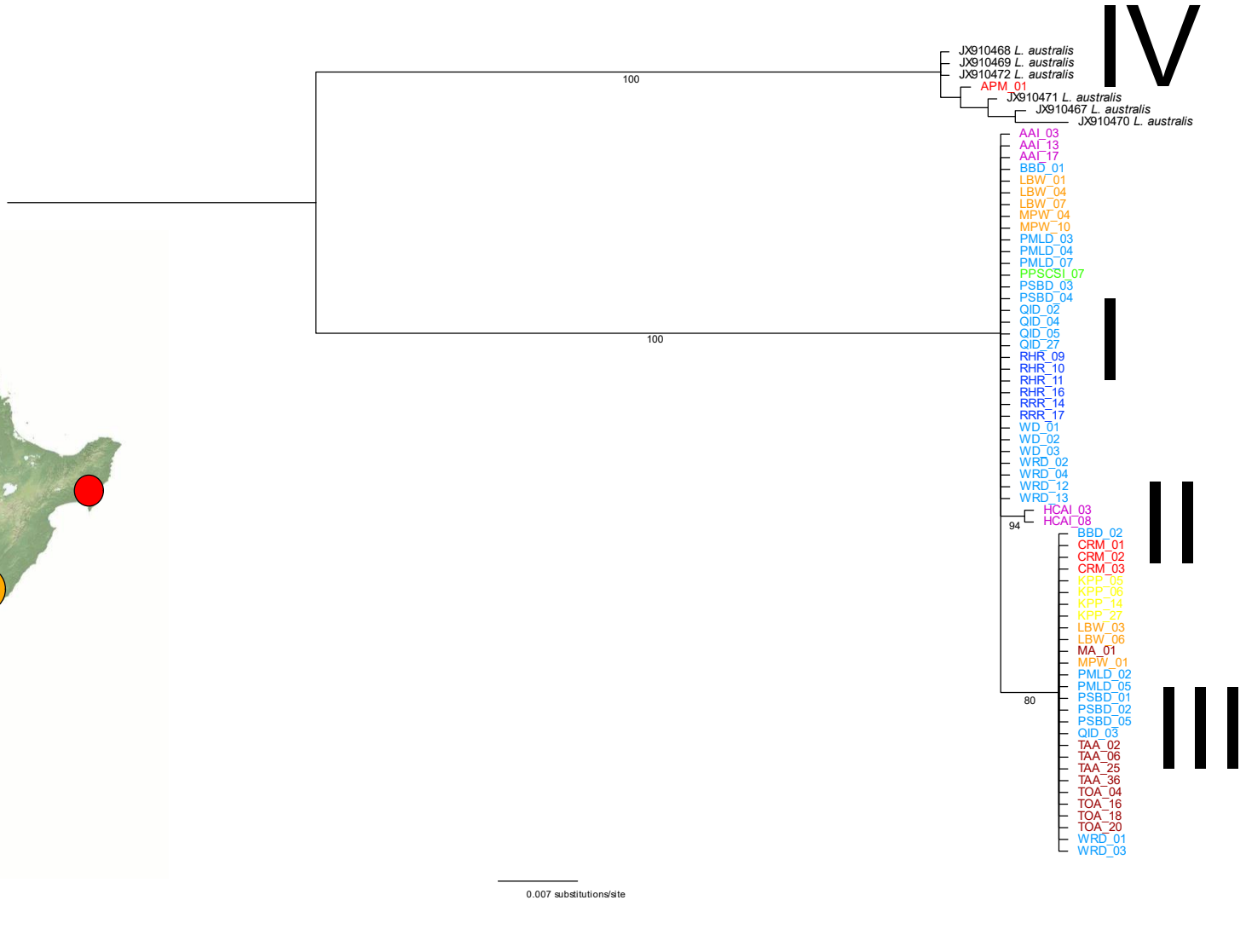

IV
I
II
III
JX910468 L. australis
JX910469 L. australis
JX910472 L. australis
APM_01
JX910471 L. australis
JX910467 L. australis
JX910470 L. australis
AAI_03
AAI_13
AAI_17
BBD_01
LBW_01
LBW_04
LBW_07
MPW_04
MPW_10
PMLD_03
PMLD_04
PMLD_07
PPSCSI_07
PSBD_03
PSBD_04
QID_02
QID_04
QID_05
QID_27
RHR_09
RHR_10
RHR_11
RHR_16
RRR_14
RRR_17
WD_01
WD_02
WD_03
WRD_02
WRD_04
WRD_12
WRD_13
HCAI_03
HCAI_08
BBD_02
CRM_01
CRM_02
CRM_03
KPP_05
KPP_06
KPP_14
KPP_27
LBW_03
LBW_06
MA_01
MPW_01
PMLD_02
PMLD_05
PSBD_01
PSBD_02
PSBD_05
QID_03
TAA_02
TAA_06
TAA_25
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
WRD_01
WRD_03
100
100
94
80
0.007 substitutions/site

Figure 2.5: Bayesian tree for ITS2 sequences, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst posterior probabilities (values below 75% are not shown) are indicated below nodes. Locations for *L. hinemoa* are grouped by main localities (localities (Auckland (TOA = Torbay, Auckland, TAA = Takapuna, Auckland), Mahia (APM = Auroa Point, Mahia, CRM = Coronation Reserve, Mahia), Wellington (MPW = Moa Point, Wellington, LBW = Lyall Bay, Wellington), Picton (KPP = Karaka Point, Picton), Akaroa (CBA = Children's Bay, Akaroa), Dunedin (WD = Warrington, Dunedin, BBD = Broad Bay, Dunedin, PSBD = Portobello South Beach, Dunedin, PMLD = Portobello Marine Lab, Dunedin, QID = Quarantine Island, Dunedin, WRD = Weller's Rock, Dunedin), Riverton (RHR = Riverton Highway, Riverton, RRR = Riverton Rocks, Riverton), Stewart Island (Port Pegasus, Shipbuilder's Cove, Stewart Island) and Antipodes Island (HCAI = Hutt Cove, Antipodes Island, AAI = Anchorage, Antipodes Island), and corresponding colours for locality are indicated on the tree. Explanation for the individual labels can be found in Table 2.2 and Appendix A. Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_03-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01).

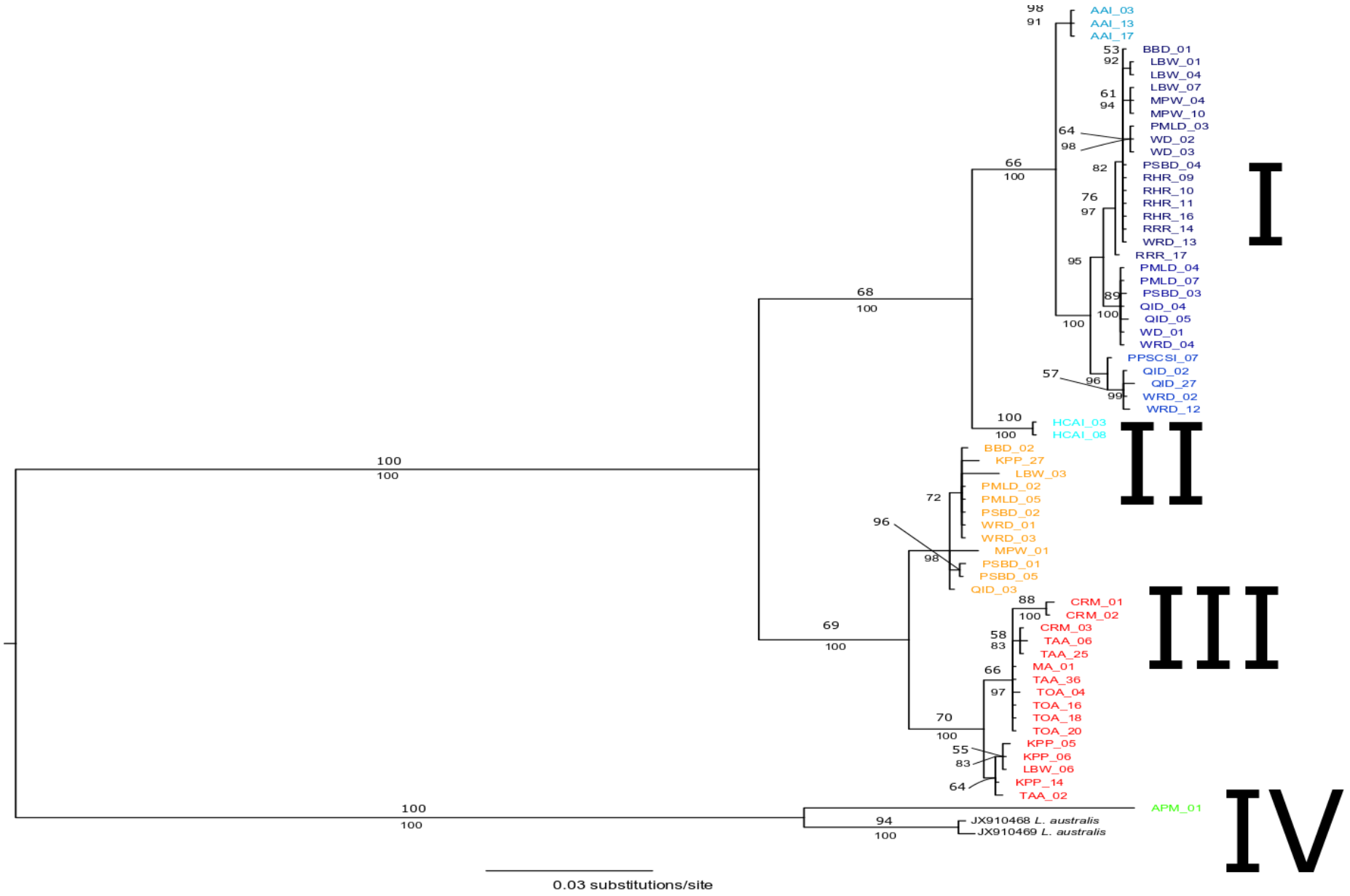

AAI_03
AAI_13
AAI_17
BBD_01
LBW_01
LBW_04
LBW_07
MPW_04
MPW_10
PMLD_03
WD_02
WD_03
PSBD_04
RHR_09
RHR_10
RHR_11
RHR_16
RRR_14
WRD_13
RRR_17
PMLD_04
PMLD_07
PSBD_03
QID_04
QID_05
WD_01
WRD_04
PPSCSI_07
QID_02
QID_27
WRD_02
WRD_12
HCAI_03
HCAI_08
BBD_02
KPP_27
LBW_03
PMLD_02
PMLD_05
PSBD_02
WRD_01
WRD_03
MPW_01
PSBD_01
PSBD_05
QID_03
CRM_01
CRM_02
CRM_03
TAA_06
TAA_25
MA_01
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
KPP_05
KPP_06
LBW_06
KPP_14
TAA_02
APM_01
JX910468 L. australis
JX910469 L. australis
I
II
III
IV
0.03 substitutions/site

Figure 2.6: Combined COIII and ITS2 Bayesian tree, Bayesian posterior probabilities (values less than 75% are not included) are indicated below the nodes whilst maximum likelihood bootstrap support values (values less than 50% are not indicated) are recorded above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A. Localities (Auckland (TOA = Torbay, Auckland, TAA = Takapuna, Auckland), Mahia (APM = Auroa Point, Mahia, CRM = Coronation Reserve, Mahia), Wellington (MPW = Moa Point, Wellington, LBW = Lyall Bay, Wellington), Picton (KPP = Karaka Point, Picton), Akaroa (CBA = Children's Bay, Akaroa), Dunedin (WD = Warrington, Dunedin, BBD = Broad Bay, Dunedin, PSBD = Portobello South Beach, Dunedin, PMLD = Portobello Marine Lab, Dunedin, QID = Quarantine Island, Dunedin, WRD = Weller's Rock, Dunedin), Riverton (RHR = Riverton Highway, Riverton, RRR = Riverton Rocks, Riverton), Stewart Island (Port Pegasus, Shipbuilder's Cove, Stewart Island) and Antipodes Island (HCAI = Hutt Cove, Antipodes Island, AAI = Anchorage, Antipodes Island)Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_03-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01)

2.4c Population Genetics Analyses

ITS2 evolves more slowly than COIII but that allows clear distinctions to be seen from the clades produced through sequencing (Brown et al., 1979). The TCS network produced shows the distinction of the four main clades from New Zealand (Figure 2.7a). The distinct Antipodes clade has diverged from one of the main clades. Interestingly, the highly divergent individual (APM_01) clusters with the Australian sequences, rather than *L. hinemoa*. *L. hinemoa* appears to have two main clades, that exhibit some geographic separation, with the blue clade (Clade I) being more dominant in the South whilst the red clade (Clade III) is more dominant in the North (Figure 2.7b).

The number of haplotypes, number of polymorphic sites, haplotypic diversity and nucleotide diversity was calculated for COIII data using Arlequin (Table 2.3). Thirty unique haplotypes were identified, and 189 polymorphic sites. Standard indices of genetic diversity across populations can be seen in Table 2.3.

COIII allows faster inferences to be drawn as it evolves more quickly than ITS2 (Brown et al., 1979). The TCS network produced shows the distinction of two individuals from Mahia and Picton from the rest of those sequenced (Figure 2.8a). Several groups cluster within the TCS network. *L. hinemoa* appears to have two main clades, that experience some geographic distinction, with the blue clade (Clade I) being more dominant in the South whilst the red clade (Clade III) is more dominant in the North (Figure 2.8b). There appear to be 15 shared haplotypes, whilst other unique haplotypes are represented by white clades.

The AMOVA tested the relationship between the Northern New Zealand groups (Auckland, Mahia, Wellington and Picton), Southern (Akaroa, Dunedin, Riverton and Stewart Island) and Antipodes Island. The variation observed was greatest within populations, and smallest among populations within groups (Table 2.4).

DAPC (Discriminant analysis of population components) (Figure 2.9) shows a distinction of two main groups for the COIII sequence data with Auckland and Mahia dominating one group. Whilst the other group was dominated by individuals from further south.

F_{ST} values were evaluated between populations (Table 2.5) and gave support for population level differences.

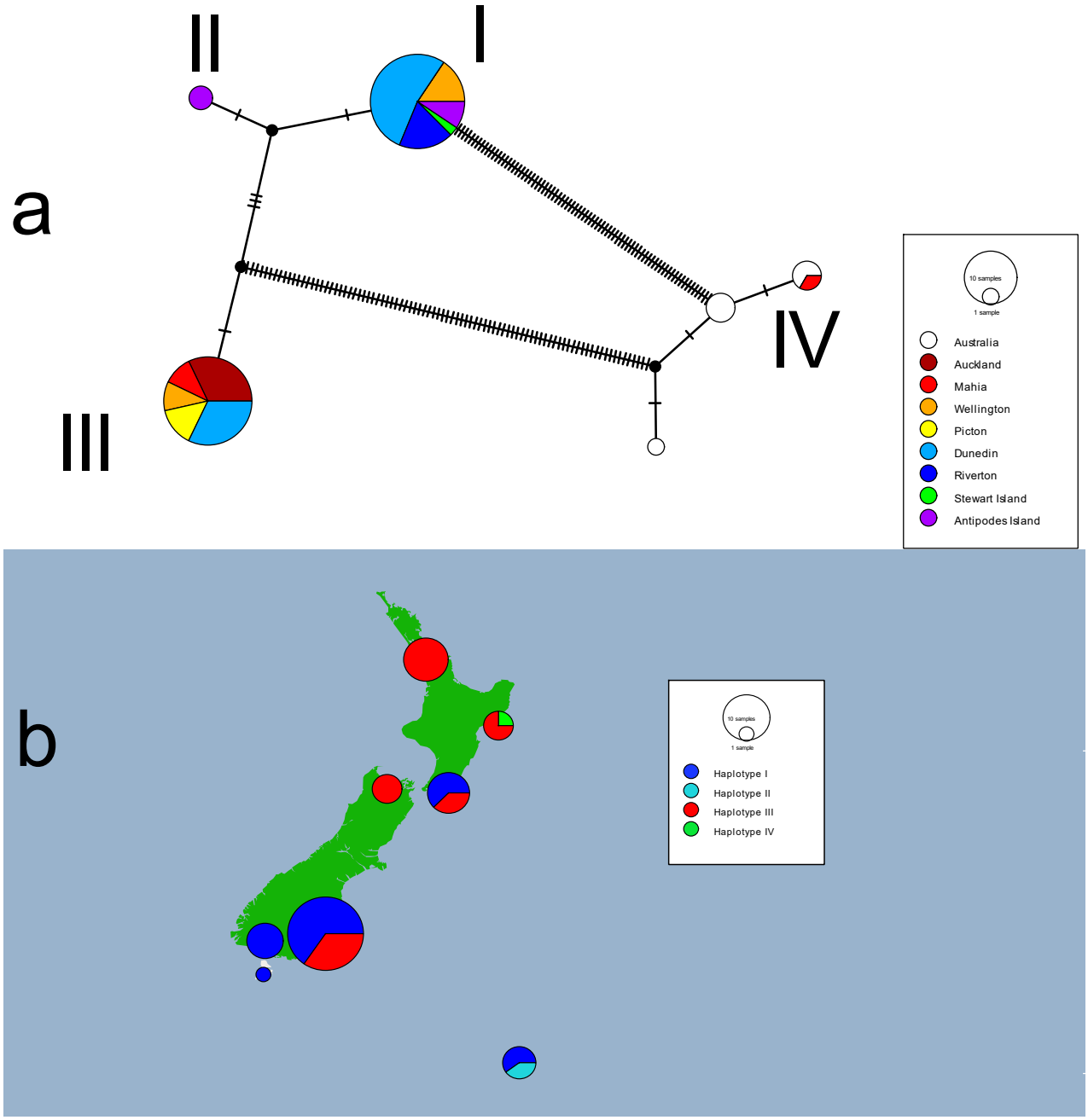


Figure 2.7: a) TCS network for ITS2 demonstrating the haplotypes observed, and the distribution of locations within those haplotypes. Hatch marks indicate number of mutational steps and dark circles represent inferred, unsampled haplotypes. b) The distribution of the four ITS2 haplotypes seen across New Zealand.

Table 2.3: Population genetic analysis summary statistics, N = number of individuals, N_h = Number of haplotypes, N_p = number of polymorphic sites, H = haplotype diversity, Π = nucleotide diversity.

Location	**N**	**N_h**	**N_p**	**H**	**Π**
Auckland	10	4	10	0.5333±0.1801	0.003605±0.002466
Mahia	6	4	145	0.8667±0.1291	0.084504±0.049460
Wellington	10	7	93	0.9111±0.0773	0.074693±0.040094
Picton	5	5	151	1.0000±0.1265	0.107860±0.065999
Akaroa	1	1	0	1.0000±0.0000	0.000000±0.000000
Dunedin	26	8	77	0.8708±0.0326	0.059974±0.030051
Riverton	10	3	3	0.3778±0.1813	0.001003±0.000982
Stewart Island	3	3	70	1.0000±0.2722	0.078595±0.059325
Antipodes Island	6	2	26	0.6000±0.1291	0.026087±0.015723

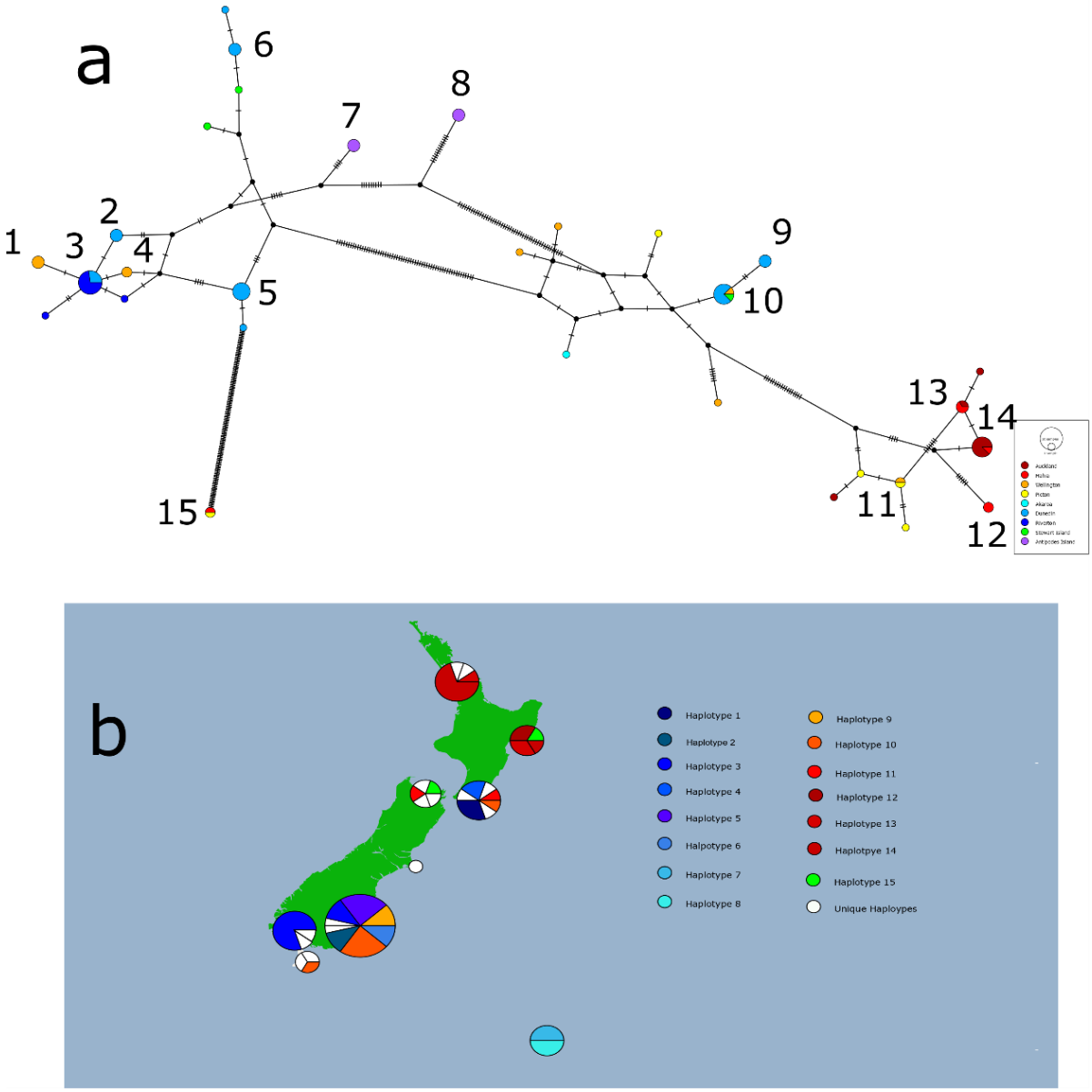


Figure 2.8: a) TCS network for COIII demonstrating the clades observed, and the distribution of locations within those clades. Hatch marks indicate number of mutational steps and dark circles represent inferred, unsampled haplotypes. b) The distribution of the COIII haplotypes seen across New Zealand.

Table 2.4: Results of AMOVA, including percentage of variation (%), fixation indices and p-values, d.f.= degrees of freedom.

