Detection of the bacterial endosymbiont *Wolbachia* and determination of super groups present within New Zealand invertebrates.

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

Masters of Science in Zoology
Massey University
Palmerston North, New Zealand

Benjamin James Bridgeman

2016

Synopsis

Wolbachia is one of the most widespread intracellular bacterium on earth, estimated to infect between 40 and 66% of arthropod species. Where significantly screened for, there is virtually no ecosystem that *Wolbachia* has not managed to invade. Their impact does not come solely from their vast distribution but in their ability to modify their hosts reproductive biology. *Wolbachia* is a maternally inherited endosymbiont that can induce a range of host phenotypic responses, including cytoplasmic incompatibility, male death, feminization, and parthenogenesis. This holds high potential for influencing genetic diversity and speciation of its host. *Wolbachia* has yet to be formally identified in New Zealand native invertebrates and therefore a gap remains in the global understanding of *Wolbachia* distribution and diversity. The first aim of this thesis (Chapter 2) was to determine if the use of high throughput sequencing (HTS) of invertebrates could be used to identify *Wolbachia* sequences and establish a *Wolbachia* infection. seven HTS produced a positive indication for *Wolbachia* sequences, with six samples originating from native New Zealand invertebrate hosts. Once *Wolbachia* was detected the second aim (Chapter 2) was to determine which super group the *Wolbachia* strains detected fall into. Comparing New Zealand *Wolbachia* sequences to sequences obtained from GenBank, it was determined that there were two distinct strains of *Wolbachia* in New Zealand hosts. One strain was related to *Wolbachia* super group A and the other matched *Wolbachia* super group B. *Wolbachia* had now been detected in New Zealand, the next aim (Chapter 3) was to determine the presence and distribution of *Wolbachia* across New Zealand and across a number of native invertebrates. *Wolbachia* was detected in both of the main islands of New Zealand across a number of species or putative species of ground and cave weta. The final aim of this thesis was to determine the diversity of the *Wolbachia* detected and compare that to the diversity of the ground weta complex *Hemiandrus maculifrons* to determine if there was any evidence for *Wolbachia* affecting the genetic diversity of its host (Chapter 4). It was observed that *H. maculifrons* complex was infected with both strains of *Wolbachia* detected in New Zealand invertebrates. There was however not enough resolution to determine if *Wolbachia* has had a significant effect on the diversity and speciation of the ground weta host.

Acknowledgements

I would like to thank my supervisors, Prof. Steve Trewick, Assoc. Prof Mary Morgan-Richards, and Dr Dave Wheeler, for their expertise and guidance during the period of this thesis. I am extremely grateful for your continuous support and encouragement throughout this work and look forward to working with you in future ventures.

I would also like to express my gratitude to Dr Briar Smith for the assistance in understanding the diversity of New Zealand Ground weta which became a large aspect of this study. I would also like to thank her for the access to raw samples, extracted DNA samples, sequence data, and access to unpublished data which this thesis would not have been feasible without.

I would like to extended my thanks to John Early for the supplying of samples of the parasitoid wasp *Archaeoteleia* to be tested as a potential vector. These individuals provided interesting aspects to the understanding of the movement of *Wolbachia* and provide a potential avenue for a more detailed investigation into vectors in New Zealand.

Contents

List of Figures

List of Tables

Chapter 1

Intercellular bacterium of arthropods

Chapter 1: Intracellular bacterium of arthropods

1.1 Introduction

The bacterium *Wolbachia* (Hertig & Wolbach, 1924; Hertig, 1936) is estimated to infect between 40 and 66% of arthropod species (Hilgenboecker, et al., 2008; Morrow, et al., 2014) making it among the most abundant intracellular bacterial genus. *Wolbachia* is a maternally inherited endosymbiont that can induce a range of host phenotypic responses, including cytoplasmic incompatibility, male death, feminization, and parthenogenesis (Hoffmann, et al., 1996; Hurst & Jiggins, 2000; Breeuwer & Werren, 1993; Rokas, et al., 2002; Werren & Windsor, 2000). *Wolbachia* infections can therefore have long-term evolutionary effects on their host lineages, in addition to immediate reproductive modifications, by providing a pathway to rapid speciation and influencing the evolution of sex-determining mechanisms (Hoffmann, et al., 1996; Hurst & Jiggins, 2000; Rokas, et al., 2002; Werren & Windsor, 2000). A better understanding of the mechanisms underlying *Wolbachia* induced changes in their hosts could lead to novel pest and disease control strategies (Brelsfoard & Dobson, 2009; McMeniman, et al., 2009) as well as a better understanding of factors contributing to speciation. This is particularly important because *Wolbachia*-mediated mating incompatibilities can create reproductive barriers in sympatric populations, thereby accelerating speciation in invertebrate hosts (Werren, et al., 2008). Induction of sperm-egg incompatibility (CI) between diverging populations could drive the evolution of new species (Bordenstein, et al., 2001; Werren, et al., 2008) and so provide a mechanism for speciation in circumstances, such as sympatric speciation events, that are currently poorly understood (Smith & Cornell, 1979).