Source of Variation	**d.f.**	**Percentage of Variation (%)**	**Fixation Indices**	**p-value**
Among Groups	2	29.44	0.28413	0.00000±0.00000
Among populations within groups	6	20.05	0.49487	0.00010±0.00010
Within populations	68	50.51	0.29438	0.04871±0.00206
Total	**76**			

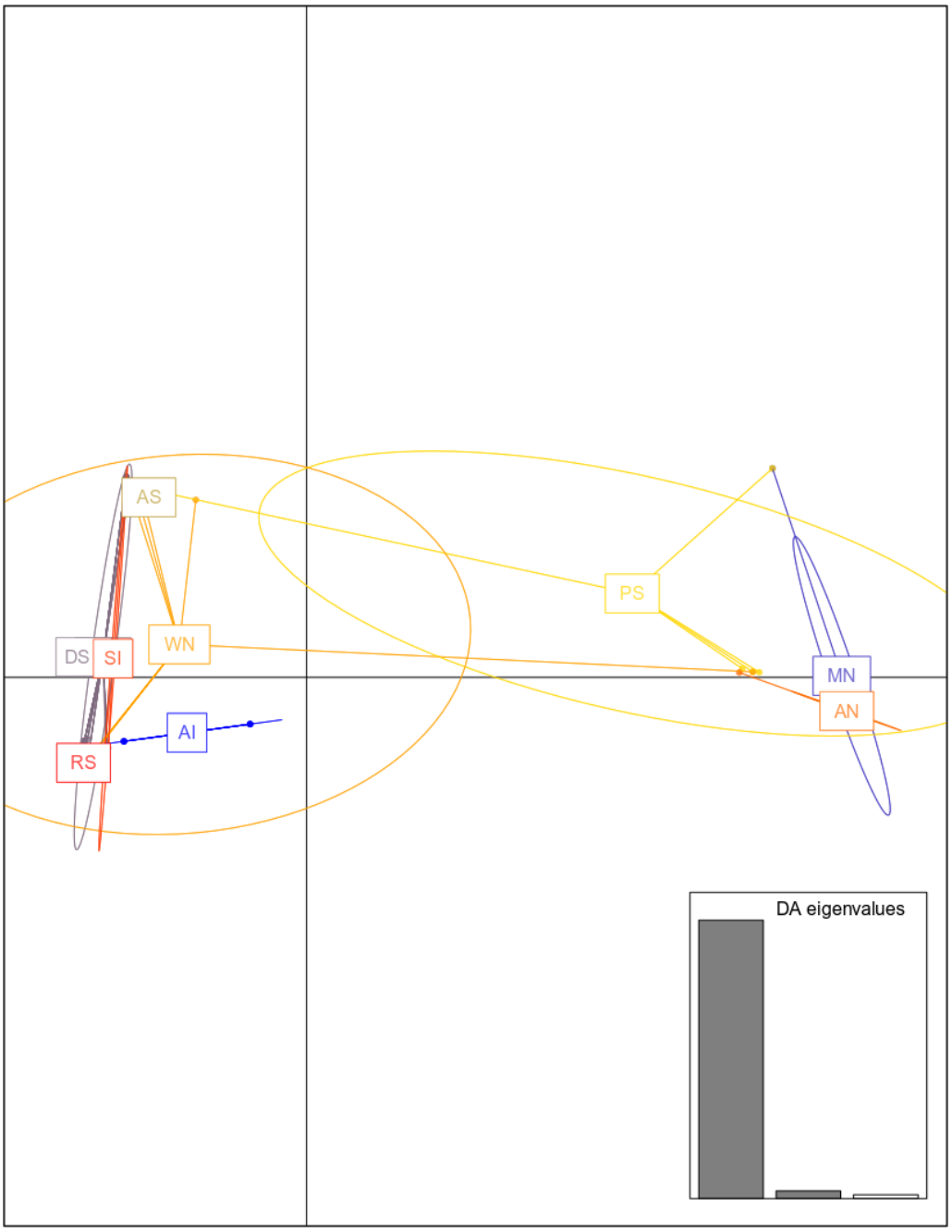


Figure 2.9: Scatterplot of the DAPC analysis, each point represents an individual (not the circles), and ellipses represent 95% confidence limits for locality-level groups (AN=Auckland, North Island, MN=Mahia, North Island, WN=Wellington, North Island, PS=Picton, South Island, AS=Akaroa, South Island, DS=Dunedin, South Island, RS=Riverton, South Island, SI=Stewart Island and AI=Antipodes Island).

Table 2.5: Population pairwise F_{ST} values calculated using the Tamura and Nei model between the main sampling locations (Auckland, Mahia, Wellington, Picton, Akaroa, Dunedin, Riverton, Stewart Island and Antipodes Island).

	Auckland	**Mahia**	**Wellington**	**Picton**	**Akaroa**	**Dunedin**	**Riverton**	**Stewart Island**	**Antipodes Island**
Auckland	0.00000	0.10977	0.55041	0.27902	0.92317	0.60501	0.98461	0.80219	0.91102
Mahia	0.10977	0.00000	0.29861	0.00000	0.00000	0.45994	0.77265	0.31495	0.60077
Wellington	0.55041	0.29861	0.00000	0.21841	0.00000	0.00000	0.40844	0.00000	0.31217
Picton	0.27902	0.00000	0.21841	0.00000	0.00000	0.40310	0.74913	0.20824	0.55938
Akaroa	0.92317	0.00000	0.00000	0.00000	0.00000	0.20613	0.99217	0.00000	0.77951
Dunedin	0.60501	0.45994	0.00000	0.40310	0.20613	0.00000	0.25753	0.00000	0.26484
Riverton	0.98461	0.77265	0.40844	0.74913	0.99217	0.25753	0.00000	0.56727	0.72749
Stewart Island	0.80219	0.31495	0.00000	0.20824	0.00000	0.00000	0.56727	0.00000	0.24910
Antipodes Island	0.91102	0.60077	0.31217	0.55938	0.77951	0.26484	0.72749	0.24910	0.00000

2.5 Conclusions

From Chapter 2 the primary conclusions centred upon the results for phylogenetic analysis and population-level genetics.

Taylor and Ó Foighil (2000) conducted the first gene sequencing for any New Zealand *Lasaea*, represented with the *L. hinemoa* that were sequenced in the course of this study (NZL1 and NZL3) (Figure 2.4). Whilst they did not positively identify which New Zealand *Lasaea* species was represented, the close clustering with subsequent *L. hinemoa* that were sequenced in the course of this study gives strong support for them belonging to the same species. From the phylogenetics analysis it was determined that the two distinct clades characterised by Taylor and Ó Foighil (2000) from Wellington were found throughout much of New Zealand (Clades I and III). Antipodes Island individuals separated into two clades with one group being more closely related to the Kerguelen Islands *Lasaea* than to the other Antipodes Island *Lasaea* group, indicating that Antipodes Island individuals are not monophyletic. In addition to these distinct clades being found throughout New Zealand, two distinct individuals (Clade IV) were identified using COIII that clustered quite far away from all other individuals sequenced in the course of this project. The nuclear gene, ITS2 broadly supported the clades seen with the mitochondrial gene, COIII. The posterior probabilities and bootstrap values gave strong support in both trees for the genetic divergence seen.

Population genetics utilized TCS networks, haplotype maps, AMOVA, DAPC and standard indices of genetic diversity. These population-based genetics methods supported a North-South split in New Zealand for *L. hinemoa*.

2.6 Limitations

Limitations were present in the course of this study and are thus recognised and discussed.

Sampling limitations were that *L. hinemoa* could not be found in many locations (see Appendix A for further details). This shortfall resulted in a reduced number that could be compared and meant that due to the lack of *L. hinemoa* found on the west coast north of Wellington no conclusions could be drawn about *L. hinemoa* in this region. Dr Bruce Marshall from Te Papa museum was contacted and confirmed that the presence of *L. hinemoa* in that region had been previously found to be low. To counteract the lack of *L. hinemoa* found in many places, large collections were taken from places where they could be sampled, and a large geographic range was examined.

DNA extractions were often poor quality. Molluscan DNA extractions are known to be difficult due to the mucopolysaccharides and polyphenolic proteins they secrete, which copurify with DNA and interfere with enzymatic processing of nucleic acids (Winnepenninckx, Backelijau, & De Wachter, 1993). They were considered to be low quality due to the low amount of genomic DNA quantified using the nanodrop.

It should also be acknowledged that misidentification or mix ups are possible and Clade IV could be a result of this. The confidence we can have in this result is built on the premise that with both markers are amplifying and successfully producing the same result it is accurate. However, mistakes can happen and without more sampling Clade IV cannot definitively represent true divergence.

As reviewed by Schrader, Schielke, Ellerbroek, and Johne (2012), PCR inhibitors are present in a diverse range of taxa, including molluscs. Polysaccarides seem to be primarily responsible for inhibition of PCR; additionally, in bivalves, the glycogen content of the tissues influences the efficiency of PCR. In order to increase the chances of amplification, Bovine Serum Albumin (BSA) (Al-Soud & Rådström, 2000) was added to COIII PCRs and a low annealing temperature of 40°C was used (Ó Foighil & Smith, 1995). When sequences were returned, they were checked against BLAST (the online database) to confirm their species identity. ITS2 was subject to more contamination issues than COIII, as high rates of negative amplification occurred, gradients were run, and the annealing temperature increased. New primers were then used to remedy this problem.

2.7 References

Al-Soud, W. A., & Rådström, P. (2000). Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *Journal of Clinical Microbiology*, *38*(12), 4463-4470.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Batham, E. (1956). Ecology of southern New Zealand sheltered rocky shore. *Transactions of the Royal Society of New Zealand*, *84*(2), 447-465.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *76*(4), 1967-1971.

Casquet, J., Thebaud, C., & Gillespie, R. G. (2012). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources*, *12*(1), 136-141.

Clement, M., Posada, D. C. K. A., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*(10), 1657-1659.

Doyle, J. J., & Dickson, E. E. (1987). Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, *36*(4), 715-722.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

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), 564-567.

Faure, B., Jollivet, D., Tanguy, A., Bonhomme, F., & Bierne, N. (2009). Speciation in the deep sea: Multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PLoS One*, *4*(8), e6485.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307-321.

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772-780.

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., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547-1549.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695-1701.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772-773.

Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110-1116.

Li, J., Ó Foighil, D., & Park, J. K. (2013). Triton's trident: cryptic Neogene divergences in a marine clam (*Lasaea australis*) correspond to Australia's three temperate biogeographic provinces. *Molecular Ecology*, *22*(7), 1933-1946.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, 1-8.

Ó Foighil, D., & Smith, M. J. (1995). Evolution of asexuality in the cosmopolitan marine clam *Lasaea*. *Evolution*, *49*(1), 140-150.

Ó Foighil, D., & Smith, M. J. (1996). Phylogeography of an asexual marine clam complex, *Lasaea*, in the Northeastern Pacific based on Cytochrome Oxidase III sequence variation. *Molecular Phylogenetics and Evolution*, *6*(1), 134-142.

Ó Foighil, D., & Jozefowicz, C. J. (1999). Amphi-Atlantic phylogeography of direct-developing lineages of *Lasaea*, a genus of brooding bivalves. *Marine Biology*, *135*(1), 115-122.

Paradis, E., Claude, J., & Strimmer, K. (2004). *APE*: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289-290.

Ponder, W. F. (1971). Some New Zealand and subantarctic bivalves of the Cyamiacea and Leptonacea with descriptions of new taxa. *Records of the Dominium Museum*, *7*(13), 119-141.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901-904.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, Liang., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539-542.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors–occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014-1026.

Sequencher, 2011. Tutorial for Windows and Macintosh, Quality Scores. Ann Arbor, MI: Gene Codes Corporation, 1–6.

Sponer, R., & Roy, M. S. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*, *56*(10), 1954-1967.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688-2690.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512-526.

Taylor, D. J., & Ó Foighil, D. (2000). Transglobal comparisons of nuclear and mitochondrial genetic structure in a marine polyploid clam (*Lasaea*, Lasaeidae). *Heredity*, *84*(3), 321-330.

Vawter, L., & Brown, W. M. (1986). Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science*, *234*(4773), 194-196.

Wade, C. M., Mordan, P. B., & Naggs, F. (2006). Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnaean Society*, *87*(4), 593-610.

Winnepenninckx, B., Backelijau, T., & De Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends in Genetics*, *9*(12), 407.

Xu, Z., Guo, X., Gaffney, P. M., & Pierce, J. C. (2001). Chromosomal location of the major ribosomal RNA genes in *Crassostrea virginica* and *Crassostrea gigas*. *The Veliger*, *44*(1), 79-83.

3. Population Genetics of *Lasaea hinemoa* utilizing Microsatellite Analysis

3.1 Abstract

Microsatellite markers provide a framework for answering fine-scale ecological questions and have been utilized to investigate the population dynamics of *Lasaea hinemoa*. Five microsatellite markers were developed and used to analyse three populations of *L. hinemoa* in Dunedin. *L. hinemoa* divides into two main groups for mitochondrial gene sequencing (COIII) as well as for microsatellite markers, indicating support for two species being present. Support was weaker for divisions based upon locality.

3.2 Introduction

Microsatellites are tandem repeats also known as simple sequence repeats (SSR), variable number tandem repeats (VNTR) and short tandem repeats (STR) that consist of repeats of 1–6 nucleotides found at high frequency in the nuclear genomes of most species (Tautz & Renz, 1984; Selkoe & Toonen, 2006). A marked number of microsatellites have high mutation rates (ranging between 10^{-2} and 10^{-6} mutations per locus per generation), that produce the high levels of genetic diversity required for genetic studies of processes acting on ecological time scales (Schlötterer, 2000). Because they are highly polymorphic, microsatellites have been used in a wide range of applications including: forensics, population genetics, genetic mapping and evolutionary studies (Chakraborty, Kimmel, Stivers, Davison, & Deka, 1997; Selkoe & Toonen, 2006; Ballantyne et al., 2010). Microsatellites are particularly interesting for ecologists because, unusually among genetic markers, they allow researchers to discern answers to fine-scale ecological queries (Selkoe & Toonen, 2006). The DNA immediately neighbouring a microsatellite locus is termed the flanking region.

Microsatellite loci are often identified by these flanking regions because they are generally conserved across individuals of the same species, and occasionally across different species. Primers can be designed to bind to the flanking region and guide the amplification of a microsatellite locus with polymerase chain reaction (PCR).

Microsatellite markers have not previously been employed to study *Lasaea* species. Other genetic markers have been used to examine populations, including allozyme data (Taylor & Ó Foighil, 2000), RAPD markers (Tyler-Walters & Hawkins, 1995) and gene sequencing (Ó Foighil & Jozefowicz, 1999).

Allozyme data has shown that in the five populations incorporated into a transglobal study that there was marked deviation from expected random mating patterns (within and among loci), frequent fixed heterozygosity and reduced genotypic diversity (Taylor & Ó Foighil, 2000). This structure and that there were multiple asymmetric allozymic heterozygotes, indicated a clonal structure consistent with allopolyploid origins for each respective population. Spatial analysis of allozyme markers also indicated strong geographic structuring and no cosmopolitan clonal lineages. Tyler-Walters and Hawkins (1995) asserted that the small size of *Lasaea* has hampered the use of allozyme electrophoresis in the past, such that only a few loci can be scored per individual.

RAPD markers (Random Amplified Polymorphic DNA) have been developed for *Lasaea rubra* (a polyploid, clonal lineage), and have the potential for more widespread utility (Tyler-Walters & Hawkins, 1995). However, RAPD markers also have limitations compared to microsatellites as they do not provide as many fine-scale insights (Sunnucks, 2000; Selkoe & Toonen, 2006).

Gene sequencing has incorporated both mitochondrial and nuclear genes for *Lasaea* (Taylor & Ó Foighil, 2000; Li, Ó Foighil & Park, 2013), and has yielded insight into the genetics of global *Lasaea* populations. However, microsatellites have higher rates of polymorphisms and can give more insight into fine-scale population differences (Schlötterer, 2000).

L. hinemoa has been studied before through mitochondrial gene sequencing and allozyme data analysis (Taylor & Ó Foighil, 2000). This prior information allows us some rudimentary insights of *L. hinemoa* based on one location from Wellington (Lyall Bay) (Taylor & Ó Foighil, 2000). Whilst the results garnered have been promising they have also had limitations; allozyme data has been found to be less reliable compared to microsatellite data and could fail to answer more fine scale ecological questions (Selkoe & Toonen, 2006; Djan, Popović, Veličković, Obreht, & Vapa, 2014). Also the conclusions drawn from Taylor and Ó Foighil (2000) feature four individuals for mitochondrial analysis and nine individuals for allozyme data, so incorporating more individuals from a wider range of sites could provide significantly better understanding of population dynamics of *L. hinemoa*.

As microsatellite markers have not previously been developed for the *Lasaea* genus they had to be developed. There are several different ways to develop microsatellite markers, but the broad steps are the same (Figure 3.1); the acquisition of knowledge pertaining to nucleotide sequences in which SSRs occur, the design of primers complementary to the regions flanking the SSR, validation of primers that had been constructed via PCR and electrophoresis, and finally detection of polymorphisms amongst individuals (Mason, 2015).

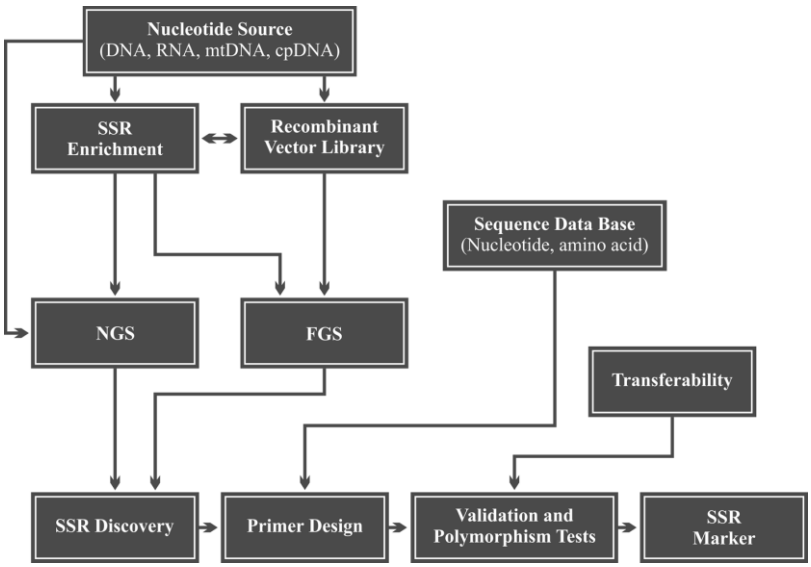


Figure 3.1: Figure taken from Vieira, Santini, Diniz, & Munhoz (2016). Schematic depicting the process of developing microsatellite markers (also known as Single sequence repeats (SSR's).

3.3 Methods

3.3a Sampling

Sampling conducted for *L. hinemoa* for microsatellite analysis followed the same system for collection as Chapter 2 (Section 2.3a), although sites were limited to three Dunedin locations; Warrington (n=4), Portobello (n=13) and Weller's Rock (n=6). Numbers of *L. hinemoa* were not evenly distributed among sites, and numbers collected were approximate (Figure 3.2). Full description of sampling can be found in Appendix A.

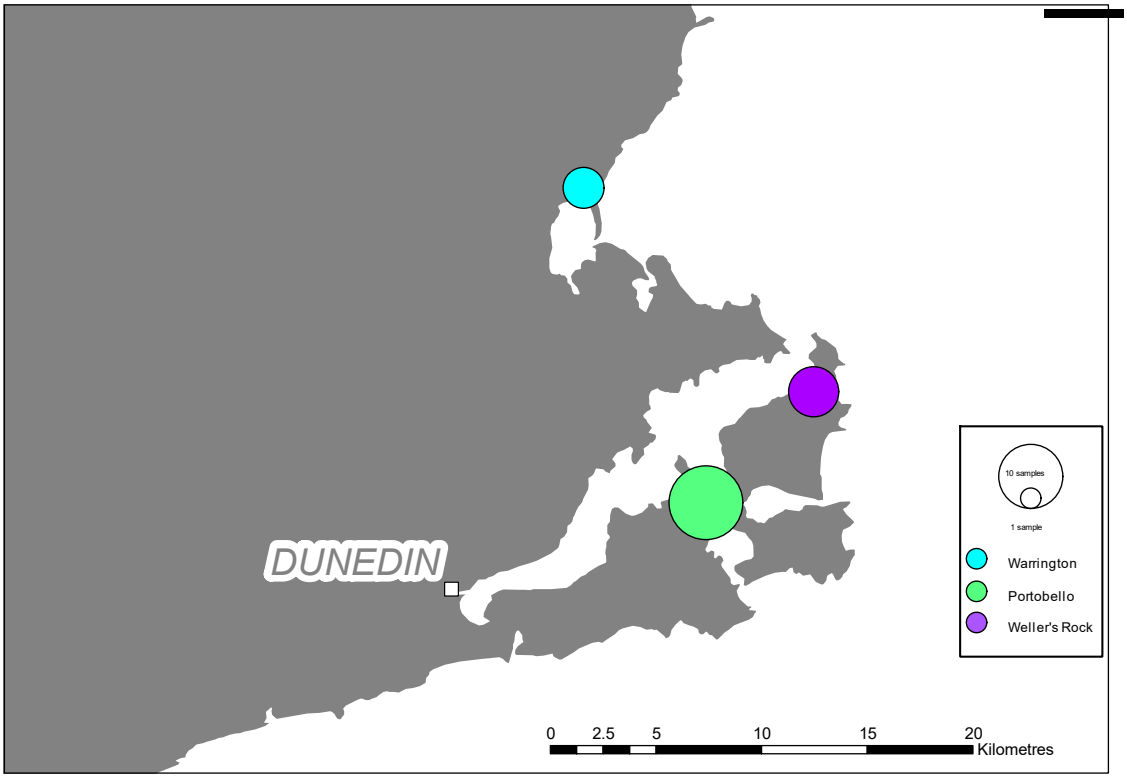


Figure 3.2: Map of locations in Dunedin where *L. hinemoa* were collected from: Warrington (45°42'31.2"S 170°35'17.0"E) (n=4), Portobello (45°50'21.4"S 170°39'05.8"E) (n=13) and Weller's Rock (45°47'51.6"S 170°42'55.5"E) (n=5).

3.3b Development of Microsatellite Markers

Initially the genome of *L. hinemoa* collected from Weller's Rock, Dunedin was sequenced via Next Generation Sequencing (NGS), in order for microsatellites to be identified and appropriate primers determined (Ekblom & Galindo, 2011; Gardner, Fitch, Bertozzi, & Lowe, 2011). An Illumina HiSeq 2500 sequencing system was utilized and produced 26.4 million paired-end reads. The sequencing length was 2 x 125bp. In brief, following a quality control step using FastQC (Andrews, 2010), reads were subsampled to 10% and then assembled using SPAdes v.3.11.1 (Bankevich et al., 2012). MSATCOMMANDER v.10.8.beta (Faircloth, 2008) was then used to identify microsatellites and extract primers. Microsatellites sequences or primers overlapping in sequence were ignored and 28 microsatellites were selected aiming for high repeat counts and a variety of PCR product size (i.e. nine tetranucleotides, nine trinucleotides and ten dinucleotides repeats) (Appendix D).

In order to analyse the respective lengths of the PCR products by electrophoresis and a laser detection system, one of these primers had to carry a fluorescent dye label which could be 6-carboxy-fluorescine (FAM), hexachloro-6-carboxy-fluorescine (HEX), 6-carboxy-X-rhodamine (ROX), or tetrachloro-6-carboxy-fluorescine (TET) (Schuelke, 2000). In order to fluorescently label the dye of PCR products in one reaction, three primers were constructed (a sequence-specific forward primer with M13(-21) tail at its 5' end (5'-TGT AAA ACG ACG GCC AGT-3'), a sequence-specific reverse primer, and the universal fluorescent- labelled M13(-21) primer) (Figure 3.3). The thermocycling conditions are chosen to ensure that during the first cycles, the forward primer with its M13(-21) sequence is incorporated into the accumulating PCR products. Later, when the forward primer is used up, the annealing temperature is lowered to facilitate annealing of the universal M13(-21) primer. Thus, the universal fluorescent-labelled M13(-21) primer "takes over" as the forward primer and incorporates the fluorescent dye into the PCR product.

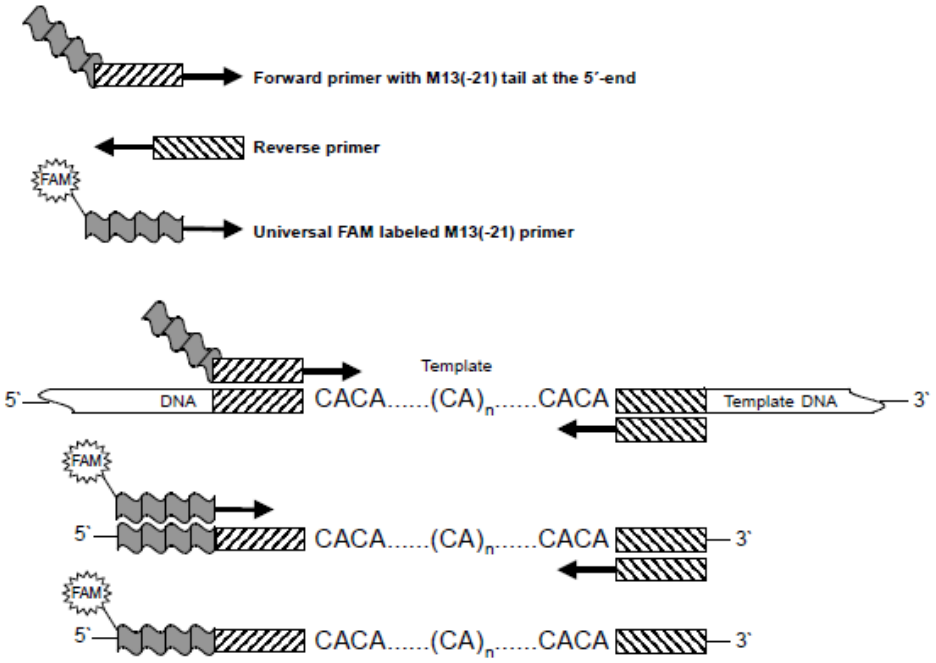


Figure 3.3: Figure taken from Schuelke (2000). Amplification scheme for the one-tube, single-reaction nested PCR method. The hatched boxes indicate the microsatellite-specific primers, the undulating grey box the universal M13(-21) sequence, and the star the fluorescent FAM label. In the first PCR cycles, the forward primer with the M13(-21) tail is incorporated into the PCR products. These products are then the target for the FAM-labelled universal M13(-21) primer, which is incorporated during subsequent cycles at a lower annealing temperature of 53°C. The final labelled product can be analysed on a laser detection system.

Primers (28 pairs) were ordered through Sigma-Aldrich Incorporated (https://www.sigmaaldrich.com/) and tested with high quality DNA to identify the microsatellites that would be used to characterise the populations of *L. hinemoa* available. Primers were tested across different samples to ascertain whether they would amplify across different populations. Five primer sets were found to amplify well enough across different populations to be used to characterise different populations (Table 3.1).

Table 3.1: Five microsatellite primers that successfully amplified and were used for the purposes of characterising *L. hinemoa* populations.

Locus	Primer Sequences	Repeat Motif	Product Length (bp)
1	F- AGCCGTTGTTGTGACTCTTC R- GAAGCAAAGCAAATATCAGCCC	AAC(11)	146
2	F- TTTAAAGAGCGGAGGGTATTCC R- GAACTCACGAACTCTGGCTTC	ATC(9)	437
3	F- AAACACTGGTATGAGGACAGC R- TGTTGGTATGTGTTCGATCGTG	AC(24)	364
4	F- AACCCTAGCCTAACCGTTTG R- ATGCGTGTAAATCCTGTGCG	AAAC(14)	409
5	F- TCGAAGATACCCATGCACAC R- GAAAGGATGTTGCGTGTTTGC	ACGC(9)	163

3.3c DNA Extraction and PCR Conditions

DNA extraction was the same for samples that were used in COIII gene sequencing or microsatellite analysis and followed the same methodology as Chapter 2 (Section 2.3b). COIII sequencing was utilized for all individuals that were analysed via microsatellite analysis as, per Chapter 2, it has already been determined that *L. hinemoa* falls into two distinct clades. We want to test the hypothesis that there is no cross-clade genetic admixture. By sequencing all individuals undergoing microsatellite analysis for COIII, we can carry out this test.

The PCR protocol for COIII was the same as in Chapter 2 (Section 2.3b).

For the purposes of microsatellite PCR set up DNA was dried into wells and PCRs were constructed (2 µl per well) consisting of a M13-tagged forward primer, reverse primer, dye, 2xType-it mix and H_2O. The thermocycler conditions consisted of

denaturation at 95°C for 5 min, followed by 8 cycles of 94°C for 30 s, 60°C annealing temperature, 72°C for 45 s for amplification, and finally a prolonged period of 25 cycles consisting of 94°C for 30 s and 52°C for 1 min and 30 s. Following the thermocycler, the PCR product was diluted, and the product was then pooled and was subsequently sent to Genetics Analysis Services, Anatomy Department, University of Otago. ABI sequencing (ABI 3730, Applied Biosystems) was performed by Genetic Analysis Services at Otago University (Dunedin, New Zealand).

It was attempted to have 20 individuals per population to be analysed for microsatellites and COIII in order to be able to draw conclusions between populations with reasonable support. However, poor amplification led to much smaller populations being analysed so only rudimentary conclusions could be drawn. This high failure rate is not completely surprising, as noted by Hedgecock et al. (2004), some species experience more amplification problems than others, notably bivalves.

3.3d Data Processing

Prior to analysis COIII sequences underwent the same process as described in Chapter 2 (Section 2.3c). GeneMapper v.4.0 (Chatterji & Pachter, 2006) was utilized to identify and characterise amplified microsatellite markers. Peaks were identified and recorded in a spreadsheet (Appendix E). Poor amplification led to a high degree of missing data.

3.3e Phylogenetic Analysis

The phylogenetic relationship between Dunedin *L. hinemoa* populations was examined by creating a single gene tree (COIII) using Bayesian and maximum likelihood methods. The analysis followed the same process as described in Chapter 2 (Section 2.3d), although the outgroups consisted only of *Kellia laperousi* (Ó Foighil & Smith, 1995), NZL1 and NZL3 (Taylor & Ó Foighil, 2000) (Table 2.1).

3.3f Formatting and Data Manipulation

MICRO-CHECKER v.2.2.3 (Van Oosterhout et al., 2004) was used to detect null alleles (when one or more alleles fail to amplify during PCR), stuttering (slight changes to allele frequencies during PCR), large allelic dropout (large alleles do not amplify as efficiently as small alleles), and base-call errors. The application uses a Monte Carlo simulation (bootstrap) method to generate expected homozygote and heterozygote allele size difference frequencies. The Hardy-Weinberg theory of equilibrium is used to calculate expected allele frequencies and the frequency of any null alleles detected.

For the purposes of utility all the samples were initially analysed together as the Warrington and Weller's Rock populations would have been too small to investigate otherwise.

3.3g Population Genetic Analyses

STRUCTURE v.2.3.4 (Pritchard, Wen, & Falush, 2003) is a program that infers population structure by implementing a model-based clustering method. The method was introduced by Pritchard, Stephens and Donnelly (2000). The model assumes that there are K populations, characterised by a set of allele frequencies at each locus. Individuals in the sample are assigned to populations, assuming that within populations the loci are at Hardy-Weinberg equilibrium, and linkage equilibrium. STRUCTURE was run with K=2, with parameters consisting of the length of Burnin Period being 5000, and the number of MCMC Reps after Burnin being 50000. The two clades previously identified by phylogenetic analysis (Figure 3.4) were assigned to separate populations to identify the structure within the microsatellite markers.

GeneAlEx v.6.5 (Smouse & Peakall, 2012) (Genetic Analysis in Excel) is a cross-platform package for running population genetic analysis. It allows analysis of codominant, haploid and binary genetic data (including microsatellite data). This program allows microsatellite data to be evaluated and provides a method of comparison for the two clades identified by COIII phylogenetics and by the three source populations. Characteristics of the data were determined for both divisions of the

data, including sample size (N), number of alleles (N_A), fixation index (F), observed heterozygosity (H_O), expected heterozygosity (H_E), and p-values.

An AMOVA (Analysis of Molecular Variance) (Excoffier, Smouse, & Quattro, 1992) was constructed in GeneAlEx to investigate whether more clearly defined differences were present when different clades were being compared as populations, or when different localities of origin were being compared to one another. The genetic distance was determined by making a genetic distance calculation for codominant data. From this genetic data a distance matrix was constructed to enable the AMOVA and PCoA (Principal Coordinates Analysis) (Anderson & Willis, 2003) analysis to be carried out. The results of the AMOVA analyses were tested for significance by 9999 (the most available) random permutations.

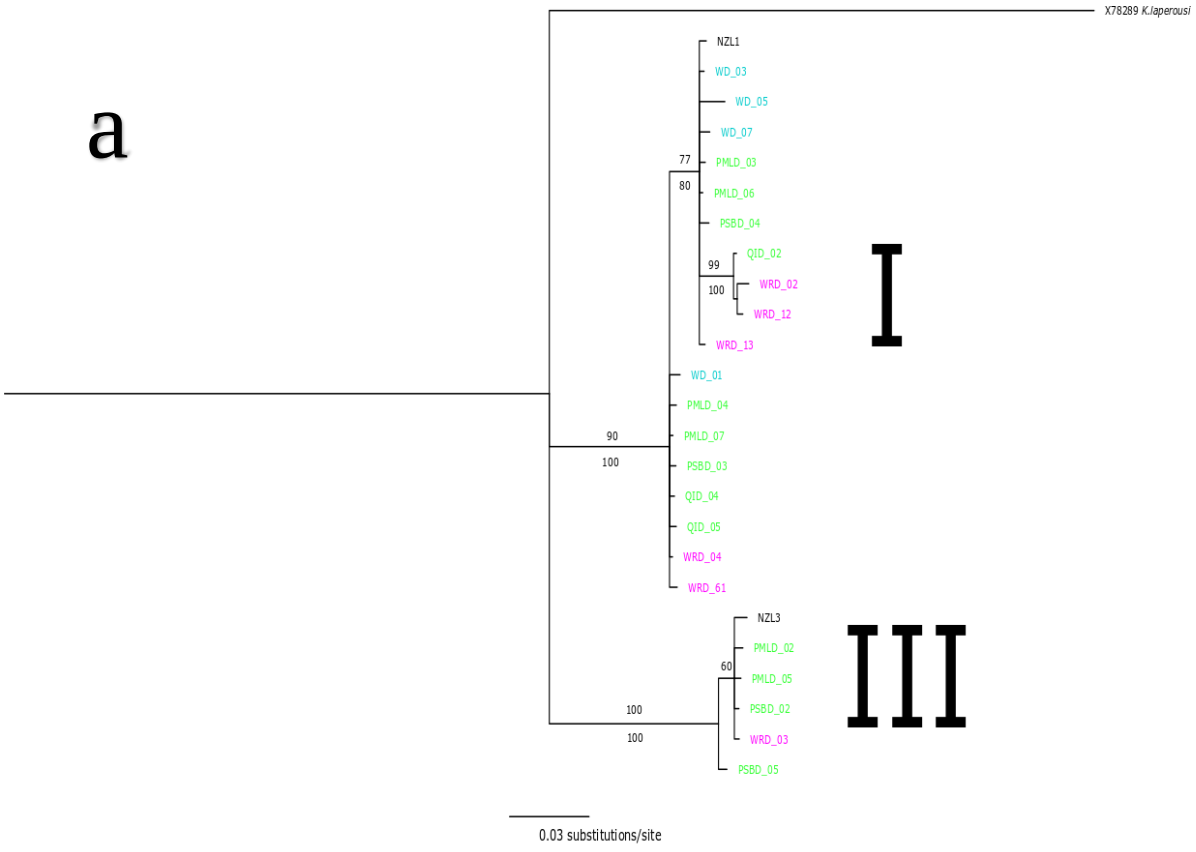
PCoA was calculated in GeneAlEx from the distance matrix previously calculated. PCoA is a method used to explore similarities/dissimilarities in data (Anderson & Willis, 2003). A distance matrix is required and each item is assigned a location in a low-dimensional space. PCoA works to find the main axes through a matrix. It calculates a series of eigenvalues and eigenvectors. Each eigenvalue has an eigenvector, and there are as many eigenvectors and eigenvalues as there are rows in the initial matrix. Eigenvalues are usually ranked from the greatest to the least. Using the eigenvectors, the main axes can be visualised through the initial distance matrix. Clusters in the data are then able to be visualised.

3.4 Results

3.4a Phylogenetic Analysis

To assess *L. hinemoa* from three populations around Dunedin (Warrington, Portobello and Weller's Rock), the individuals that were assessed for microsatellite analysis also had to be assessed via single-gene sequencing of the mitochondrial gene COIII. In both phylogenetic analyses conducted, Bayesian and maximum likelihood methods yielded

very similar topologies for the single-gene dataset, and so only the Bayesian 50% majority-rule inference tree is shown here (Figure 3.4) (maximum likelihood trees are shown in Appendix B). Two clades could easily be identified (Clade I and Clade III) from the three populations.



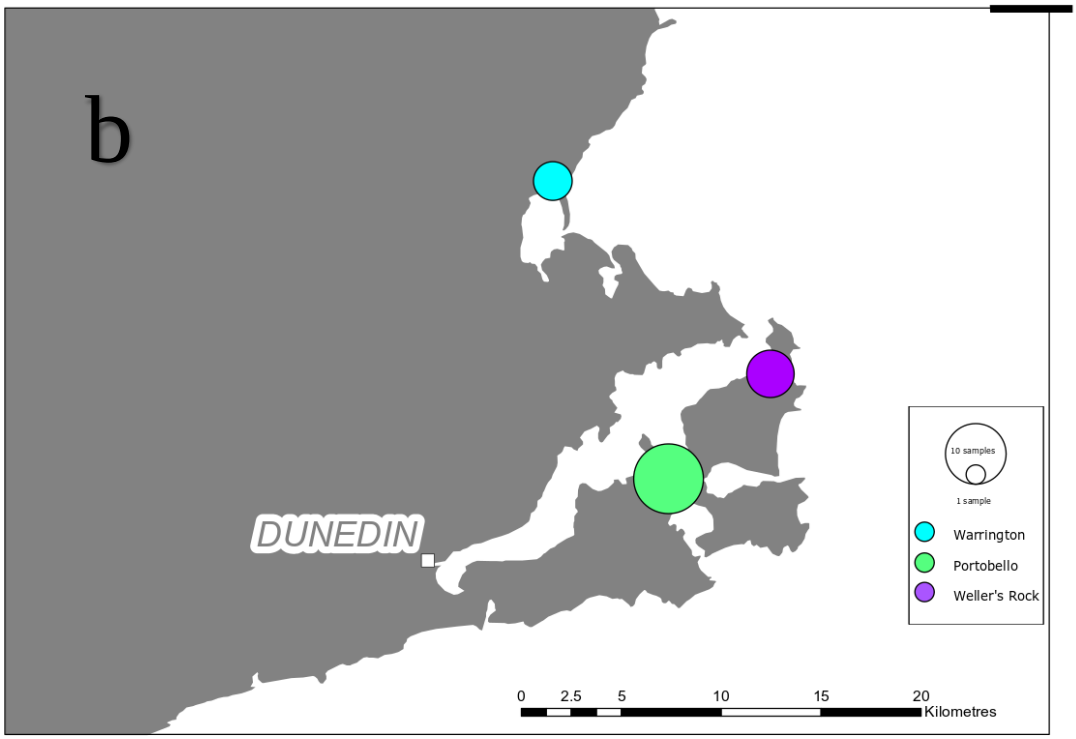


Figure 3.4: a) Bayesian tree for COIII sequences, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst Bayesian posterior probabilities (values below 75% are not shown) are indicated below nodes. Black samples indicate outgroups. b) Locations for where the individuals that comprise the phylogenetic tree were sourced from are displayed (Light blue = Warrington, Green = Portobello, and Purple = Weller's Rock).

3.4b Formatting and Data Manipulation

MICRO-CHECKER showed the frequencies of alleles identified for five loci for the combined data (Table 3.2). MICRO-CHECKER gave indications of high frequencies of null alleles, totalling four of the five loci (Table 3.3). This could indicate that the two clades present represent separate species. Removing individuals from Clade III showed that null alleles were no longer an issue for locus three (Table 3.4). The estimated null allele frequency for each locus is compared to the null allele frequencies obtained using methods by Chakraborty (Chakraborty, Andrade, Daiger, & Budowle, 1992) and Brookfield (Brookfield, 1996).

Table 3.2: The frequency of the alleles identified for the five loci.

Loci	**Size (bp)**	**Observed allele frequency**
Locus 1	167	0.364
	185	0.455
	188	0.182
Locus 2	337	0.531
	453	0.469
Locus 3	379	0.722
	383	0.278
Locus 4	349	0.028
	353	0.083
	397	0.028
	413	0.167

	418	0.083
	422	0.056
	426	0.361
	431	0.139
	453	0.028
	456	0.028
Locus 5	159	0.350
	167	0.150
	170	0.125
	180	0.225
	186	0.100
	237	0.050

Table 3.3: Whole dataset analysed for evidence of null alleles with several different methods of estimation.

Locus	**Null Present**	**Van Oosterhout**	**Chakraborty**	**Brookfield 1**	**Brookfield 2**
1	Yes	0.4408	1	0.3858	0.8183
2	No	-0.3964	-0.2399	-0.2099	0.3712
3	Yes	0.3769	1	0.2863	0.5624
4	Yes	0.1939	0.2322	0.1678	0.4572
5	Yes	0.3819	0.6761	0.3526	0.5049

Table 3.4: Clade I analysed for evidence of null alleles with several different methods of estimation.

Locus	Null Present	Van Oosterhout	Chakraborty	Brookfield 1	Brookfield 2
1	Yes	0.3797	1	0.2899	0.8423
2	No	-0.3548	-0.2247	-0.1925	0.2907
3	No	0	0	0	0.527
4	Yes	0.182	0.2195	0.1636	0.3961
5	Yes	0.349	0.5906	0.313	0.4562

3.4c Population Genetic Analyses

Microsatellite data for individuals that been identified through phylogenetic analysis as Clade I and Clade III (Figure 3.4) were compared through the program GeneAlEx to identify whether clade or locality was a better fit for explaining variation observed. Characteristics of all five loci were evaluated by clade (Table 3.5) and by locality (Table 3.6). The expected heterozygosity (H_E) (the proportion of heterozygosity expected under random mating), and the observed heterozygosity (H_O) (the proportion of *N* samples that are heterozygous at a given locus) differ at most of the loci when evaluated under either locality or by clade. H_O values ranged from 0 to 0.786 by clade, and from 0 to 1 by locality. H_E values ranged from 0 to 0.833 by clade, and from 0 to 0.813 by locality. P-values are significantly different for more of the loci evaluated by clade. This difference indicated that when evaluated by clade the assumptions of HWE are not being met. The Fixation Index (F), also known as the Inbreeding Coefficient exhibit values that range from -1 to 1. Random mating would predict values close to zero, whilst substantial positive values indicate inbreeding or undetected null alleles. However, negative values indicate excess of heterozygosity, due to negative assortative mating, or selection for heterozygotes.

Structure was utilized to identify the underlying genetic structure of the individuals incorporated into the microsatellite analysis (Pritchard et al., 2003). Each bar represents an individual, and the bars are filled by colours that represent the likelihood of membership to each cluster. Five distinct individuals emerged, all from Clade III, showing the distinction between the two clades represented by phylogenetic analysis (Figure 3.5).

An AMOVA (Analysis of Molecular Variance) was run by clade (Table 3.7) and by locality (Table 3.8). The p-value was more significant when the AMOVA was run between clades rather than locality.

Finally, a Principal Coordinates analysis (PCoA) was run for the dataset divided into Clade I and Clade II (Figure 3.6), and for the dataset divided by locality origin (Figure 3.7). More structure for the data was visible when the data was divided by clades, rather than locality.

Table 3.5: Sample size (N), Number of alleles (N_A), Fixation Index (F), observed heterozygosity (H_O), expected heterozygosity (H_E), and p-values observed when data was divided into the two clades identified by phylogenetic analysis (Clade I and Clade III).

	Loci	**N**	**N_A**	**F**	**H_O**	**H_E**	**P-value**
Clade I							
	Locus 1	7	2	1	0	0.408	0.008 **
	Locus 2	14	2	-0.579	0.786	0.497	0.030 *
	Locus 3	13	1	N/A	0	0	N/A
	Locus 4	15	9	0.360	0.533	0.833	0.072
	Locus 5	16	6	0.743	0.188	0.729	0.0 ***
Clade III							
	Locus 1	4	2	1	0	0.5	0.046 *
	Locus 2	2	2	-1	1	0.5	0.157
	Locus 3	5	1	N/A	0	0	N/A

	Locus 4	3	2	-0.2	0.333	0.278	0.729
	Locus 5	4	1	N/A	0	0	N/A

Table 3.6: Sample size (N), Number of alleles (N_A), Fixation Index (F), observed heterozygosity (H_O), expected heterozygosity (H_E), and p-values observed for data when divided by locality collected from (Warrington, Portobello and Weller's Rock).

	Loci	**N**	**N_A**	**F**	**H_O**	**H_E**	**p-value**
Warrington							
	Locus 1	1	1	N/A	0.000	0.000	N/A
	Locus 2	2	2	-1	1.000	0.500	0.157
	Locus 3	1	1	N/A	0.000	0.000	N/A
	Locus 4	3	4	0	0.667	0.667	0.387
	Locus 5	3	3	0.455	0.333	0.611	0.343
Portobello							
	Locus 1	7	3	1	0.000	0.571	0.003
	Locus 2	11	2	-0.467	0.727	0.496	0.122
	Locus 3	12	2	1	0.000	0.444	0.001
	Locus 4	11	7	0.413	0.364	0.620	0.002
	Locus 5	13	5	0.889	0.077	0.695	0.0
Weller's Rock							
	Locus 1	3	1	N/A	0.000	0.000	N/A
	Locus 2	3	2	-1	1.000	0.500	0.083
	Locus 3	5	2	1	0.000	0.320	0.025
	Locus 4	4	6	0.077	0.750	0.813	0.382
	Locus 5	4	4	0.652	0.250	0.719	0.207

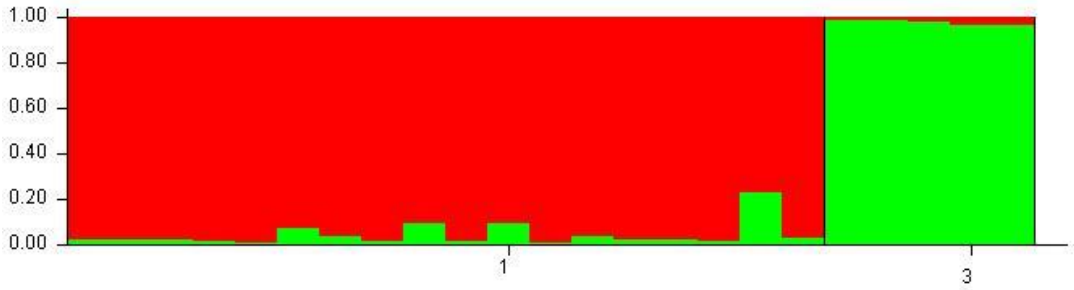


Figure 3.5: Structure plot indicating the distinction between two groups within the data evaluated 1 = Clade I, 3 = Clade III.

Table 3.7: AMOVA carried out between the two clades identified by phylogenetic analysis (Clade I and Clade III).

Source	**d.f.**	**SS**	**MS**	**Est. Var.**	**%**
Among populations	1	26.060	26.060	2.650	33%
Within populations	21	111.722	5.320	5.320	67%
Total	**22**	**137.783**		**7.970**	**100%**

p-value=0.001

Table 3.8: AMOVA carried out between the data for the different localities (Warrington, Portobello and Weller's Rock).

Source	**d.f.**	**SS**	**MS**	**Est. Var.**	**%**
Among populations	2	18.545	9.273	0.494	8%
Within populations	20	119.237	5.962	5.962	92%
Total	**22**	**137.783**		**6.456**	**100%**

p-value = 0.057

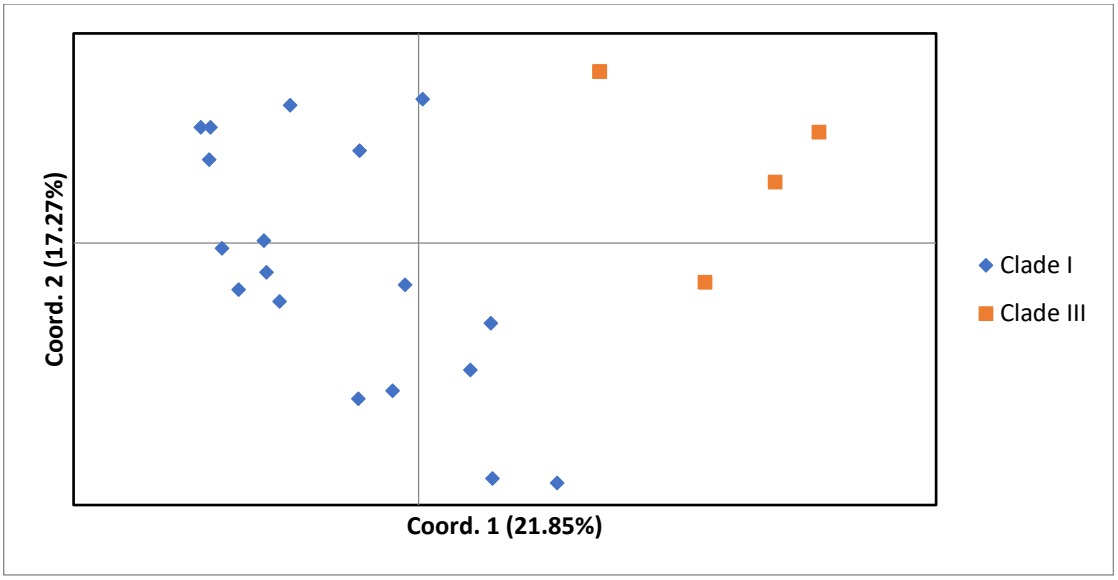


Figure 3.6: Principal Coordinates analysis (PCoA) comparing clades that were identified by phylogenetic analysis (Clade I and Clade III).