Interactions between *Wolbachia* and the parasitoid wasp *Nasonia vitripennis (Nasonia)* have been extensively studied with CI being the mode by which *Wolbachia* maintains high infection frequencies in this host. This holds the potential to be a mechanism of speciation consistent with the Bateson-Dobzhansky-Muller model. Infection by CI-inducing *Wolbachia* has been shown to precede the evolution of post-mating reproductive barriers in two closely related parasitic wasps *Nasonia giraulti* and *Nasonia longicornis*. (Bordenstein, et al., 2001) Removal of *Wolbachia* infection via antibiotics results in the production of hybrids, with F1 and F2 hybrids completely viable and fertile (Bordenstein, et al., 2001) (Dedeine, et al., 2001). *Nasonia giraulti* and *N. longicornis* have been

determined to be sister species with a third species *N. vitripennis* diverging much earlier. Previous studies testing *Wolbachia*-induced incompatibility between *N. vitripennis* and *N. giraulti* (Breeuwer & Werren, 1990) showed that hybrids were not formed unless *Wolbachia* was removed, similar to interactions between *N. giraulti* and *N. longicornis*. However, several other isolating barriers exist between these species including F2 hybrid lethality, behavioural sterility, and partial premating isolation. These may reflect the longer period of separation (based on ITS2 sequence substitutions) between *N. vitripennis* and the other two species (Campbell, et al., 1993). Initial, reproductive barriers induced by *Wolbachia* may provide the isolation needed for Bateson-Dobzhansky-Muller incompatibilities to accumulate and keep the species reproductively separated in the absence of their respective *Wolbachia* infections.

The method by which *Wolbachia* moves between species has yet to be identified, however, similarity between the *Wolbachia* within parasitoids and the *Wolbachia* within the parasitoids' host (Heath, et al., 1999; Vavre, et al., 1999; Werren, et al., 1995) suggest horizontal transmission of *Wolbachia*. It has been shown that microinjection of *Wolbachia* infected cells can facilitate the transfer of *Wolbachia* (Watanabe, et al., 2013). This provides opportunity for movement of *Wolbachia* via parasitoids introducing infected tissue during oviposition. Should the egg fail to develop *Wolbachia* would be required to move in to the parasitoids' host to continue into further generations or alternatively be transferred through the digestive system of invertebrates feeding on *Wolbachia* infected hosts.

The type species of *Wolbachia* is *Wolbachia pipientis* (Hertig, 1936), first described in the mosquito *Culex pipiens* (Hertig & Wolbach, 1924). *Wolbachia* spp. have since been divided into eight monophyletic super groups (A-H) based on DNA sequences of 16S ribosomal RNA and ftsZ regions of genomic DNA(Werren, et al., 2008) (Lo, et al., 2002). Extensive recombination between super groups complicates interpretation of the evolutionary relationships of the groups and the question of whether all bacteria within the *Wolbachia* Clade should be given the *W. pipientis* designation or whether a different species nomenclature should be applied has been debated. Until this is resolved the convention is to refer to the bacteria as *Wolbachia*, with strain designation that is based on host and super group identification (Werren, et al., 2008). Super groups C and D are commonly found within filarial nematodes, whereas the other six super groups are found primarily in arthropods.

To date no *Wolbachia* infections have been reported from New Zealand native invertebrate species and the presence of *Wolbachia* in New Zealand has not previously been formally investigated. The present study aimed to help fill this gap in the international understanding of *Wolbachia* distribution by providing a preliminary data on the presence of *Wolbachia* among a sample of native and introduced invertebrate species found in New Zealand. Two different approaches were employed to survey potential hosts for *Wolbachia* infection bioinformatics and molecular ecology.

1.2 Thesis plan

1.2.1 Chapter 2: Using high-throughput DNA data sets to investigate whether NZ endemic arthropods are infected by Wolbachia

The use of High Throughput Sequencing datasets was used to search for *Wolbachia* sequences using PAUDA, a pseudo BLASTX algorithm (Huson & Xie, 2013). Data samples post PAUDA were analysed using MEGAN (Huson, et al., 2011), a metagenomic program. Sequences identified as *Wolbachia* were extracted from the data set and analysed against a reference *Wolbachia* genome to determine coverage, quality, and if fragments representing MLST regions are present.