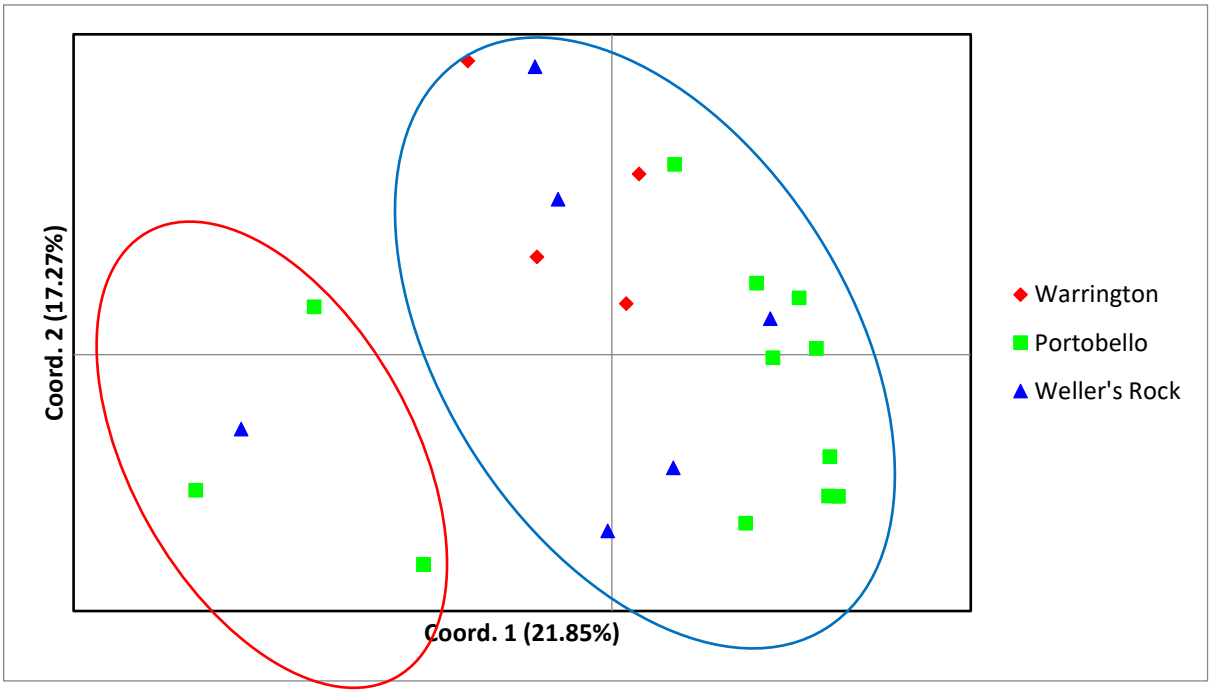


Figure 3.7: Principal components analysis (PCoA) for data compared between different localities (Warrington, Portobello and Weller's Rock). Clade 1 = blue circle, Clade III = red circle.

3.5 Conclusions

Microsatellite analysis supports that the greatest amount of difference comes from whether the individual sequenced falls into Clade I or Clade III, rather than source locality.

However, results must be treated as rudimentary. The data suffered from high rates of missing information due to poor amplification, a not altogether rare occurrence with bivalves (Winnepenninckx, Backelijau, & De Wachter, 1993). High rates of null alleles were present in the data, as has been the case in other bivalves (Rico et al., 2017). Deviation from HWE and heterozygote deficiency has also been recorded in other bivalves. Despite technical issues and possible subsequent under-estimate of diversity that could be present, some results did emerge. The phylogenetic groups Clade I and Clade III served to provide the best fit for the data. Structure supported this division (Figure 3.5). Data analysis that was run through GeneAlEx showed through data characteristics, AMOVA and PCoA that the greatest division in the data can be seen from the clades identified through phylogenetic analysis.

3.6 Limitations

DNA extraction and PCR limitations were the same as described in Chapter 2 (Section 2.6).

New marker isolation is fraught with a high failure for some taxa (in particular marine invertebrates) (Cruz, Perez, & Presa, 2005). As noted by Reece, Ribeiro, Gaffney, Carnegie, and Allen (2004) bivalve molluscs, often exhibit non-Mendelian segregation ratios of alleles, which can confound the creation of a linkage map. In molluscs, and in particular bivalves, microsatellite analysis is often subject to heterozygote deficiencies and departures from Hardy-Weinberg equilibrium (HWE). Inbreeding, genetic patchiness (Walhund effect) and/or null alleles are often held accountable (Johnson & Black, 1984; Lemer, Rochel, & Planes, 2011).

3.7 References

Anderson, M. J., & Willis, T. J. (2003). Canonical analysis of Principal Coordinates: a useful method of constrained ordination for ecology. *Ecology*, *84*(2), 511-525.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M. R., & Kayser, M. (2010). Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *The American Journal of Human Genetics*, *87*(3), 341-353.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., & Pevzner, P.A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455-477.

Brookfield, J. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, *5*(3), 453-455.

Chakraborty, R., Andrade, M. D., Daiger, S. P., & Budowle, B. (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics*, *56*(1), 45-57.

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J., & Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences*, *94*(3), 1041-1046.

Chatterji, S., & Pachter, L. (2006). Reference based annotation with GeneMapper, *Genome Biology, 7*(4), R29.

Cruz, F., Pérez, M., & Presa, P. (2005). Distribution and abundance of microsatellites in the genome of bivalves. *Gene*, *346*(1), 241-247.

Djan, M., Popović, D., Veličković, N., Obreht, D., & Vapa, L. (2014). Comparison of allozyme and microsatellite variability in brown hare populations (*Lepus europaeus Pallas*) from Vojvodina (Serbia). *Biologia Serbica*, *35*(1-2), 24-30.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1-15.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

Faircloth, B. C. (2008). MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, *8*(1), 92-94.

Gardner, M. G., Fitch, A. J., Bertozzi, T., & Lowe, A. J. (2011). Rise of the machines–recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, *11*(6), 1093-1101.

Hedgecock, D., Li, G., Hubert, S., Bucklin, K., & Ribes, V. (2004). Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *Journal of Shellfish Research*, *23*(2), 379-386.

Johnson, M. S., & Black, R. (1984). The Wahlund effect and the geographical scale of variation in the intertidal limpet *Siphonaria sp*. *Marine Biology*, *79*(3), 295-302.

Li, J., Ó Foighil, D., & Park, J. K. (2013). Triton's trident: cryptic Neogene divergences in a marine clam (*Lasaea australis*) correspond to Australia's three temperate biogeographic provinces. *Molecular Ecology*, *22*(7), 1933-1946.

Lemer, S., Rochel, E., & Planes, S. (2011). Correction method for null alleles in species with variable microsatellite flanking regions, a case study of the black-lipped pearl oyster *Pinctada margaritifera*. *Journal of Heredity*, *102*(2), 243-246.

Mason, A. S. (2015). SSR genotyping. *Plant Genotyping.* New York, NY: Humana Press.

Ó Foighil, D., & Smith, M. J. (1995). Evolution of asexuality in the cosmopolitan marine clam *Lasaea*. *Evolution*, *49*(1), 140-150.

Ó Foighil, D., & Jozefowicz, C. J. (1999). Amphi-Atlantic phylogeography of direct-developing lineages of *Lasaea*, a genus of brooding bivalves. *Marine Biology*, *135*(1), 115-122.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Pritchard, J. K., Wen, W., & Falush, D. (2003). Documentation for *structure* software: version 2.

Reece, K. S., Ribeiro, W. L., Gaffney, P. M., Carnegie, R. B., & Allen Jr, S. K. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity*, *95*(4), 346-352.

Rico, C., Cuesta, J. A., Drake, P., Macpherson, E., Bernatchez, L., & Marie, A. D. (2017). Null alleles are ubiquitous at microsatellite loci in the Wedge Clam (Donax trunculus). *PeerJ*, *5*, e3188.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, *109*(2), 365-371.

Schuelke, M. (2000). An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology*, *18*(2), 233-234.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615-629.

Smouse, R. P. P., & Peakall, R. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics, 28*(19), 2537-2539.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution, 15*(5), 199-203.

Tautz, D., & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, *12*(10), 4127-4138.

Taylor, D. J., & Ó Foighil, D. (2000). Transglobal comparisons of nuclear and mitochondrial genetic structure in a marine polyploid clam (*Lasaea*, Lasaeidae). *Heredity, 84*(3), 321-330.

Tyler-Walters, H., & Hawkins, A. R. (1995). The application of RAPD markers to the study of the bivalve mollusc *Lasaea rubra*. *Journal of the Marine Biological Association of the United Kingdom, 75*(3), 563-569.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535-538.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312-328.

Winnepenninckx, B., Backelijau, T., & De Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends in Genetics*, *9*(12), 407.

4. General Discussion

4.1 Clade I and Clade III

4.1a Species-level differences

Molluscs have frequently caused confusion taxonomically, often displaying homoplasy and phenotypic plasticity of shell characteristics (Puillandre, Sysoev, Olivera, Couloux, & Bouchet, 2010). Some taxa turn out to comprise several cryptic species within one morphologically indistinguishable form (Li, Ó Foighil, & Park, 2013). For example, the intertidal limpets *Notoacmea helmsi* was found to consist of five genetically separable species; simultaneously, *N. scapha* consisted of individuals with two distinctly different shell phenotypes (Nakano & Spencer, 2007).

Ponder (1971) identified two primary ectotypes for *L. hinemoa* that were distinguished by colour and concentric sculpture. These ecotypes were environmentally influenced by how much moisture and light they were exposed to. However, intermediate conditions produced intermediate phenotypes. My observations of the clades of the clades detected by genetic analysis showed that both phenotypes were present in each major clade

In the course of this investigation *L. hinemoa* fell into four clades for both COIII and ITS2 data (even though ITS2 data gave less structural information) (Figures 4.1 and 4.2). However, most of the *L. hinemoa* divided into two major clades, which seems to indicate the presence of two cryptic species within the New Zealand *L. hinemoa* species complex. This division into two groups was supported by microsatellite data (Figure 4.3). Clustering methods allowed this division for COIII data and microsatellite data to be shown visually (Figure 4.4). This study was conducted by performing genetic analyses with two genes and with microsatellite markers. By have multiple genetic markers, problems linked to the single-gene approach were reduced (e.g. the presence of pseudogenes (Lorenz, Jackson, Beck, & Hanner, 2005)).

Genetics has become integral to discerning species level relationships (Avise 1989; Goldstein, Desalle, Amato, & Vogler, 2000), and has resulted in increased importance being placed upon concepts that incorporate this (e.g. the Phylogenetic Species Concept) (Donoghue, 1985). As genetic tools have become more common cryptic species have become more readily identifiable (Bickford et al., 2007). Cryptic species have been a taxonomic challenge for centuries, owing to their shared morphology and apparent lack of selective advantage. A lack of morphological variation between two cryptic species could represent morphologically static cladogenesis despite genetic divergence (Lee & Frost, 2002), a not altogether rare occurrence in the marine environment (Payo et al., 2013). The morphological similarity should lead to a lack of co-occurrence as the two species should not be able to able co-exist due to the pressure of competition (Gittenberger, 1991).

It is possible they are accessing resources differently or are not experiencing competition with each other for other reasons. *L. rubra* differentiated into several species of the basis of tidal level preference; although they occurred in very close environment, they did not directly compete (Crisp & Standen, 1988). Resources might also be abundant, allowing them to occupy the same ecological niche. Alternatively, there may be physiological, behavioural or other undetected differences that provide selective differences (Derycke et al., 2016). In Section 4.1b this matter is investigated in more depth.

Species delimitation has often varied by taxonomic group and this is an important point to consider when making claims about the species level relationships for *Lasaea*. Taxonomic studies have often been biased towards vertebrates (Pante, Schoelinck, & Puillandre, 2014). Genetic distance data for COIII shows genetic distinction between Clade I and Clade III (Table 4.1), which could support species-level delimitation. Whilst the genetic data that has been gathered in the course of this project is promising, more data would be needed to warrant a formal classification.

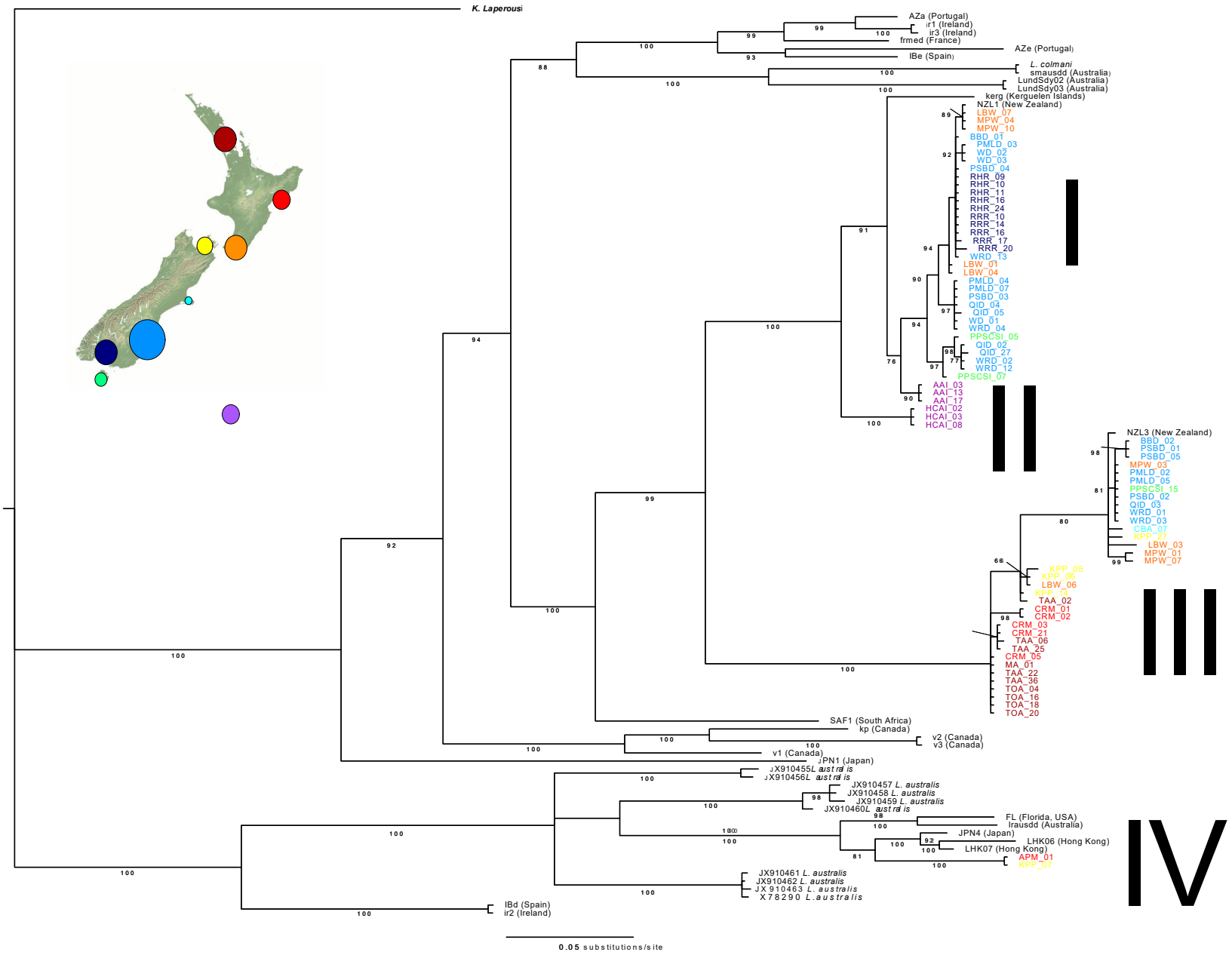

K. Laperousi
AZa (Portugal)
ir1 (Ireland)
ir3 (Ireland)
frmed (France)
AZe (Portugal)
IBe (Spain)
L. colmani
smausdd (Australia)
LundSdy02 (Australia)
LundSdy03 (Australia)
kerg (Kerguelen Islands)
NZL1 (New Zealand)
LBW_07
MPW_04
MPW_10
BBD_01
PMLD_03
WD_02
WD_03
PSBD_04
RHR_09
RHR_10
RHR_11
RHR_16
RHR_24
RRR_10
RRR_14
RRR_16
RRR_17
RRR_20
WRD_13
LBW_01
LBW_04
PMLD_04
PMLD_07
PSBD_03
QID_04
QID_05
WD_01
WRD_04
PPSCSI_05
QID_02
QID_27
WRD_02
WRD_12
PPSCSI_07
AAI_03
AAI_13
AAI_17
HCAI_02
HCAI_03
HCAI_08
NZL3 (New Zealand)
BBD_02
PSBD_01
PSBD_05
MPW_03
PMLD_02
PMLD_05
PPSCSI_15
PSBD_02
QID_03
WRD_01
WRD_03
CBA_07
KPP_27
LBW_03
MPW_01
MPW_07
KPP_05
KPP_06
LBW_06
KPP_13
TAA_02
CRM_01
CRM_02
CRM_03
CRM_21
TAA_06
TAA_25
CRM_05
MA_01
TAA_22
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
SAF1 (South Africa)
kp (Canada)
v2 (Canada)
v3 (Canada)
v1 (Canada)
JPN1 (Japan)
JX910455 L. australis
JX910456 L. australis
JX910457 L. australis
JX910458 L. australis
JX910459 L. australis
JX910460 L. australis
FL (Florida, USA)
Irausdd (Australia)
JPN4 (Japan)
LHK06 (Hong Kong)
LHK07 (Hong Kong)
APM_01
KPP_07
JX910461 L. australis
JX910462 L. australis
JX910463 L. australis
X78290 L. australis
IBd (Spain)
ir2 (Ireland)
0.05 substitutions/site
I
II
III
IV

Figure 4.1: Bayesian tree for COIII sequences. maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst Bayesian posterior probabilities (values below 75% are not shown) are indicated below nodes. Locations for *L. hinemoa* are grouped by main localities (Auckland (TOA = Torbay, Auckland, TAA = Takapuna, Auckland), Mahia (APM = Auroa Point, Mahia, CRM = Coronation Reserve, Mahia), Wellington (MPW = Moa Point, Wellington, LBW = Lyall Bay, Wellington), Picton (KPP = Karaka Point, Picton), Akaroa (CBA = Children's Bay, Akaroa), Dunedin (WD = Warrington, Dunedin, BBD = Broad Bay, Dunedin, PSBD = Portobello South Beach, Dunedin, PMLD = Portobello Marine Lab, Dunedin, QID = Quarantine Island, Dunedin, WRD = Weller's Rock, Dunedin), Riverton (RHR = Riverton Highway, Riverton, RRR = Riverton Rocks, Riverton), Stewart Island (Port Pegasus, Shipbuilder's Cove, Stewart Island) and Antipodes Island(HCAI = Hutt Cove, Antipodes Island, AAI = Anchorage, Antipodes Island), and corresponding colours for locality are indicated on the tree. Explanation for the individual labels can be found in Table 2.2 and Appendix A Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_02-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01-KPP_07).

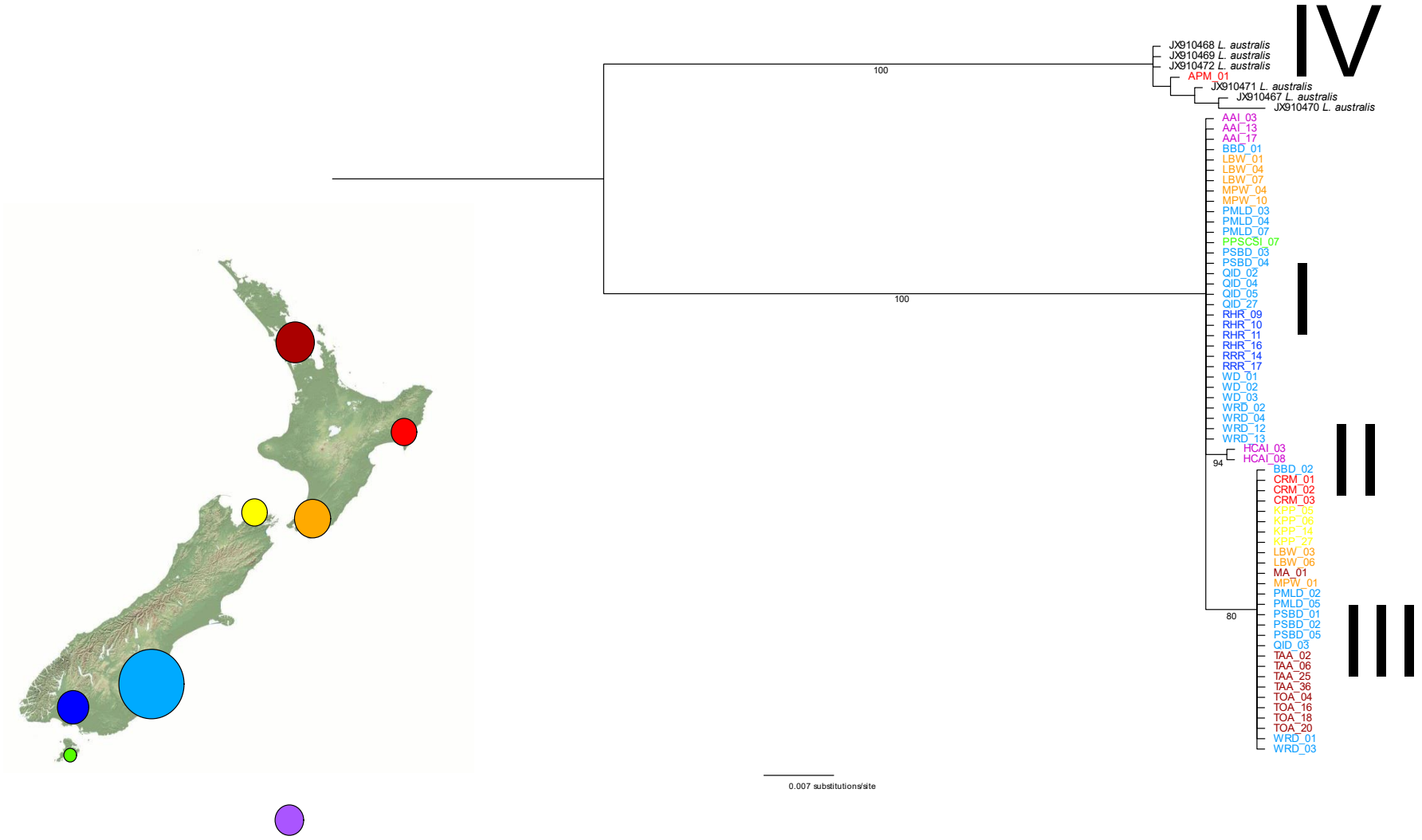

IV
I
II
III
JX910468 L. australis
JX910469 L. australis
JX910472 L. australis
APM_01
JX910471 L. australis
JX910467 L. australis
JX910470 L. australis
AAI_03
AAI_13
AAI_17
BBD_01
LBW_01
LBW_04
LBW_07
MPW_04
MPW_10
PMLD_03
PMLD_04
PMLD_07
PPSCSI_07
PSBD_03
PSBD_04
QID_02
QID_04
QID_05
QID_27
RHR_09
RHR_10
RHR_11
RHR_16
RRR_14
RRR_17
WD_01
WD_02
WD_03
WRD_02
WRD_04
WRD_12
WRD_13
HCAI_03
HCAI_08
BBD_02
CRM_01
CRM_02
CRM_03
KPP_05
KPP_06
KPP_14
KPP_27
LBW_03
LBW_06
MA_01
MPW_01
PMLD_02
PMLD_05
PSBD_01
PSBD_02
PSBD_05
QID_03
TAA_02
TAA_06
TAA_25
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
WRD_01
WRD_03
100
100
94
80
0.007 substitutions/site

Figure 4.2: Bayesian tree for ITS2 sequences, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst posterior probabilities (values below 75% are not shown) are indicated below nodes. Locations for *L. hinemoa* are grouped by main localities (Auckland (TOA = Torbay, Auckland, TAA = Takapuna, Auckland), Mahia (APM = Auroa Point, Mahia, CRM = Coronation Reserve, Mahia), Wellington (MPW = Moa Point, Wellington, LBW = Lyall Bay, Wellington), Picton (KPP = Karaka Point, Picton), Akaroa (CBA = Children's Bay, Akaroa), Dunedin (WD = Warrington, Dunedin, BBD = Broad Bay, Dunedin, PSBD = Portobello South Beach, Dunedin, PMLD = Portobello Marine Lab, Dunedin, QID = Quarantine Island, Dunedin, WRD = Weller's Rock, Dunedin), Riverton (RHR = Riverton Highway, Riverton, RRR = Riverton Rocks, Riverton), Stewart Island (Port Pegasus, Shipbuilder's Cove, Stewart Island) and Antipodes Island(HCAI = Hutt Cove, Antipodes Island, AAI = Anchorage, Antipodes Island), and corresponding colours for locality are indicated on the tree. Explanation for the individual labels can be found in Table 2.2 and Appendix A. Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_03-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01).