1.2.2 Chapter 3: The distribution and abundance of Wolbachia infections in New Zealand ground weta (Hemiandrus spp.) and cave weta (Rhaphidophoridae)

Reporting of the use of Polymerase Chain Reaction (PCR) with standard *Wolbachia* Multi Locus Sequence Typing (MLST) (Baldo, et al., 2006) primers to test for *Wolbachia* infection in a range of New Zealand invertebrates. Species will be assessed for infection, and if infection is detected, the level of infection will be determined. The geographic origins of host individuals and their infection status were input into QGIS (QGIS Development Team, 2015) to determine if any spatial patterning is apparent within species or New Zealand as a whole.

1.2.3 Chapter 4: Distribution of Wolbachia strains and genetic diversity of their hosts in New Zealand

DNA sequences for one of the MLST loci (ftsZ) from individuals that gave positive PCR results for *Wolbachia* (Chapter 3). The number of strains and their distribution within host species was determined using phylogenetic analysis. Genetic diversity of the New Zealand

Wolbachia was analysed and phylogenetic placement among global *Wolbachia* determined. The diversity of the *Wolbachia* was compared to the genetic diversity of host lineages to seed evidence for any pattern that might have resulted from the *Wolbachia* effecting the hosts genetic diversity.

1.3 References

Baldo, L. et al., 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Applied and Environmental Microbiology, Volume 76(11), pp. 7098–7110.

Bordenstein, S. R., O'Hara, F. P. & Werren, J. H., 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature, Volume 409, pp. 707-710.

Breeuwer, J. A. J. & Werren, J. H., 1990. Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. Nature, Volume 346, pp. 558-560.

Breeuwer, J. A. J. & Werren, J. H., 1993. Cytoplasmic incompatibility and bacterial density in *Nasonia vitripenni*. Genetics, Volume 135, pp. 565-574.

Brelsfoard, C. L. & Dobson, S. L., 2009. *Wolbachia*-based strategies to control insect pests and disease vectors. Asia-Pacific Journal of Molecular Biology and Biotechnology, Volume 17(3), pp. 55-63.

Campbell, B. C., Steffen-Campbell, J. D. & Werren, J. H., 1993. Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28s rDNA sequences. Insect Molecular Biology, Volume 2(4), pp. 225-237.

Dedeine, F. et al., 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. PNAS, Volume 98(11), pp. 6247–6252.

Heath, B., Butcher, R., Whitfield, W. & Hubbard, S., 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Current. Biology, Volume 9, pp. 313-316.

Hertig, M., 1936. The Rickettsia, *Wolbachia pipientis* (gen. et sp.n.) and Associated Inclusions of the Mosquito, *Culex pipiens*. Parasitology, Volume 28(4), pp. 453-486.

Hertig, M. & Wolbach, S. B., 1924. Studies on *rickettsia*-like microorganism in insects. Journal of Medical Research, Volume 44(3), pp. 329-374.

Hilgenboecker, K. et al., 2008. How many species are infected with *Wolbachia*? – a statistical analysis of current data. FEMS Microbiology Letters, Volume 281, pp. 215–220.

Hoffmann, A. A., Clancy, D. & Duncan, J., 1996. Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity, Volume 76, pp. 1-8.

Hurst, G. D. & Jiggins, F. M., 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. Emerging Infectious Diseases, Volume 6, pp. 329–336.

Huson, D. H. & Xie, C., 2013. A poor man's BLASTX—high-throughput metagenomic protein database search using PAUDA. Bioinformatics, Volume 30(1), pp. 38-39.

Huson, D. et al., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Research, Volume 21, pp. 1552-1560.

Lo, N. et al., 2002. How many *Wolbachia* supergroups exist?. Molecular Biology and Evolution, Volume 19, pp. 341-346.

McMeniman, C. J. et al., 2009. Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. Science, Volume 323, pp. 141-144.

Morrow, J. L., Frommer, M., Shearman, D. C. A. & Riegler, M., 2014. Tropical tephritid fruit fly community with high incidence of shared *Wolbachia* strains as platform for horizontal transmission of endosymbionts. Environmental Microbiology, Volume 16(12), pp. 3622–3637.

QGIS Development Team, 2015. QGIS Geographic Information System

Rokas, A. et al., 2002. The incidence and diversity of *Wolbachia* in gall wasps (Hymenoptera; Cynipidae) on oak. Molecular Ecology, Volume 11(9), pp. 1815-1829.

Smith, M. A. & Cornell, H. V., 1979. Hopkins host-selection in *Nasonza Vitripennis* and its implications for sympatric speciation. Animal Behaviour, Volume 27, pp. 365-370.

Vavre, F. et al., 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. Molecular Biology and Evolution, Volume 16(12), pp. 1711-1723.

Watanabe, M., Kageyama, D. & Miura, K., 2013. Transfer of a parthenogenesis-inducing *Wolbachia* endosymbiont derived from *Trichogramma dendrolimi* into *Trichogramma evanescens*. Journal of Invertebrate Pathology, Volume 112(1), pp. 83-87.

Werren, J. H., Clark, M. E. & Baldo, L., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature, Volume 6, pp. 741-751.

Werren, J. H. & Windsor, D. M., 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium?. Proceedings of the Royal Society of London B, Volume 267(1450), pp. 1277-1285.

Werren, J. H., Zhang, W. & Guo, L., 1995. Evolution of *Wolbachia*: reproductive parasites of arthropods Proceedings of the Royal Society of London B, Volume 251, pp. 55-63.

Chapter 2

Using high-throughput DNA data sets to investigate whether NZ endemic arthropods and are infected by *Wolbachia*

Chapter 2: Using high-throughput DNA data sets to investigate whether NZ endemic arthropods are infected by *Wolbachia*

2.0 Abstract

High throughput sequencing (HTS) has provided an exponential amount of data for the molecular biology field. To analysis the vast amount of data the use of computer processing and the development of Bioinformatics was required. HTS is often used for the construction of genomes (nuclear and mitochondrial) however it will also pick up sequences from all sources of DNA included in the extraction. This includes infections such as the endoparasite *Wolbachia*. PAUDA and MEGAN were used together to determine the sources for the DNA sequenced and determine if there was *Wolbachia* DNA sequenced. *Wolbachia* was detected in seven individuals using HTS raw data (*Klapopteryx kuscheli*, *Macropathus sp*, *Hemiandrus sp*, *Talitropsis edilloti*, *Miotopus sp*, and two *Neonetus sp)*, six of which are native to New Zealand. These are the first known examples of *Wolbachia* infections in New Zealand native invertebrates. *Wolbachia* sequences were extracted from the HTS data and compared to GenBank samples indicating the presence of both super group A and super group B in New Zealand as well as the detection of super group E in the Chillan *K. kuscheli*.

Key words

Bioinformatics, high throughput sequencing

Abbreviations

High throughput sequencing = HTS

2.1 Introduction

Despite the potential for dramatic effects on the population dynamics of *Wolbachia* infected species, we currently have no data available for the prevalence of *Wolbachia* in New Zealand endemic species. This lack of knowledge has implications in understanding the current distribution of species and their evolutionary history. From a practical prospective *Wolbachia* can also introduce incompatibilities between individuals seemingly of the same species thus complicating conservation based breeding programs.

Phylogenetic studies have identified eight globally distributed super groups of *Wolbachia* (Lo, et al., 2002; Werren, et al., 2008). Incongruence between the *Wolbachia* and host phylogenies suggest many strains have been transferred horizontally and thus unrelated hosts in the same region can share similar strains of *Wolbachia*. However, phylogenetic relationships are also complicated by genetic exchanges between *Wolbachia* strains (Werren, 1997; Lo, et al., 2002; Jiggins, et al., 2001) and host – parasite coevolution (Casiraghi, et al., 2005). For this reason, Baldo et al (2006) developed a Multilocus Sequence Typing (MLST) system that allows differentiation between even closely related strains of *Wolbachia*. This core set of loci target five genes, ftsZ, coxA, fpbA, hpcA, and gatB, have been used to provide molecular data that allows distinction between the different *Wolbachia* super groups.

2.1.1 DNA sequencing technology

The large-scale, broad-scope biosystematics projects such as the barcode of life initiative (Huson, et al., 2011) have mostly relied on sanger sequencing technologies. However, traditional sequencing has many drawbacks when it comes to analysing whole genomes or multiple genomes. It has relatively low throughput (Shokralla, et al., 2012) and this makes it very expensive and time consuming to sequence whole genomes or sequence DNA from all organisms in a sample (Shokralla, et al., 2012). Since its origin in 2005, with the introduction of the Roche454 Genome Sequencer, High Throughput Sequencing (HTS) has become the dominant methodology used in molecular biosciences. The major difference between traditional sequencing and HTS is the ability to process millions of sequence reads in parallel (Mardis, 2007). This feature of HTS allows the acquisition of a large number of sequences across all organisms found within a sample. Not only can the reads be used to identify species the read counts provide a measure of the abundance of the different organisms in the sample.