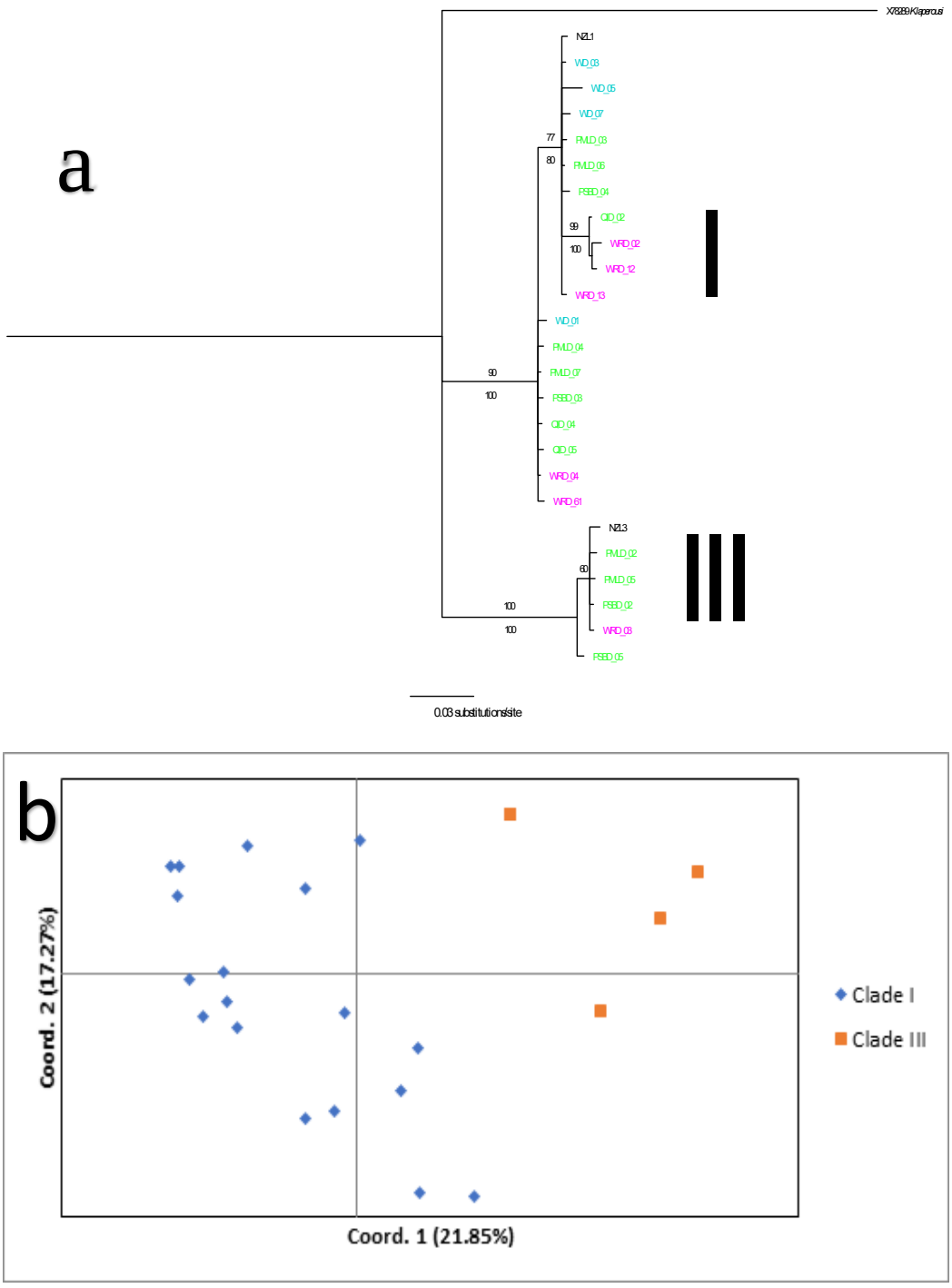


Figure 4.3: a) Bayesian tree for COIII sequences indicating primary clades (Clade I and Clade III) for *L. hinemoa* from three Dunedin locations (Warrington=blue, Portobello=green and Weller's Rock=purple). Maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst Bayesian posterior probabilities (values below 75% are not shown) are indicated below nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A..b) PCoA plot is shown displaying how the individuals sequenced in the phylogenetic tree cluster into 2 groups in microsatellite analysis.

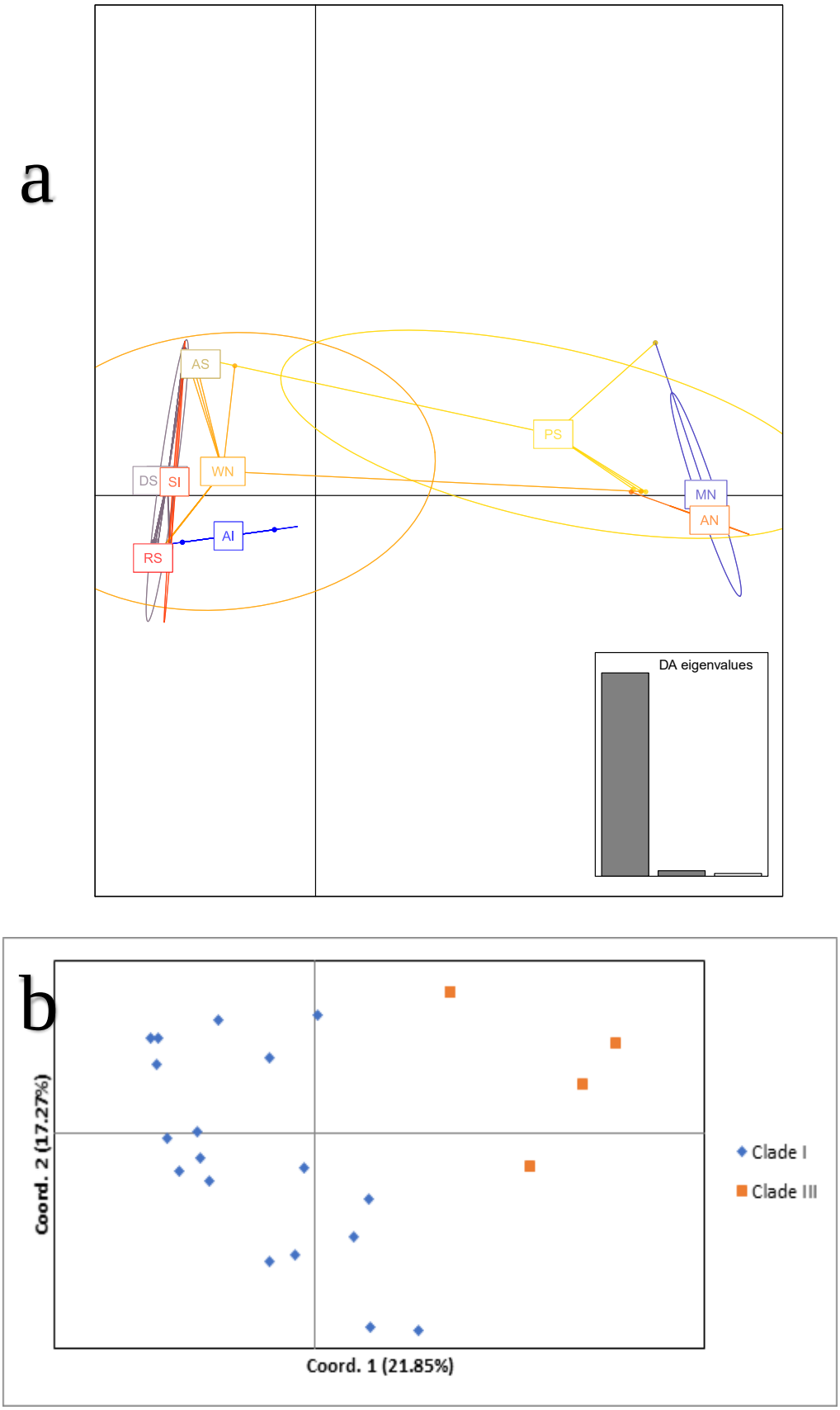


Figure 4.4: Two clustering methods for different sets of data to show the broad support for two main groups of *L. hinemoa*. a) Figure 2.9: Scatterplot of the DAPC analysis, each point represents an individual, and ellipses represent 95% confidence limits for locality-level groups (AN=Auckland, North

Island, MN=Mahia, North Island, WN=Wellington, North Island, PS=Picton, South Island, AS=Akaroa, South Island, DS=Dunedin, South Island, RS=Riverton, South Island, SI=Stewart Island and AI=Antipodes Island). b) PCoA plot is shown displaying how the individuals sequenced in the phylogenetic tree cluster into 2 groups in microsatellite analysis.

Table 4.1: Genetic distance data for the COIII gene measured by between group mean distance using the Tamura-Nei model.

	Clade I	**Clade II**	**Clade III**
Clade I			
Clade II	0.0582		
Clade III	0.1386	0.1319	
Clade IV	0.2977	0.2917	0.2916

4.1b Population-level differences

Within the broader scheme of biogeography another discipline has emerged (Figure 4.5).

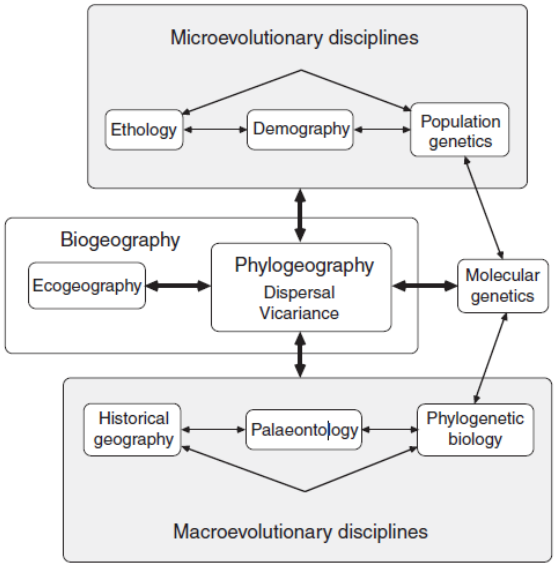
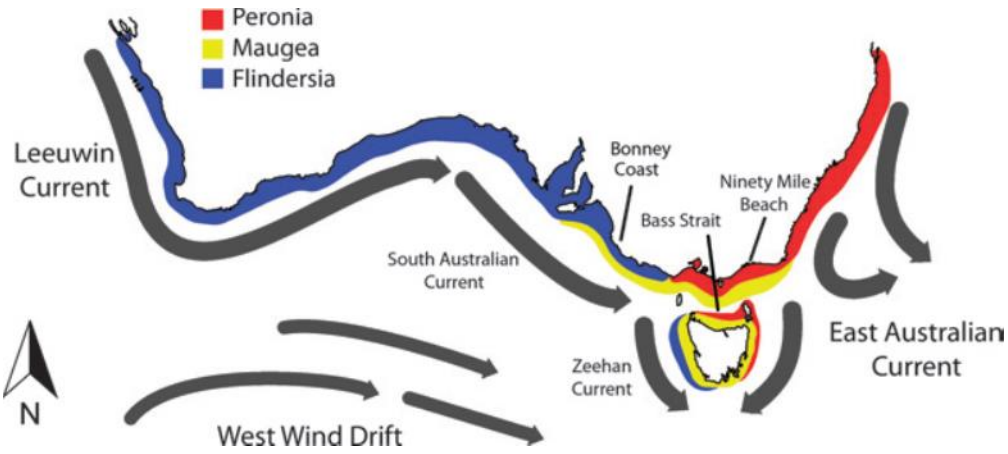


Figure 4.5: Figure taken from Avise (2009). The general place of phylogeography, and some of its empirical and conceptual bridging functions, within the biodiversity sciences.

As reviewed by Avise (2009) the discipline of phylogeography is relatively new and focuses on the spatial arrangements of genetic lineages, particularly within and among closely related species. Phylogeography gives a framework in which to understand the interconnectedness of genealogy and geography. Many phylogeographic splits have arisen as a consequence of geographic barriers, with later genetic or behavioural barriers maintaining the distinctiveness of populations. Phylogeography does appear to be a branch of biogeography with its focus on the spatial and temporal dimensions of genealogy but it remains distinct in important ways. It retains a special focus on conspecific populations and explicit genealogical information that distinguishes it from traditional biogeography; it also remains distinct from ecogeography (Gaston et al., 2008) due to retaining a focus on historical causation in addition to selective forces and ecological processes at work in more contemporary timeframes.

Marine life is no exception to exhibiting phylogeographic trends. Ancient environmental changes have had long term effects that continue to influence the contemporary distribution and genetic composition of species. As reviewed by Schmitt (2007), phylogeographic trends have arisen out of Pleistocene geographic breaks that have commonly contributed to long-term genetic breaks. Ancient climatic conditions have influenced the current phylogeography of species within *Lasaea*, with *L. australis*, for example, exhibiting cryptic genetic structuring with high fidelity to biogeographic province (Li et al., 2013). The geographic separation and temperature differences are thought to have contributed to their genetic divergence. Despite contemporary currents that would allow *L. australis* access to other provinces, they haven't colonised these different areas. The current temperature differences are thought to have contributed to this failure to disperse, as the various genetic forms may have specifically adapted to their own province (Figure 4.6).



4.6: Figure taken from Li et al. (2013). Major current systems of the southern Australian coast.

Theoretically, *L. hinemoa* could have been influenced by the temperature differences seen in New Zealand, but ocean currents will also have affected distribution. Northern areas of New Zealand are much warmer than further south, and this difference could contribute to the distribution seen (Francis, 1996) (Figure 4.7).

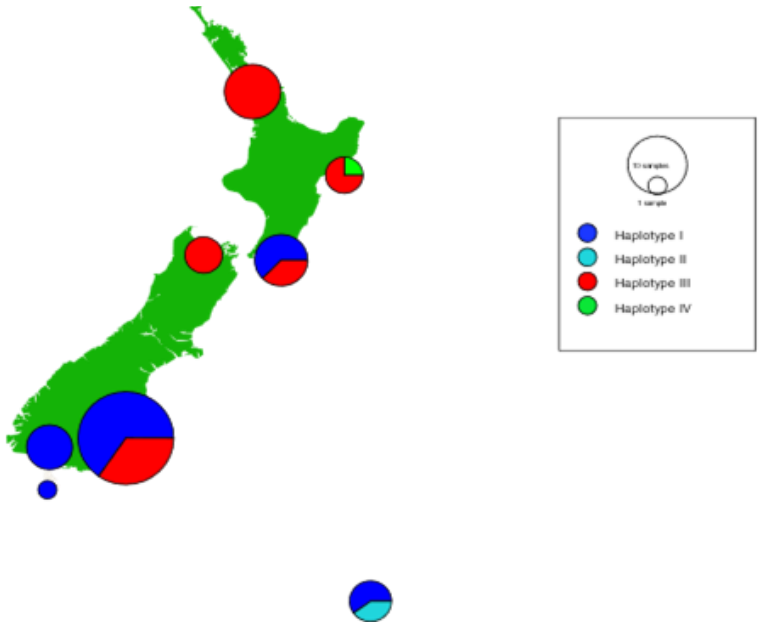


Figure 4.7: ITS2 haplotype map across New Zealand, indicating the separation of four groups.

Long distance dispersal has often been found to play a role in broad distributions (Helmuth, Veit, & Holberton, 1994). Ancient migrations of *Lasaea* have been found to impact the current biogeography of *Lasaea*. The distribution of direct developing Azores *Lasaea* appears to have been informed by ancient tidal currents rather than the

contemporary ones, (and possible anthropogenic introduction) (Ó Foighil, & Jozefowicz, 1999).

The distribution of two cryptic species around New Zealand and as far offshore as the Antipodes Islands may be the result of rafting (Figure 4.8). As discussed in Chapter 1 (Section 1.1c) dispersal often influences the range of marine invertebrates. Direct-developing *Lasaea* are hypothesised to raft in order to gain their current distribution; by contrast, *L. australis* has a restricted range despite possessing pelagic larvae (Ó Foighil, 1989). It is important to note though that *L. australis* is a cross-fertilizing species and would have the added difficulty of maintaining adequate breeding densities. Species with non-pelagic larvae also have better survival as adults protect the larvae (Booth, 1979).

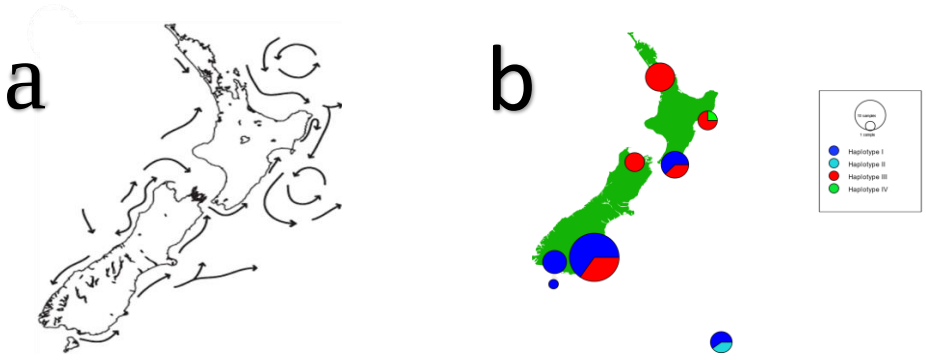


Figure 4.8: a) Figure from Wallis and Trewick (2009) indicating water currents around New Zealand. b) ITS2 haplotype map, indicating separation of the four groups.

Within *L. hinemoa* one species appears to be more dominant in the north, whilst the other appears to be restricted to the south (Figure 4.7). This distribution appears to fit with what is known on New Zealand biogeographic patterns.

New Zealand has been the focus on many biogeographic studies looking to uncover the processes that govern the observed phylogeographic splits. Over the course of several decades, different biogeographic splits have been proposed for New Zealand (Figure 4.9) (Moore, 1949; Knox, 1975; Nelson, 1994; Walls, 1995; Francis, 1996; Apte & Gardner, 2002).

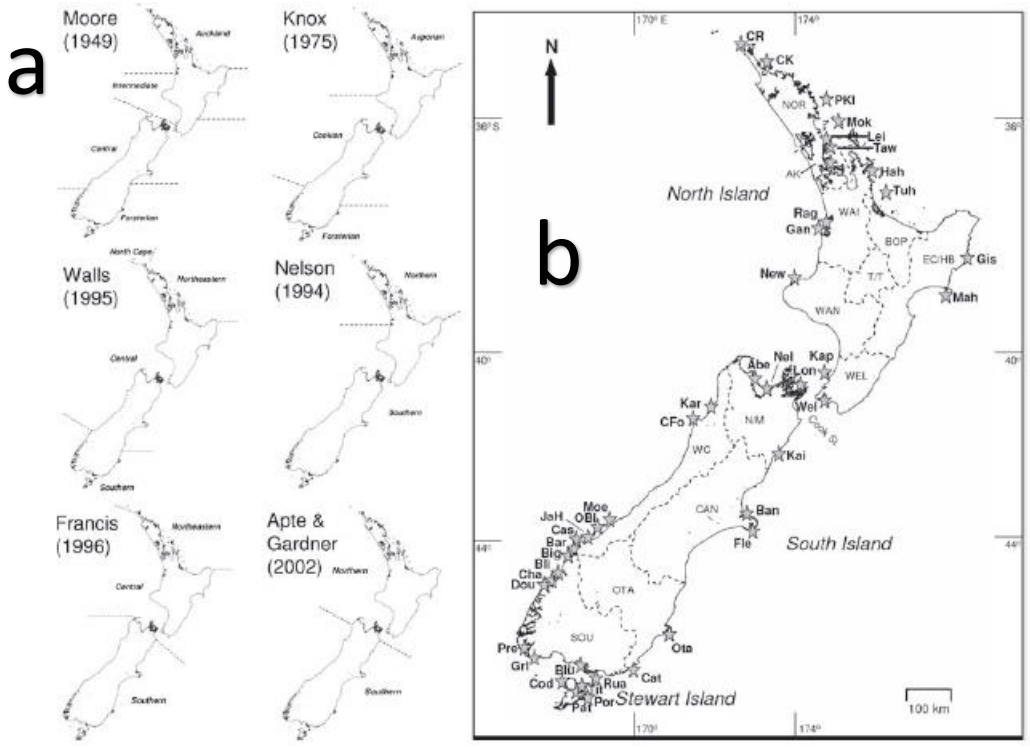


Figure 4.9. Figure taken from Shears, Smith, Babcock, Duffy and Villouta, (2007).
a) New Zealand marine biogeographic classification schemes evaluated in this study (dashed lines, proposed biogeographic boundaries. b) locations where reef communities were sampled around New Zealand in the present study.

However, Shears et al. (2008) found that when these different schemes were evaluated the highest support was garnered for distinct northern and southern groupings. The boundary point differed between differing studies, which emphasised the problems associated with New Zealand's nearshore marine environments. The New Zealand region is bathed by two main water masses: subtropical in the north and subantarctic in the south (Francis 1996). Cook Strait has been found to be the southern limit for many warm temperate fishes, and the northern limit for several cooler temperate ones. However, this zone is crossed by many species and is thus considered an area where northern and southern elements of fish fauna interact and overlap, rather than a strict biogeographic boundary.

In many species in New Zealand there is a biogeographic North-South split centred around Kaikoura and Banks Peninsula (Figure 4.10a) (Ayers & Waters, 2002; Sponer & Roy, 2002). This split has been found for reef fish (Francis, 1996), echinoderms

(Ayers & Waters, 2002; Sponer & Roy, 2002) and molluscs (Apte, & Gardner, 2002). This genetic break coincides with the divergence of the Southland Current offshore towards the Chatham Islands, and seems to represent a major marine barrier on the eastern coast of New Zealand (Sponer & Roy, 2002). In *L. hinemoa* biogeographic differences have been observed between the north and south, fitting with this biogeographic split (Figure 4.10b).

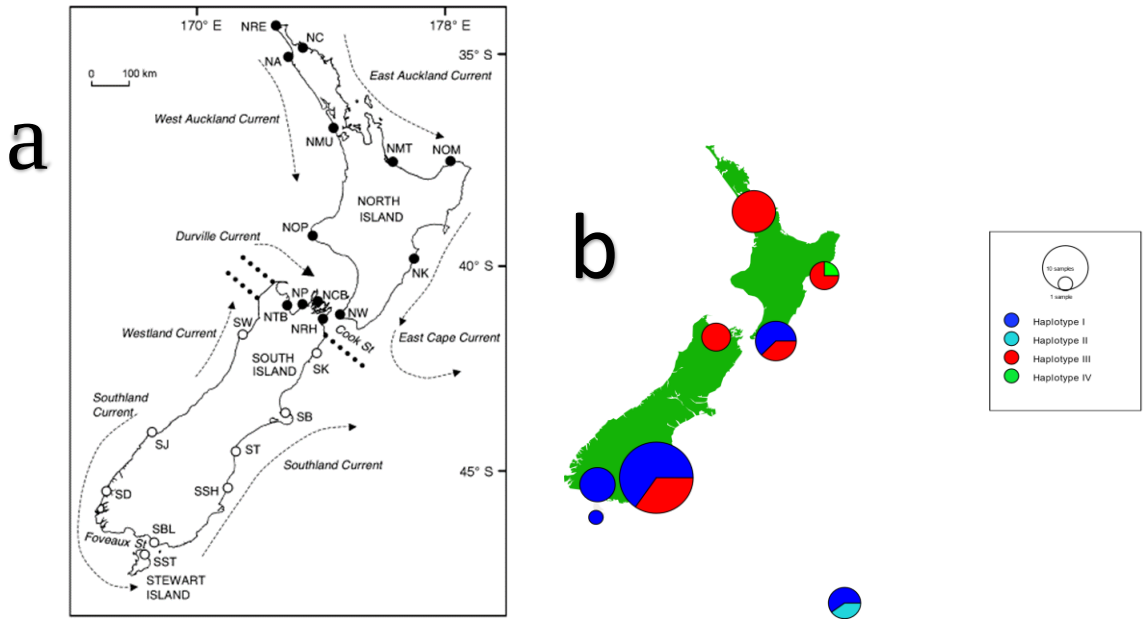


Figure 4.10: a) Figure taken Ayers and Waters (2002). Dotted lines indicate upwelling zones in northern South Island. Sampling localities north of the upwelling zone are by black circles, whereas localities to the south are indicated by open circles. b) ITS2 haplotype map across New Zealand, indicating the separation of four groups.

4.2 Clade II

Antipodes Island lies approximately 740km south-east from Dunedin, New Zealand (49°41'39.4"S 178°45'44.7"E) and have presented an interesting source of genetic variation in this study (Taylor, 1992) (Figure 4.11).

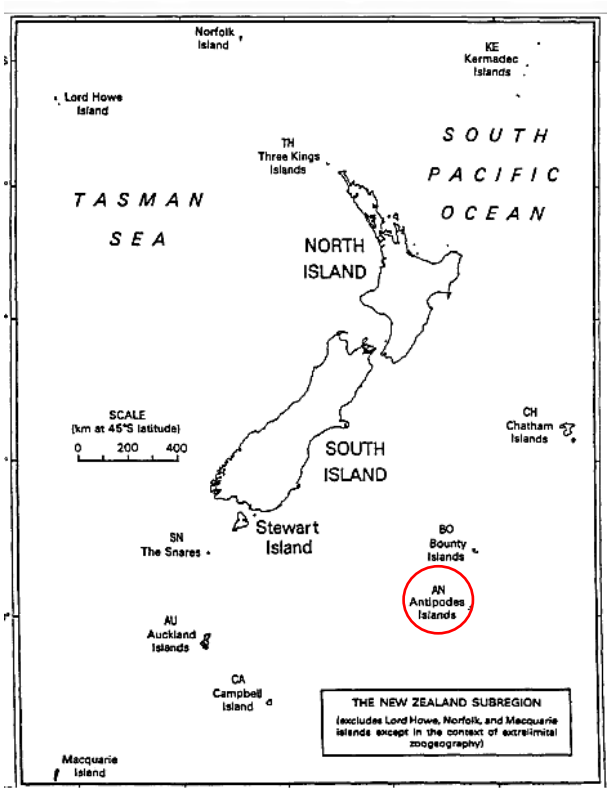


Figure 4.11: Figure adapted from Crosby, Dugdale and Watt (1998). Antipodes Island encircled.

L. hinemoa were collected from the two geographically close locations Anchorage Bay and Hutt Cove on Antipodes Island (Figure 4.12).

Figure 4.12: Figure adapted from Taylor (1992), stars indicate locations that *L. hinemoa* were sampled from (Anchorage Bay and Hutt Cove).

Intriguingly whilst three individuals of the Antipodes Island are in Clade II, three other individuals from the Antipodes Island are in Clade I. However, Clade II whilst close to Clade I is distinct, as shown for the mitochondrial gene COIII (Figure 4.1) and the nuclear gene ITS2 (Figure 4.2). The distinction between these two clades found on Antipodes Island is particularly curious, as those in Clade I are closer phylogenetically to those from the far-distant subantarctic Kerguelen Islands.

Durvillea antarctica is a buoyant species of bull kelp found throughout the subantarctic (Smith, 2002). Kelp is an integral feature of benthic environments of subantarctic, providing an important habitat for invertebrate species. Several marine invertebrate species disperse via kelp rafting, allowing them to access a broader geographic range than would otherwise be anticipated (Fraser, Nikula, & Waters, 2010). Helmuth, Veit and Holberton (1994) showed that dispersal via kelp rafting for the brooding bivalve *Gaimardia trapesina*, serves as an important dispersal mechanism via the Antarctic Circumpolar Current (ACC). Warham and Johns (1975) noted on their 1969 expedition of the Antipodes Island, when they tried to come ashore via Stella Bay (close to Hutt Cove) that the beach was densely covered with the bull kelp *D. antarctica*. The ACC current could explain the movement of genetically similar but geographically distant *Lasaea*, transporting them across the vast the distance between the Kerguelen Islands and Antipodes Island (Figure 4.13).

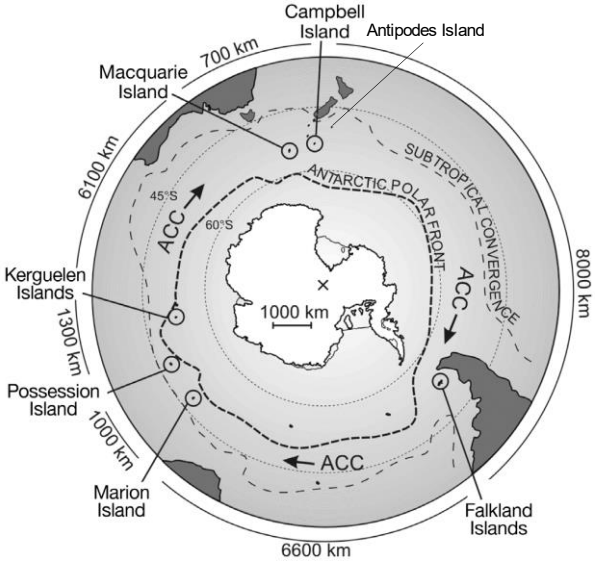


Figure 4.13: Figure Adapted from Nikula, Fraser, Spencer and Waters (2004), showing the path of the Antarctic Circumpolar Current (ACC) around the subantarctic.

4.3 Clade IV

Of all 77 *Lasaea* sequenced for COIII and 69 individuals sequenced for ITS2, the majority of individuals divided neatly into two clades. Those clades that had been previously identified by Taylor and Ó Foighil (2000). However, two individuals that were sequenced for COIII (APM_01 & KPP_07) and one individual sequenced for ITS2 (APM_01) were found to be genetically distant and instead clustered more closely to foreign *Lasaea*. In the combined-gene tree, both Bayesian and maximum likelihood methods showed a clear distinction for this grouping, with strong bootstrap and posterior probability support (Figure 4.14).

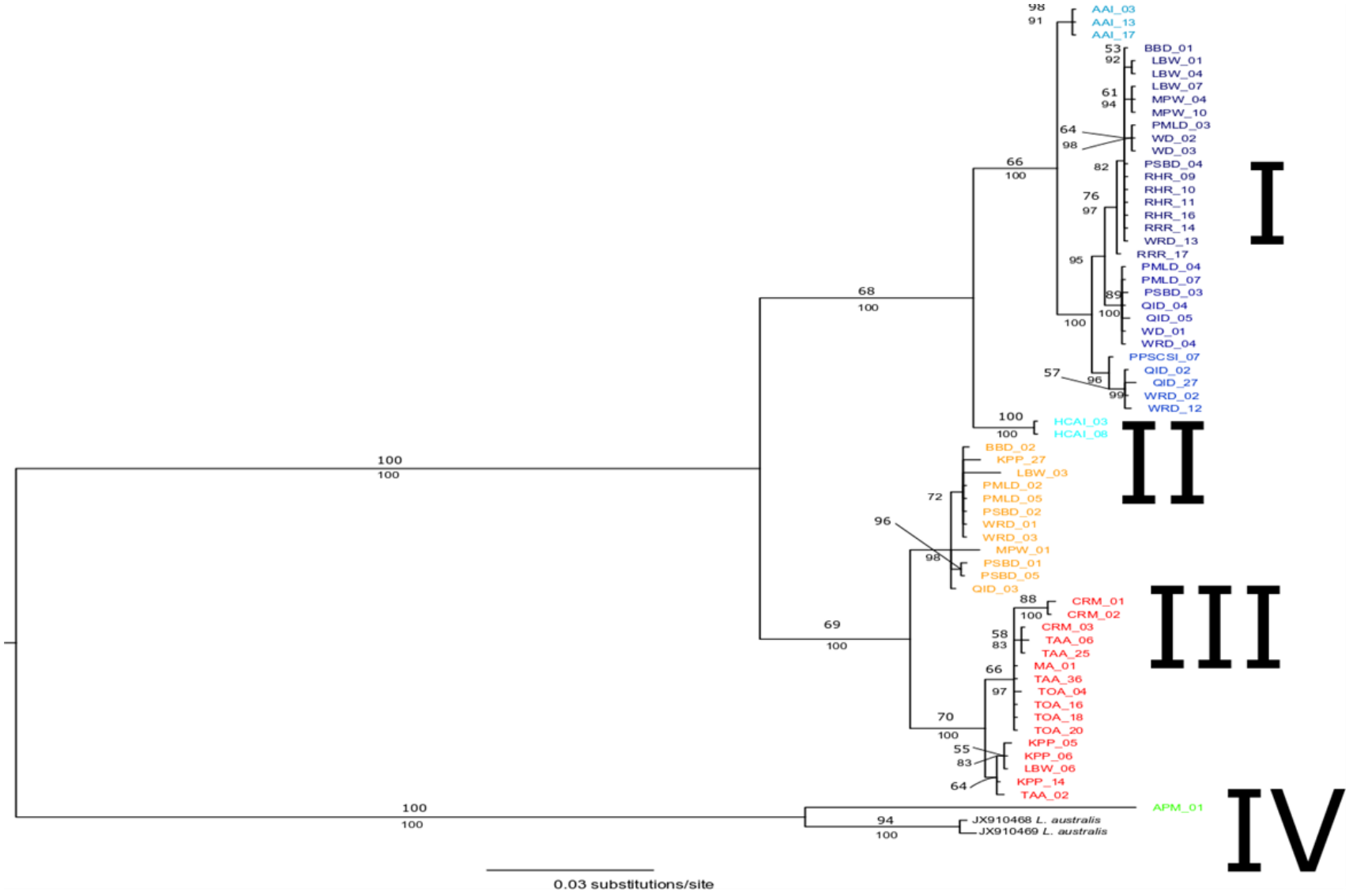

AAI_03
AAI_13
AAI_17
BBD_01
LBW_01
LBW_04
LBW_07
MPW_04
MPW_10
PMLD_03
WD_02
WD_03
PSBD_04
RHR_09
RHR_10
RHR_11
RHR_16
RRR_14
WRD_13
RRR_17
PMLD_04
PMLD_07
PSBD_03
QID_04
QID_05
WD_01
WRD_04
PPSCSI_07
QID_02
QID_27
WRD_02
WRD_12
I
HCAI_03
HCAI_08
II
BBD_02
KPP_27
LBW_03
PMLD_02
PMLD_05
PSBD_02
WRD_01
WRD_03
MPW_01
PSBD_01
PSBD_05
QID_03
CRM_01
CRM_02
CRM_03
TAA_06
TAA_25
MA_01
TAA_36
TOA_04
TOA_16
TOA_18
TOA_20
KPP_05
KPP_06
LBW_06
KPP_14
TAA_02
III
APM_01
JX910468 L. australis
JX910469 L. australis
IV
0.03 substitutions/site

Figure 4.14: Combined COIII and ITS2 Bayesian tree, Bayesian posterior probabilities (values less than 75% are not included) are indicated below the nodes whilst maximum likelihood bootstrap support values (values less than 50% are not indicated) are recorded above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A. Clades are divided into four main groupings; I (LBW_07-AAI_17), II (HCAI_03-HCAI_08), III (BBD_02-TOA_20) and IV (APM_01).

Having such support in both genes examined, gives strong support for this result being genuine.

There are two main lines of explanation for this result, each with lines of support. One possibility is that this clade could comprise *L. australis* that has made the trans-Tasman journey, either naturally or through anthropogenic means. Another possibility for this clade's presence in New Zealand is that they are direct developers that have dispersed from much further away or arrived through anthropogenic means.

The phylogenetic support for this clade consisting of *L. australis* comes from the sequencing completed for the COIII gene (Figure 4.15) and the ITS2 gene (Figure 4.16). The ITS2 phylogenetic tree places clade IV within three clades of *L. australis* and when a TCS network is constructed, the green individual found in New Zealand clusters with two *L. australis* that had been located in the Flindersian biogeographic province on the Southern coast of Australia (Li et al., 2013).

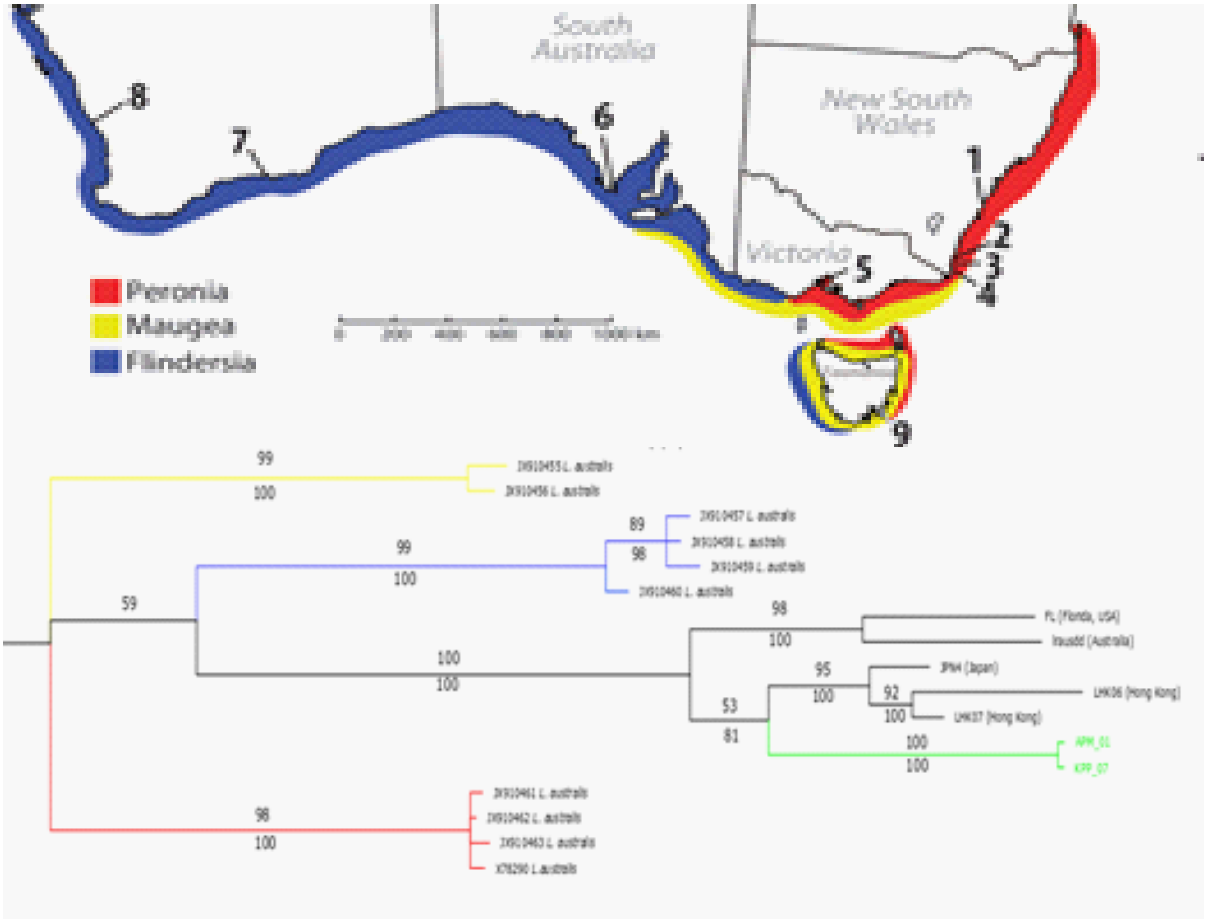


Figure 4.15: Bayesian tree for COIII sequences for *L.* australis and Clade IV, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst Bayesian posterior probabilities (values below 75% are not shown) are indicated below nodes The three *L. australis* clades indicated by Li et al. (2013); colours on the phylogenetic tree correspond to biogeographic province.

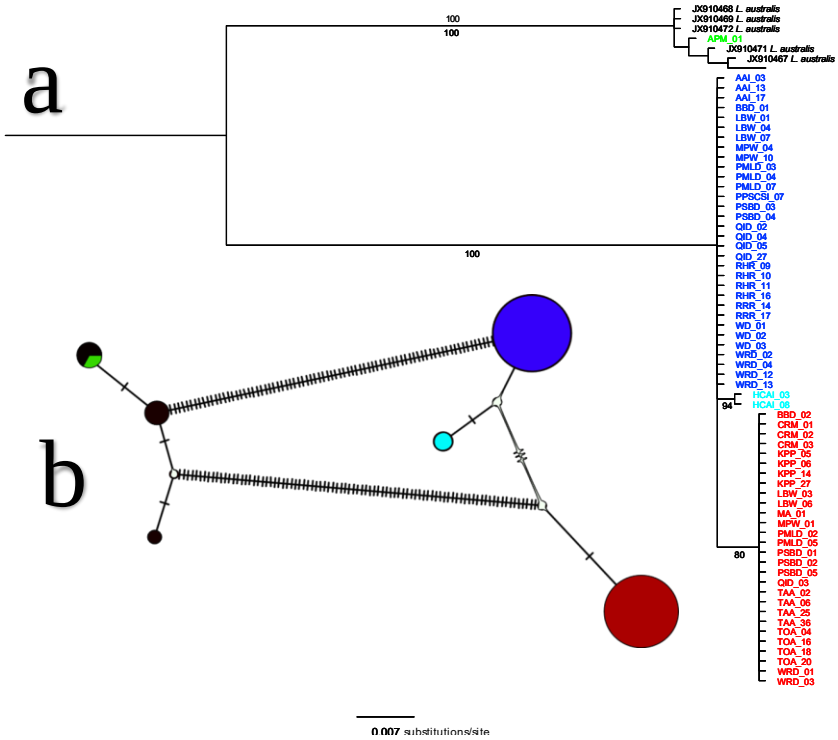


Figure 4.16: Bayesian tree for ITS2 sequences, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes, whilst posterior probabilities (values below 75% are not shown) are indicated below nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A. and ITS2 TCS network for *L. hinemoa*
(*L. australis* serves as the outgroup).

New Zealand biota has a long history of colonisation by species of Australian origin (Sanmartín & Ronquist, 2004), marine life being no exception. As discussed in Chapter 1 (Section 1.1c), *L. australis*, unlike all other *Lasaea* species has planktotrophic development and pelagic larvae. Animals with pelagic larvae such as *Onchidella nigricans* have dispersed trans-Tasman before via a complex system of oceanic currents (Cumming, Nikula, Spencer, & Waters, 2016). If Clade IV consists of *L. australis* it could have dispersed by being swept up by the EAC (East Australian Current) into the Tasman front and into New Zealand waters (Figure 4.17).

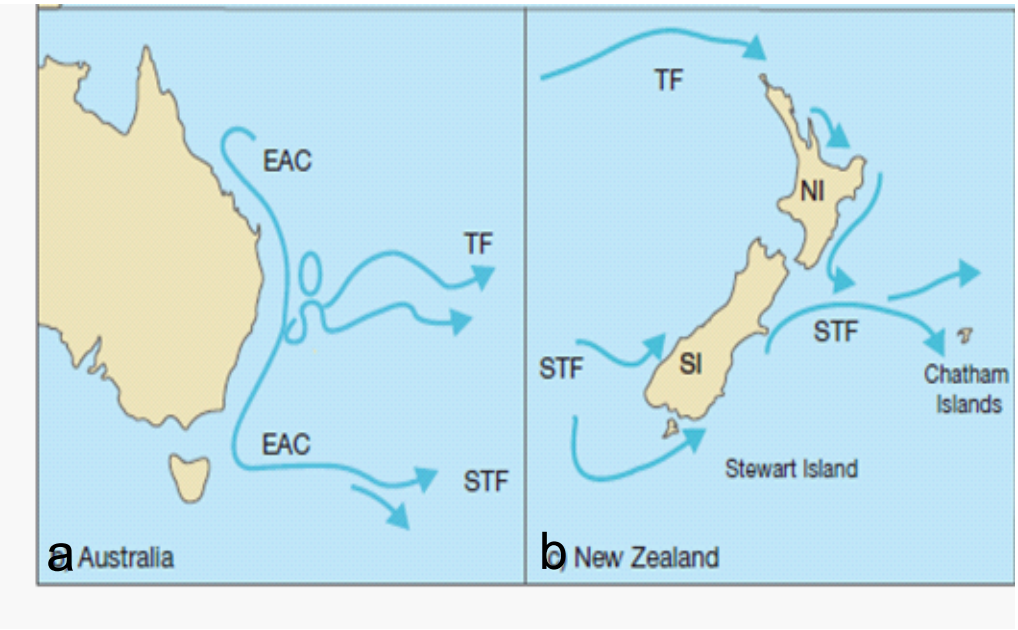


Figure 4.17: Figure adapted from Cumming et al. (2016). a) Australia and b) New Zealand oceanic currents (EAC= East Australian Current; TF= Tasman Front; STF=Sub-Tropical Front; NI=North Island; SI=South Island).

Floating in the water column and possible subsequent attachment to drifting objects has also been suggested as methods of dispersal for other bivalve molluscs (Highsmith, 1985; Martel & Chia, 1991). The Antarctic brooding bivalve *Mysella charcoti* has even been found to disperse via unsuccessful predation by *Notothenia coriiceps*, having been found alive in expelled faeces (Domaneschi, Da Silva, Neto, & Passos, 2002).

However, it is important to note that the New Zealand individual sequenced falls closer to the Flindersian clade, rather than either of the more likely Peronian and Maugean clades, which are found geographically closer to the hypothesised dispersing currents. The Flindersian clade is further away from the EAC, and it would be harder for them to be swept up into that current. This inconsistency could be due to the other two clades being more adapted to warmer surface water and thus unable to survive in colder New Zealand waters. *L. australis* are thought not to disperse into each other's biogeographic provinces due to water temperature differences despite having the currents as a dispersal mechanism enabling them to do so (Li et al., 2013). It is also important to note that this individual sequenced for ITS2 is but one individual, and with more intensive sampling and successful sequencing other individuals from the other two clades could come to light. However, as this individual appears come closer to the Flindersian clade,

anthropogenic introduction also remains a possibility. Godwin (2003) found that the natural barriers to marine species invasions can be overcome by human-mediated dispersal such as maritime vessel hull fouling. The location of this sample, Picton, is a port town and subject to the Durville current that runs through Cook Strait (Bowman, Kibblewhite, Murtagh, Chiswell, & Sanderson, 1983); if *L. australis* was non-intentionally brought into this area of New Zealand it could easily be swept up towards Mahia and produce the distribution seen for Clade IV by COIII sequencing (Figure 4.18).

Figure 4.18: Figure from Wallis and Trewick (2009) indicating water currents around New Zealand.

Clade IV's restricted distribution in New Zealand could be due to the colder water temperatures of southern New Zealand, but sampling difficulties could also provide an explanation.

However, there is more than one explanation for the composition of the Clade IV. Another possibility is that this clade could comprise direct developing polyploid *Lasaea*. The evidence for this idea is primarily based upon the COIII gene phylogenetic tree. The New Zealand clade whilst placed close to the *L. australis* clade, is nestled with direct developing polyploids with a global distribution (Figure 4.15).

These direct-developing polyploids that cluster so close to Clade IV mitochondrially are from Florida (USA), Australia, Hong Kong and Japan. This distribution is rather broad, so it seems reasonable that they have rafted trans-oceanically or they have arrived via anthropogenic means. Clade IV might have made the trans-Tasman journey but as a direct developer. As an asexual direct-developer it would be easier to colonise rapidly, as it would not have to find other individuals for cross-fertilization.

4.4 Conclusions

There are several conclusions that can be made after this study of *L. hinemoa*. There is evidence via several genetic markers that *L. hinemoa* consists of two main distinct species (Clade I and Clade III).

Antipodes Island consists of two distinct *L. hinemoa* clades (Clade I and Clade II), and Clade II might have rafted via the Antarctic Circumpolar Current (ACC) to the Antipodes Island.

Clade IV consists of individuals that are not *L. hinemoa*, but instead appear to represent either *L. australis* or a trans-oceanic group of direct-developing, asexual *Lasaea*. These individuals have arrived in New Zealand by trans-oceanic rafting or through anthropogenic introduction.

4.5 Future Directions

There are several future directions that became apparent from this study.
Clade I and III could represent separate cryptic species but further investigation is warranted to whether they have different resource requirements in some way so that they can co-exist in many places.

Clade IV warrants more investigation in New Zealand. The developmental mode should be studied to elucidate whether it exhibits planktotrophic or direct development. By understanding what developmental mode it exhibits it will be easier to ascertain which species it should be considered as.

Toman and Flegr (2018) posit that asexual *Lasaea* experience less selection pressure than sexual *Lasaea*, arguing that sexual *Lasaea* experience greater selection pressure through predation and parasitism (as with the red queen hypothesis). However, this theory hasn't been empirically tested, and doing so is warranted to ascertain the mechanisms maintaining the dominance of asexual *Lasaea* species.

4.6 References

Apte, S., and J. P. A.Gardner. 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*(9), 1617–1628.

Avise, J. C. (1989). A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology & Evolution*, *4*(9), 279-281.

Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, *36*(1), 3-15.

Ayers, K. L., & Waters, J. M. (2005). Marine biogeographic disjunction in central New Zealand. *Marine Biology*, *147*(4), 1045-1052.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., ... & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148-155.

Booth, J. D. (1979). Common bivalve larvae from New Zealand: leptonacea. *New Zealand Journal of Marine and Freshwater Research*, *13*(2), 241-254.

Bowman, M. J., Kibblewhite, A. C., Murtagh, R. A., Chiswell, S. M., & Sanderson, B. G. (1983). Circulation and mixing in greater Cook Strait, New Zealand. *Oceanologica acta*, *6*(4), 383-391.

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography*. Sunderland, MA: Sinauer Associates.

Crisp, D. J., & Standen, A. (1988). *Lasaea rubra* (Montagu)(Bivalvia: Erycinacea), an apomictic crevice-living bivalve with clones separated by tidal level preference. *Journal of Experimental Marine Biology and Ecology*, *117*(1), 27-45.

Crosby, T. K., Dugdale, J. S., & Watt, J. C. (1998). Area codes for recording specimen localities in the New Zealand subregion. *New Zealand Journal of Zoology*, *25*(2), 175-183.

Cumming, R. A., Nikula, R., Spencer, H. G., & Waters, J. M. (2016). Trans-Tasman genetic connectivity in the intertidal air-breathing slug *Onchidella nigricans*. *Marine Ecology Progress Series*, *562*(1), 93-100.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879-886.

Derycke, S., De Meester, N., Rigaux, A., Creer, S., Bik, H., Thomas, W. K., & Moens, T. (2016). Coexisting cryptic species of the *Litoditis marina* complex (Nematoda) show differential resource use and have distinct microbiomes with high intraspecific variability. *Molecular Ecology*, *25*(9), 2093-2110.

Domaneschi, O., Da Silva, J., Neto, L., & Passos, F. (2002). New perspectives on the dispersal mechanisms of the Antarctic brooding bivalve *Mysella charcoti* (Lamy, 1906). *Polar Biology*, *25*(7), 538-541.

Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, *88*(3), 172-181.

Francis, M. P. (1996). Geographic distribution of marine reef fishes in the New Zealand region. *New Zealand Journal of Marine and Freshwater Research*, *30*(1), 35-55.

Fraser, C. I., Nikula, R., & Waters, J. M. (2010). Oceanic rafting by a coastal community. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1706), 649-655.

Gaston, K. J., Chown, S. L., & Evans, K. L. (2008). Ecogeographical rules: elements of a synthesis. *Journal of Biogeography*, *35*(3), 483-500.

Gittenberger, E. (1991). What about non-adaptive radiation?. *Biological Journal of the Linnaean Society*, *43*(4), 263-272.

Godwin, L. S. (2003). Hull fouling of maritime vessels as a pathway for marine species invasions to the Hawaiian Islands. *Biofouling*, *19*(1), 123-131.

Goldstein, P. Z., Desalle, R., Amato, G., & Vogler, A. P. (2000). Conservation genetics at the species boundary. *Conservation Biology*, *14*(1), 120-131.

Helmuth, B., Veit, R. R., & Holberton, R. (1994). Long-distance dispersal of a subantarctic brooding bivalve (*Gaimardia trapesina*) by kelp-rafting. *Marine Biology*, *120*(3), 421-426.

Highsmith, R. C. (1985). Floating and algal rafting as potential dispersal mechanisms in brooding invertebrates. *Marine Ecology Progress Series*, *25*(2), 169-179.

Knox, G. A. (1975). *Biogeography and Ecology in New Zealand*. Dordrecht: Springer.

Lee, C. E., & Frost, B. W. (2002). Morphological stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia*, *480*(1-3), 111-128.

Li, J., Ó Foighil, D., & Park, J. K. (2013). Triton's trident: cryptic Neogene divergences in a marine clam (*Lasaea australis*) correspond to Australia's three temperate biogeographic provinces. *Molecular Ecology*, *22*(7), 1933-1946.

Lorenz, J. G., Jackson, W. E., Beck, J. C., & Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1869-1877.

Martel, A., & Chia, F. S. (1991). Drifting and dispersal of small bivalves and gastropods with direct development. *Journal of Experimental Marine Biology and Ecology*, *150*(1), 131-147.

Moore, L. B. (1949). The marine algal provinces of New Zealand. *Transactions of the Royal Society of New Zealand*, *77*(1), 187-189.

Nakano, T., & Spencer, H. G. (2007). Simultaneous polyphenism and cryptic species in an intertidal limpet from New Zealand. *Molecular Phylogenetics and Evolution*, *45*(2), 470-479.

Nelson, W. A. (1994). Distribution of macroalgae in New Zealand—an archipelago in space and time. *Botanica Marina*, *37*(3), 221-234.

Nikula, R., Fraser, C. I., Spencer, H. G., & Waters, J. M. (2010). Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Marine Ecology Progress Series*, *405*(1), 221-230.

Ó Foighil, D. (1989). Planktotrophic larval development is associated with a restricted geographic range in *Lasaea*, a genus of brooding, hermaphroditic bivalves. *Marine Biology*, *103*(3), 349-358.

Ó Foighil, D., & Jozefowicz, C. J. (1999). Amphi-Atlantic phylogeography of direct-developing lineages of *Lasaea*, a genus of brooding bivalves. *Marine Biology*, *135*(1), 115-122.

Pante, E., Schoelinck, C., & Puillandre, N. (2014). From integrative taxonomy to species description: one step beyond. *Systematic Biology*, *64*(1), 152-160.

Payo, D. A., Leliaert, F., Verbruggen, H., D'hondt, S., Calumpong, H. P., & De Clerck, O. (2013). Extensive cryptic species diversity and fine-scale endemism in the marine red alga *Portieria* in the Philippines. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1753), 2012-2660.

Ponder, W. F. (1971). Some New Zealand and subantarctic bivalves of the Cyamiacea and Leptonacea with descriptions of new taxa. *Records of the Dominium Museum*, *7*(13), 119-141.

Puillandre, N., Sysoev, A. V., Olivera, B. M., Couloux, A., & Bouchet, P. (2010). Loss of planktotrophy and speciation: geographical fragmentation in the deep-water gastropod genus *Bathytoma* (Gastropoda, Conoidea) in the western Pacific. *Systematics and Biodiversity*, *8*(3), 371-394.

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, *4*(1), 11.

Sanmartín, I., & Ronquist, F. (2004). Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic biology*, *53*(2), 278-298.

Schönrogge, K., Barr, B., Wardlaw, J. C., Napper, E., Gardner, M. G., Breen, J., Elmes, G. W., & Thomas, J. A. (2002). When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnaean Society*, *75*(3), 291-300.

Shears, N. T., Smith, F., Babcock, R. C., Duffy, C. A., & Villouta, E. (2008). Evaluation of biogeographic classification schemes for conservation planning: Application to New Zealand's coastal marine environment. *Conservation Biology*, *22*(2), 467-481.

Simon, D. (2008). Biogeography-based optimization. *IEEE Transactions on Evolutionary Computation*, *12*(6), 702-713.

Smith, S. D. (2002). Kelp rafts in the Southern Ocean. *Global Ecology and Biogeography*, *11*(1), 67-69.

Sponer, R., & Roy, M. S. (2002). Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution*, *56*(10), 1954-1967.

Taylor, R. H. (1992). New Zealand fur seals at the Antipodes Islands. *Journal of the Royal Society of New Zealand*, *22*(2), 107-122.

Toman, J., & Flegr, J. (2018). General environmental heterogeneity as the explanation of sexuality? Comparative study shows that ancient asexual taxa are associated with both biotically and abiotically homogeneous environments. *Ecology and Evolution*, *8*(2), 973-991.

Wallis, G. P., & Trewick, S. A. (2009). New Zealand phylogeography: evolution on a small continent. *Molecular ecology*, *18*(17), 3548-3580.

Walls, K. (1995). The New Zealand experience in developing a marine biogeographic regionalisation. *Towards a marine regionalisation for Australia. Townsville: Great Barrier Reef Marine Park Authority*, 33-48.

Warham, J., & Johns, P. M. (1975). The university of Canterbury antipodes island expedition 1969. *Journal of the Royal Society of New Zealand*, *5*(2), 103-131.

Appendices

Appendix A: Sampling Information for Lasaea hinemoa

Table A.1: Sites in Auckland (36°50'53.8"S 174°45'39.7"E) explored for *L. hinemoa* sampling.

Time	Location (area)	Location (specific)	Description
07/01/2018 (5pm-9pm)	Auckland	• Long Bay Beach (LBBA) • Between Long Bay Beach and Waiake Beach (including Winstone's Cove) (BLBWA) • Waiake Beach (WBA)	LBBA was predominately searched at the Southern end. BLBWA lots of reef and rocks – occasional individual *L. hinemoa* found (walked between the beaches on the reef at low tide). WBA more *L. hinemoa* found around the Tor. *This region was later combined and recategorized as Torbay (TOA).
08/01/2018 (6.30am-3pm)	Auckland	• Takapuna Beach (TBA) • Takapuna Reef (TRA)	TBA had many *L. hinemoa* on the rocks near the reef. TRA had *L. hinemoa* in abundance. More invertebrates

		• Milford (MA) • Castor Bay (CtBA) • Campbells Bay (CpBA) • Mairangi Bay Beach (MgBA) • Murrays Bay Beach (MBBA) • Rothesay Bay (RBBA) • Browns Bay Beach (BBBA)	than in other places in the alga. MA had small numbers of *L. hinemoa*, tide was coming up. CtBA was a small bay, and reef wasn't easy to access, tide was coming up. CpBA had some rocky ground but no suitable alga/habitat. MBBA was a wide bay, nothing found. BBBA might have had *L. hinemoa* out at reef but tidal conditions were too high to check. *Takapuna beach and reef were combined and recategorized as Takapuna (TAA)

Table A.2: Sites in Hawkes Bay (39°13'29.9"S 177°15'41.8"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
28/12/2017 (7pm-9pm)	Napier	• Gannet Beach (GBN) • Te Awanga Beach (TABN) Napier Beach (NBN)	Low Tide = 8.07pm. Beaches were all loose stone beaches, and suitable habitat for *L. hinemoa* could not be found.
29/12/2017 (8.30am-10am) (~6pm)	Napier- Mahia	• Bayview Beach (BBN) • Waipatiki Beach (WBH) • Tangoio Beach (TBH) • Mahanga Beach (MgBM)	BBN was a loose stone beach with no suitable habitat for *L. hinemoa*. WBH and TBH had suitable conditions but no *L. hinemoa* present in alga. MgBM had suitable habitat but no *L. hinemoa*.
30/12/2017 (9am-12.30pm)	Mahia	• Auroa Point (APM) • Left of Coronation Reserve Area (CRLM) • Right of Coronation Reserve Area (CRRM) • Mahia Beach (MaBM) • Opoutama beach (OBM)	APM was a reef but after much searching small numbers of *L. hinemoa* could be found. CRLM and CRRM had suitable habitat, numbers were more abundant at CRRM. MaBM had a rocky outcrop but no *L. hinemoa* found. OBM had rocks and alga but no *L. hinemoa*).

Table A.3: Sites from New Plymouth (39°03'20.4"S 174°04'57.9"E) to Opunake (39°27'20.5"S 173°50'59.0"E`) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
03/01/2018 West coast (4pm-8.30pm)	New Plymouth – Opunake	• Fitzroy Beach (FBNP) • East End Beach (EEBNP) • Beach area near Wind wand (WWNP) • Ngamotu Beach (NBNP) • Oakura Beach (OBO) • Opunake Beach (OBO) • Opunake Beach (Dog's beach) (DBO)	Beaches had suitable rocky shore but no suitable alga. Flea mussels seem to dominate rocks.

Table A.4: Sites in Whanganui (39°55'49.4"S 175°02'52.6"E) explored for *L. hinemoa* sampling. Highlighting indicates sites not explored personally.

Date	Location (Area)	Location (specific)	Description
04/01/2018 (6pm-8pm)	Whanganui	• Ototoka Beach (OBW) • Kai iwi beach (KIBW)	OBW had none of the right alga. KIBW had rocks but none of the right alga. Weather conditions were very wet.

15/01/2018	Whanganui	• Castlecliff Beach (CBW) • South Beach (SBW)	N/A

Table A.5: Sites in Wellington (41°17'22.6"S 174°46'59.7"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
09/01/2018 (6.30am-10.15am)	Wellington	• Lyall Bay (LBW) • Moa Point (MPW) • Breaker Bay (BBW) • Between Scorching Bay and Mahanga Bay (BSBMW)	LBW had a natural rock formation where *L. hinemoa* were found in abundance. MPW had lots of available habitat and rocks, and abundant *L. hinemoa*. BBW had no *L. hinemoa* on the rocks. BSBMW produced small numbers of *L. hinemoa*.

Table A.6: Sites in the Picton (41°17'33.4"S 174°00'02.9"E)-Havelock (41°17'05.4"S 173°46'09.3"E)-Rarangi (41°23'38.8"S 174°02'43.3"E) area explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
10/01/2018	Picton	• Shakespeare Bay (SBP)	SBP had no suitable alga.
11/01/2018	Picton- Havelock	• Waikawa Bay (WBP) • Karaka Point (KPP) • Whatamango Bay (WmBP)	WBP was built up for locals and had no suitable habitat. KPP had an exposed rocky shore and abundant *L. hinemoa*. WmBP was a mudflat area and was not

		• Governors Bay (GBP) • Ngakuta Bay (NBP) • Momorangi Bay (MrP) • Aussie Bay (ABP) • Anakiwa (AkP)	suitable for *L. hinemoa*. GBP, NBP, ABP and AkP had no suitable habitat. MrP had rocks but no alga.
12/01/2018	Picton – Rarangi	• Ahuriri Bay (AhBP) • Oyster Bay (OyBP) • Ocean Bay (OBP) • Robin Hood Bay (RHBP) • Monkey Bay (MkBP) • Rarangi (RBR)	Heavy rain and no suitable habitat available for *L. hinemoa*.

Table A.7: Sites in Kaikoura (42°24'02.7"S 173°41'03.2"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
14/01/2018	Kaikoura	• Esplanade (KEK)	Loose stone beach and no suitable alga.

Table A.8: Sites in Akaroa (43°48'13.8"S 172°58'10.9"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
15/01/2018	Akaroa	• Barry's Bay (BBA) • Duvauchelle Bay (DBA) • Robinson's Bay (RBA) • Akaroa Harbour (AHA) • Akaroa Beach (French Bay) (FBA) • Children's Bay (CBA)	Most beaches were very sandy with no suitable habitat. Small numbers of *L. hinemoa* were found at CBA.

Table A.9: Sites in Dunedin (45°52'47.0"S 170°30'21.0"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
	Dunedin	• Warrington (WD)	Abundant *L. hinemoa* found on the rocks.
	Dunedin	• Broad Bay (BBD) • Weller's Rock (WRD)	BBD small numbers of *L. hinemoa*, WRD abundant *L. hinemoa*.
14/02/2018	Dunedin	• Portobello Marine Lab (PMLD) • Quarantine Island (QID) • Portobello South Beach (PSBD)	Abundant numbers of *L. hinemoa* at all locations.

Table A.10: Sites in Riverton (46°21'51.2"S 168°00'47.9"E) explored for *L. hinemoa* sampling.

Date	Location (Area)	Location (specific)	Description
27/03/2018	Riverton	• Riverton Highway (RHR) • Riverton Rocks (RRR)	Very abundant *L. hinemoa* found on rocks.

Table A.11: Sites in Stewart Island (46°59'43.1"S 167°51'17.2"E) explored for *L. hinemoa* sampling. Highlighting indicates sites not explored personally.

Date	Location (Area)	Location (specific)	Description
04/2018	Stewart Island	• Port Pegasus, Shipbuilder's Cove (PPSCSI)	N/A

Table A.12: Sites in the Antipodes Islands (49°41'39.4"S 178°45'46.7"E) explored for *L. hinemoa* sampling. Highlighting indicates sites not explored personally.

Date	Location (Area)	Location (specific)	Descrition
06/03/2018	Antipodes Islands	• Hut Cove (HCAI) • Anchorage (AAI)	HCAI, Hand collection in intertidal zone (low tide) from under rocks/in cracks and algal scraping. AAI hand collection on rock platform from under rocks/in cracks, and algal scraping

Table A.13: Locations were *L. hinemoa* where found, and approximate numbers recorded. Highlighting indicates areas where *L. hinemoa* where not personally found.

Location (area)	Location (specific)	Total numbers
Auckland	• Between Long Bay and Waiake Beach (~20) • Waiake Beach (~15) • Milford (1) • Takapuna Beach (~50) • Takapuna Reef (~50)	~136 *L. hinemoa*
Mahia	• Coronation Reserve (left) (~10) • Coronation Reserve (Right) (~50) • Auroa Point (2)	~62 *L. hinemoa*
Wellington	• Lyall Bay (~50) • Moa Point (~50) • Between Scorching Bay and Mahanga Bay (1)	~101 *L. hinemoa*
Picton	• Karaka Point (~50)	~50 *L. hinemoa*
Akaroa	• Children's Bay (~10)	~10 *L. hinemoa*
Dunedin	• Warrington (~10) • Broad Bay (~5) • Portobello Marine Lab (~60)	~185 *L. hinemoa*

	• Quarantine Island (~50) • Portobello South Beach (~10) • Weller's Rock (~50)	
Catlins	• Kaka Point (~5)	~5 *L. hinemoa*
Riverton	• Riverton Highway (~50) • Riverton Rocks (~50)	~100 *L. hinemoa*
Stewart Island	• Port Pegasus (35)	35 *L. hinemoa*
Antipodes Island	• Hut Cove (16) • Anchorage (17)	33 *L. hinemoa*

Appendix B: Phylogenetic Trees

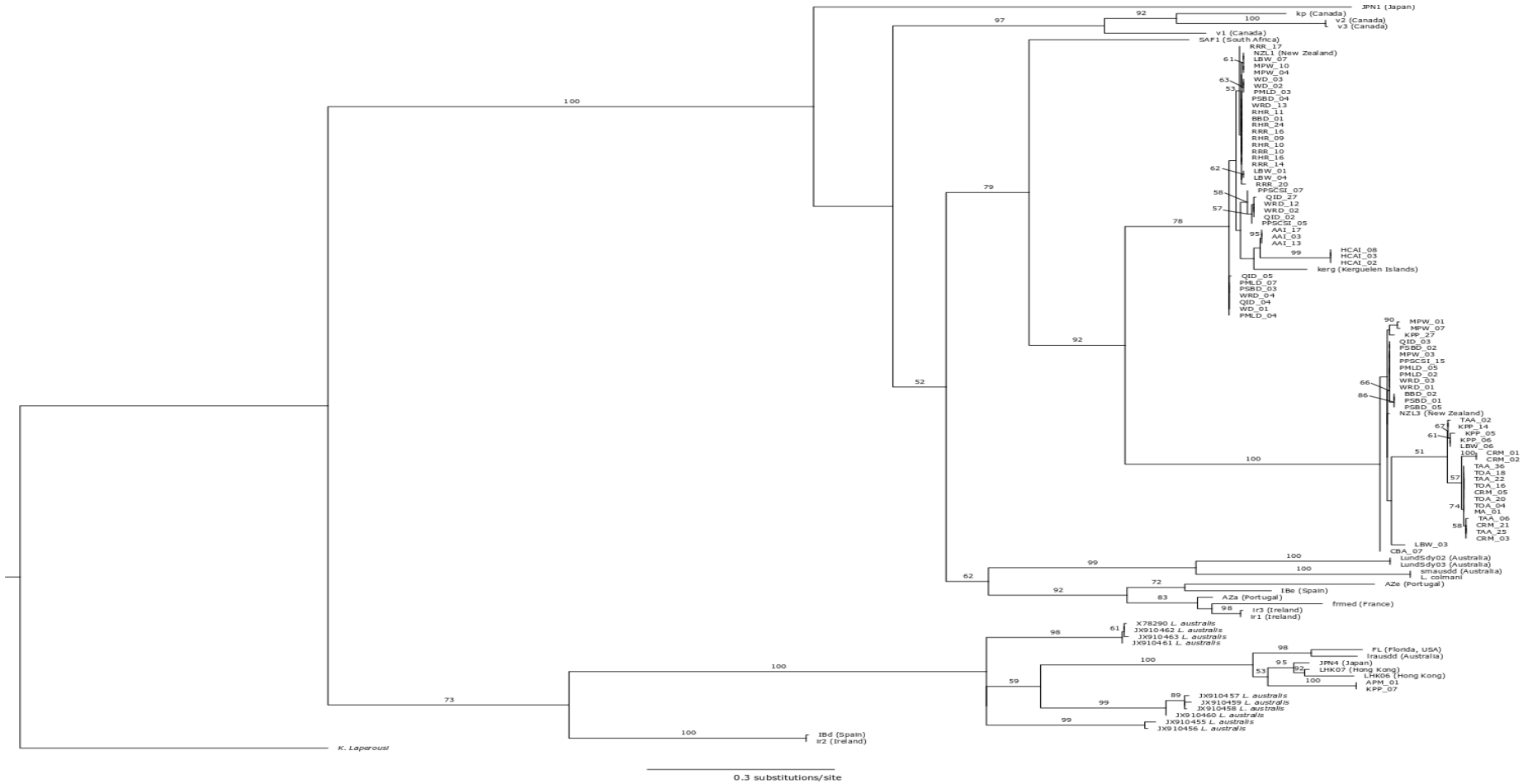


Figure B.1: Maximum likelihood tree for COIII sequences, bootstrap support values (values below 50% are not shown) are indicated above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A

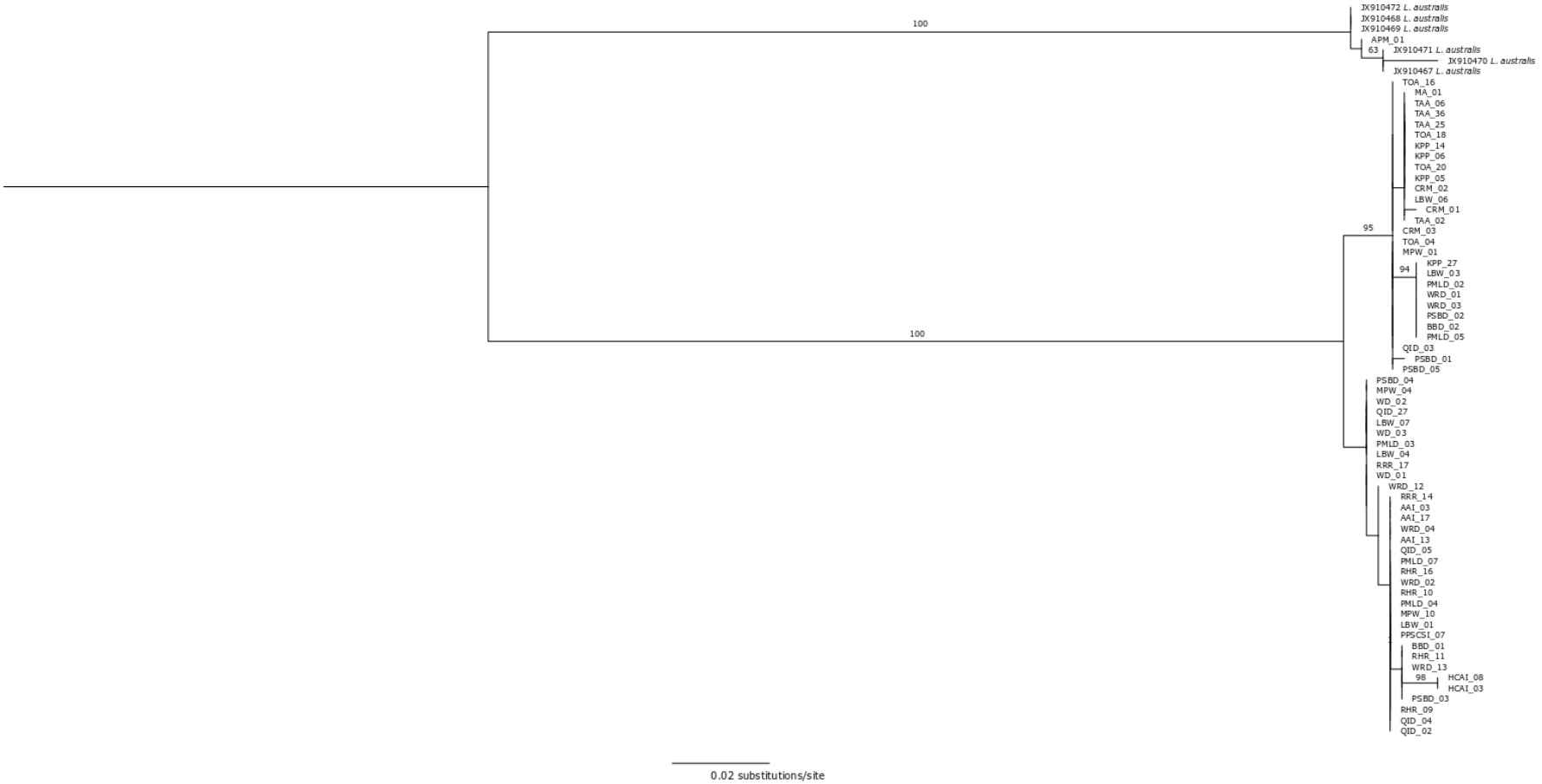


Figure B.2: Maximum likelihood tree for ITS2 sequences, maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A

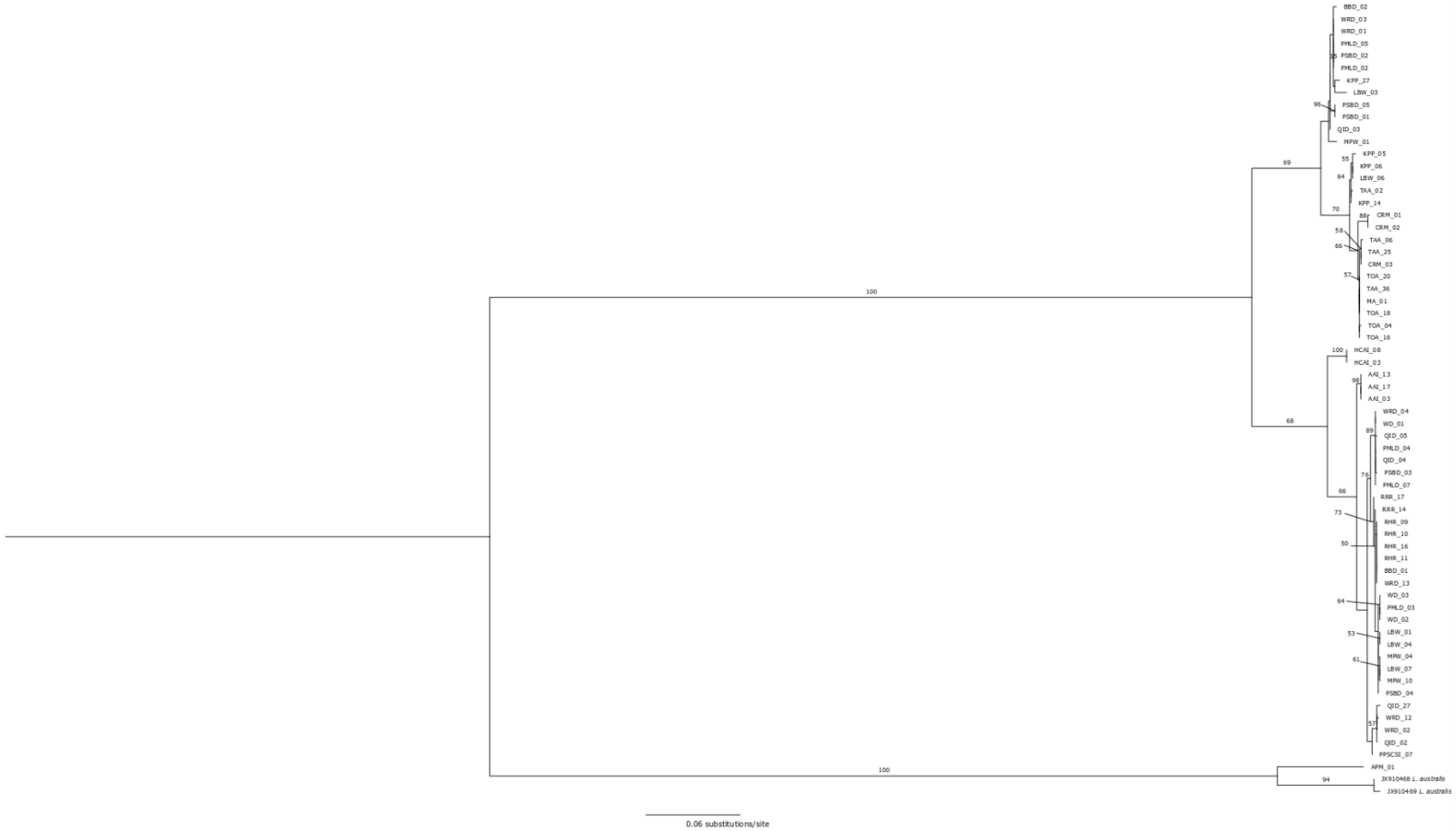


Figure B.3: Combined COIII and ITS2 maximum likelihood tree, maximum likelihood bootstrap support values (values less than 50% are not indicated) are recorded above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A

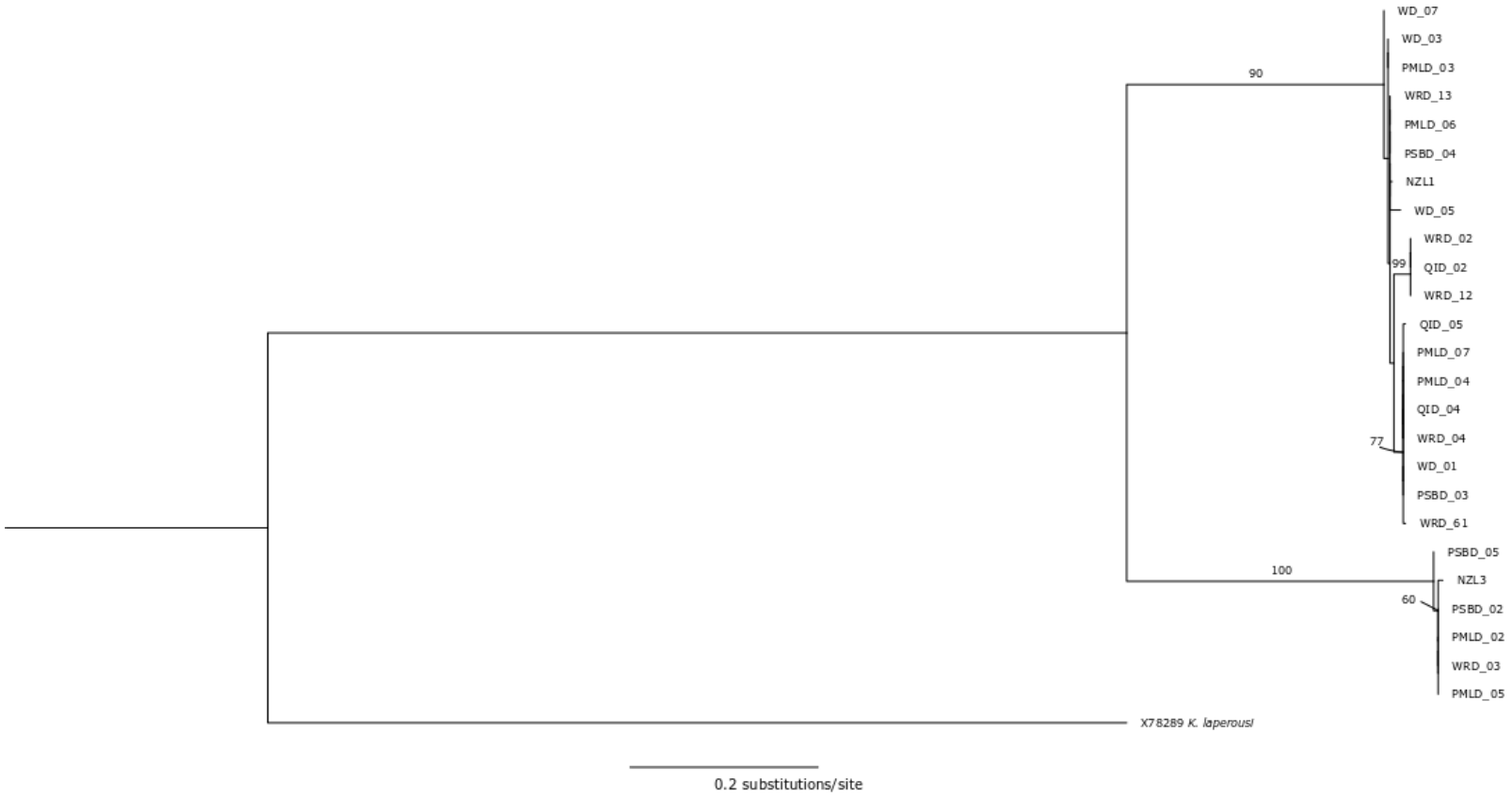


Figure B.4: Maximum likelihood tree for COIII sequences for *L. hinemoa* that were also examined in microsatellite analysis), maximum likelihood bootstrap support values (values below 50% are not shown) are indicated above the nodes. Explanation for the individual labels can be found in Table 2.2 and Appendix A

Appendix C: DAPC (Discriminant Analysis of Population Components) Analysis

```
#Population Analysis
#clear the workspace
rm(list=ls())

#set working directory
getwd()

#install adegenet with dependencies
install.packages("adegenet", dep=TRUE)

#Load the packages
library("ape")
library("pegas")
library("seqinr")
library("ggplot2")
library("adegenet")
library("mmod")

#To check the version of the package being used
packageDescription("adegenet", fields = "Version")
# "2.1.1"

#Loading the DNA file into R
library(ape)
myDNA <-read.dna ("LasaeaCOIII.fasta",format="fasta")

#Check to see it is loaded in correctly
myDNA
# 77 DNA sequences in binary format stored in a matrix.
#All sequences of same length:598
```

#Base composition:

a c g t

0.206 0.135 0.243 0.416

class(myDNA)

myDNA<-as.matrix(myDNA)

Polymorphism can be characterized using snpposi.plot and snpposi.test : the first plots SNP density along the alignment, the second tests whether these SNPs are randomly distributed.

snpposi.plot(myDNA,codon=FALSE)

#By default, the function differentiates nucleotide positions:

snpposi.plot(myDNA)

In adegenet, only polymorphic loci are conserved to form a genind object. This conversion is achieved by DNAbin2genind.

obj <- DNAbin2genind(myDNA)

obj

#/// GENIND OBJECT /////////

#// 77 individuals; 190 loci; 417 alleles; size: 221.8 Kb

#// Basic content

#@tab: 77 x 417 matrix of allele counts

#@loc.n.all: number of alleles per locus (range: 2-4)

#@loc.fac: locus factor for the 417 columns of @tab

#@all.names: list of allele names for each locus

#@ploidy: ploidy of each individual (range: 1-1)

#@type: codom

#@call: DNAbin2genind(x = myDNA)

#// Optional content

#- empty -

#Positions of the SNPs are stored as names of the loci:

head(locNames(obj))

#for the populations you need to create a table with the sequence name, and location and export in .cvs.

myDNA_loc<-read.table("LasaeaCOIII.csv", header = TRUE, sep =",")

myDNA_loc

myDNAloc<-as.data.frame(myDNA_loc)

myDNAloc

Set the locality data in DNA data strata slot.

strata(obj)<-myDNA_loc

obj

#/// GENIND OBJECT /////////

#// 77 individuals; 190 loci; 417 alleles; size: 228.6 Kb

#// Basic content

#@tab: 77 x 417 matrix of allele counts

#@loc.n.all: number of alleles per locus (range: 2-4)

#@loc.fac: locus factor for the 417 columns of @tab

#@all.names: list of allele names for each locus

#@ploidy: ploidy of each individual (range: 1-1)

#@type: codom

#@call: DNAbin2genind(x = myDNA)

#// Optional content

#@strata: a data frame with 2 columns (seq.name, loc)

 # Discriminant Analyses of Principal Components (DAPC)

DAPC necessitates defined prior groups. However, prior groups are often unknown/ uncertain, there is a need to identify genetic clusters before describing them. A

clustering algorithm can be used (k-means) which finds a given number (e.g. k) of groups maximizing the variation between groups. To ascertain the optimal number of clusters, k-means is run sequentially with increasing values of k, and different clustering solutions are compared using Bayesian Information Criterion (BIC). The optimal clustering solution should correspond to the lowest BIC. The best available BIC is typically indicated by an elbow in the curve of BIC values as a function of k. While k-means could be performed on the raw data, instead the algorithm is run after transforming the data using PCA. This reduces the number of variables and consequently speeds up the clustering algorithm.

The clusters are identified by find.clusters. This function transforms the data using PCA, asking the user to specify the number of retained PCs interactively unless the argument n.pca is provided. As there can be more groups than localities, a large number can be utilized to examine clustering (e.g. 40).

```
grp <- find.clusters(obj, max.n.clust=40)
```

The function will display a graph of cumulated variance explained by the eigenvalues of the PCA. There is no reason for keeping a small number of components; all the information can be kept, so it can be specified to retain all PC's. The function displays a graph of BIC values for increasing values of k. The elbow of the curve should match the smallest BIC, and hopefully clearly indicates the number of clusters that should be retained. The output of find.clusters is a list:

```
names(grp)
head(grp$grp, 10)
```

DAPC aims to provide an efficient description of genetic clusters using a few synthetic variables. These are constructed as linear combinations of the original variables (alleles) which have the largest between-group variance and the smallest within-group variance.

```
obj
```

```
pop(obj)<- obj$strata$loc
myDNA_dapc <- dapc(obj)
#Choose the number PCs to retain (>=1):
3
#Choose the number discriminant functions to retain (>=1):
2

myDNA_dapc

################################################
# Discriminant Analysis of Principal Components #
################################################
#class: dapc
#$call: dapc.genind(x = obj)
#$n.pca: 3 first PCs of PCA used
#$n.da: 2 discriminant functions saved
#$var (proportion of conserved variance): 0.865
#$eig (eigenvalues): 81.81 2.223 1.098  vector    length content
#1 $eig      3      eigenvalues
#2 $grp      77     prior group assignment
#3 $prior    9      prior group probabilities
#4 $assign   77     posterior group assignment
#5 $pca.cent 417    centring vector of PCA
#6 $pca.norm 417    scaling vector of PCA
#7 $pca.eig  28     eigenvalues of PCA
#data.frame    nrow ncol content
#1 $tab         77   3    retained PCs of PCA
#2 $means       9    3    group means
#3 $loadings    3    2    loadings of variables
#4 $ind.coord   77   2    coordinates of individuals (principal components)
#5 $grp.coord   9    2    coordinates of groups
#6 $posterior   77   9    posterior membership probabilities
#7 $pca.loadings 417  3    PCA loadings of original variables
```

#8 $var.contr 417 2 contribution of original variables

#Basic scatterplots can be produced by the scatterplot function.

scatter(myDNA_dapc)

Appendix D: Complete set of Microsatellite Primers

Table D.1: Microsatellite primers ordered and subsequently tested for *L. hinemoa*.

Lhin_001_F_M13	TGTAAAACGACGGCCAGTAACCCTAGCCTAACCGTTTG
Lhin_002_F_M13	TGTAAAACGACGGCCAGTGAAGACGTCAAAGCCGCTG
Lhin_003_F_M13	TGTAAAACGACGGCCAGTTCCGACTGAAGGTGACAAGG
Lhin_004_F_M13	TGTAAAACGACGGCCAGTCCTTCCATGCATCATTCCATTC
Lhin_005_F_M13	TGTAAAACGACGGCCAGTTCTGTCTGCCTGTTAGTCGG
Lhin_006_F_M13	TGTAAAACGACGGCCAGTTCGAAGATACCCATGCACAC
Lhin_007_F_M13	TGTAAAACGACGGCCAGTTGCCTATCTACCTGTCTGCC
Lhin_008_F_M13	TGTAAAACGACGGCCAGTGGCCATATGTGTGTTAGTCGG
Lhin_009_F_M13	TGTAAAACGACGGCCAGTCCCGCTCCATAATTACAGCG
Lhin_010_F_M13	TGTAAAACGACGGCCAGTAGTCACTTACACGCACTACC
Lhin_011_F_M13	TGTAAAACGACGGCCAGTGCAGATGAACTAGGTCTACGC
Lhin_012_F_M13	TGTAAAACGACGGCCAGTTAGATCCGCCCTAATGCTGC
Lhin_013_F_M13	TGTAAAACGACGGCCAGTTGCTCTTCCGATCTAGTAGTGG
Lhin_014_F_M13	TGTAAAACGACGGCCAGTAGCCGTTGTTGTGACTCTTC
Lhin_015_F_M13	TGTAAAACGACGGCCAGTTTGGTGGCGCTAATGGTAAC
Lhin_016_F_M13	TGTAAAACGACGGCCAGTCTTGGTGTTACATTTCTGGCTC

Lhin_017_F_M13	TGTAAAACGACGGCCAGTTGGTAGTGGAGAAGCTGCTG
Lhin_018_F_M13	TGTAAAACGACGGCCAGTTTTAAAGAGCGGAGGGTATTCC
Lhin_019_F_M13	TGTAAAACGACGGCCAGTAAACACTGGTATGAGGACAGC
Lhin_020_F_M13	TGTAAAACGACGGCCAGTCCTTTGTTGGTTGGTGGGTC
Lhin_021_F_M13	TGTAAAACGACGGCCAGTAGCGAGTGAGTGGAGTTCAG
Lhin_022_F_M13	TGTAAAACGACGGCCAGTTTGGTGGCTTTAAACAAACCG
Lhin_023_F_M13	TGTAAAACGACGGCCAGTGCCTGTGTTGCCAGTAAGTC
Lhin_024_F_M13	TGTAAAACGACGGCCAGTTACTCGCTCTCTCTTGGACC
Lhin_025_F_M13	TGTAAAACGACGGCCAGTGTGCACGCGGTTTAGACTTC
Lhin_026_F_M13	TGTAAAACGACGGCCAGTTGTGTGATGTCCAATGAAATGC
Lhin_027_F_M13	TGTAAAACGACGGCCAGTAGCCGTTTGAAGTTTACAAGC
Lhin_028_F_M13	TGTAAAACGACGGCCAGTGGCTCTGCGACAAATCCAC
Lhin_001_R_M13	ATGCGTGTAAATCCTGTGCG
Lhin_002_R_M13	TGGGCTTTCGTCTGGAGTTG
Lhin_003_R_M13	ATCACTTTGGCGGAACAGTG
Lhin_004_R_M13	GGGTGGTACGAACAGACTGG
Lhin_005_R_M13	CTGACAGCCTTTCAGACAGAC
Lhin_006_R_M13	GAAAGGATGTTGCGTGTTTGC
Lhin_007_R_M13	ACAGACAGAAATATAGCCAGGC

Lhin_008_R_M13	AGGGAACACAGTCAAACAACC
Lhin_009_R_M13	GGGCTAAGTCGTGGCTTTAC
Lhin_010_R_M13	AATGAAGGCCGCTAAACAGG
Lhin_011_R_M13	AATGACGTTACTGCTGCTCG
Lhin_012_R_M13	ACAGAGCAATGTAACAAACTGC
Lhin_013_R_M13	ACAGTGAAATATTGGGTGGTGC
Lhin_014_R_M13	GAAGCAAAGCAAATATCAGCCC
Lhin_015_R_M13	TGCTCTTCCGATCTGCTACC
Lhin_016_R_M13	TTATGATGGTTGCGTTGGAATG
Lhin_017_R_M13	CTACTACCATCCACGCCTCC
Lhin_018_R_M13	GAACTCACGAACTCTGGCTTC
Lhin_019_R_M13	TGTTGGTATGTGTTCGATCGTG
Lhin_020_R_M13	GAGTGAGCAAACCACAGTCG
Lhin_021_R_M13	CTCTCTTCGGTTGCCCTTTC
Lhin_022_R_M13	CGCGCGGGAACATTATTATTC
Lhin_023_R_M13	AGGCAGAGCACATTTGGAAC
Lhin_024_R_M13	AGATGAGGGCCTAGGGAGAG
Lhin_025_R_M13	CCTTCCTCTCTTCTCCGAAAC
Lhin_026_R_M13	TGGGACGGTAAGTTCGACAC

Lhin_027_R_M13	CTCTCCCGATCTCACACATATG
Lhin_028_R_M13	GCTCTCTCTCGTTTCCCATAG

Appendix E: Microsatellite Dataset

Table E.1: Microsatellite dataset for Dunedin (WD=Warrington, PMLD, PSBD, QID=Portobello, WRD=Weller's Rock).

Population	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
Dunedin	Locus 1		Locus 2		Locus 3		Locus 4		Locus 5	
WD1	0	0	337	453	0	0	349	431	186	186
WD3	185	185	337	453	0	0	413	413	0	0
WD5	0	0	0	0	0	0	413	418	170	170
WD7	0	0	0	0	379	379	0	0	170	237
PMLD2	188	188	337	453	383	383	426	426	180	180
PMLD3	185	185	337	453	379	379	397	418	159	159
PMLD4	0	0	337	453	379	379	426	426	167	167
PMLD5	188	188	337	453	383	383	426	426	180	180
PMLD6	185	185	337	453	379	379	426	426	159	159
PMLD7	185	185	337	453	379	379	0	0	167	167
PSBD2	167	167	0	0	383	383	0	0	180	180
PSBD3	0	0	337	337	379	379	426	426	159	159
PSBD4	0	0	337	453	379	379	353	353	186	186
PSBD5	0	0	0	0	383	383	426	453	180	180

QID2	0	0	453	453	0	0	413	426	159	159
QID4	185	185	337	337	379	379	431	431	159	237
QID5	0	0	337	453	379	379	413	426	159	159
WRD2	167	167	337	453	379	379	431	431	170	170
WRD3	167	167	0	0	383	383	0	0	0	0
WRD4	0	0	0	0	379	379	418	422	0	0
WRD12	0	0	337	453	379	379	353	456	167	167
WRD13	167	167	337	453	379	379	0	0	159	180
WRD61	0	0	0	0	0	0	413	422	159	159