The HiSeq platform (Illumina, Inc, 2010) was introduced in 2010 and has become the most widely used HTS approach. Illumina sequencing uses a stable, reversible terminator sequencing by synthesis (SBS) method (BGI, 2014). SBS is the most widely adopted HTS technology; it incorporates the use of four fluorescently-labelled nucleotides to sequence DNA located on millions of clusters bound to a glass slide. During each cycle, a single labelled deoxynucleotide triphosphate (dNTP) is added to the nucleic acid chain. The nucleotide label serves as a terminator for polymerization as well as a fluorescent label that is imaged at the end of each cycle in order to identify the incorporated base. Enzymatic cleavage of the dye allows the incorporation of the next nucleotide and the imaging cycle is repeated (Illumina, Inc, 2010). Early versions of this technology produced DNA sequences of only 30 bp, but through improvements to the sequencing chemistry read lengths have increased to 300 bp with the MiSeq series (Illumina, Inc, 2010).

2.1.2 Tools for metagenomics

In the field of metagenomics, millions of DNA or cDNA reads are sequenced from environmental samples and these are then analysed in an attempt to determine the functional or taxonomic content of a sample (Handelsman, 1998). Two popular approaches are rRNA amplicon sequencing and whole genome metagenomics. Amplicon sequencing involves the amplification of the rDNA sequences in a sample using universal primers followed by shotgun library preparation and HTS. This targeted nature of amplicon sequencing maximises the number of taxa that be identified in a single experiment and this can be further refined through the use of specialised rDNA primers sets that target a particular group of organisms (e.g. 16S region of bacteria and archaea). A drawback of this approach is that classifications based on a single loci provide limited information on the genomic makeup of the organisms found in the sample. In contrast, whole genome metagenomics approach randomly sequences all genomic regions found in the sample. As multiple genes can be identified better predictions can be made about the functional composition of the microbial community being studied. Another limitation of amplicon sequencing is that preferential binding of the primers to different rDNA genes can also bias the species classifications and lead to underestimation of the true taxonomic diversity of the sample. Thousands of low-abundance taxa account for most of the observed phylogenetic diversity in any environment. This ‘‘rare biosphere’’ contains a large amount of phylogenetic diversity and represents an enormous contribution to genetic distinctiveness and evolutionary innovation (Sogin, et al., 2006).

The development of new methods that use HTS to investigate the distribution and abundance of microorganisms is advancing rapidly as the cost of sequencing large amounts of DNA continues to decline. One approach employed in the sorting of the millions of sequence reads in to their corresponding taxonomic sources is interrogation of protein database searches. The Basic Local Alignment Search Tool (BLASTX) translates each nucleotide sequence query into all six possible amino acid reading frames and compares these to existing protein sequences on the BLAST database (Altschul, et al., 1990). One limitation of this approach is that searching millions of reads against the NCBI database is computationally expensive and time consuming. Traditional BLASTX searches require an extensive amount of available memory and CPU hours to process a typical environmental sample. To reduce the time and resources required to analyse metagenomic samples, alternative methods are being employed to produce BLASTX-like alignments such as RAPSearch2 (Zhao, et al., 2012) and PAUDA (Huson & Xie, 2013).

The BLASTX-like aligner PAUDA is a protein search tool that produces alignments up to 10,000 times faster than the traditional BLASTX algorithm (Huson & Xie, 2013). These speed improvements are achieved by converting all proteins sequences in the database to 'pseudo DNA' or 'pDNA' for short. The pDNA is generated by converting the standard '21 amino acid' alphabet into a four-lettered alphabet, this is completed by grouping amino acids that are likely to replace each other into one of the four new groups as determined in significant BLASTX alignments (Huson & Xie, 2013). As there are now only four potential options at each position large data sets can be processed in a computationally efficient way. Although PAUDA requires less computational resources it is typically only able to taxonomically assign one third as many reads as BLASTX analyses run to completion. (Huson & Xie, 2013).

MEGAN (Huson, et al., 2011) is a metagenomic program designed to efficiently isolate and separate the sequences by organism as determined by its corresponding GenBank match. MEGAN takes the file produced by PAUDA and creates a taxonomic distribution indicating what species' sequences matched those in the raw sequence file and the number of sequences associated with that species. This can be visualised a number of ways in addition to the default taxonomic distribution. Outputs such as word clouds (Fig. 2.1) and bar graphs (Fig. 2.2) provide representations of the most common species identified in the dataset. Once sequences have been assigned to the level of species, genus, or family, they can be extracted and further analysed to determine the robustness of the assignments.

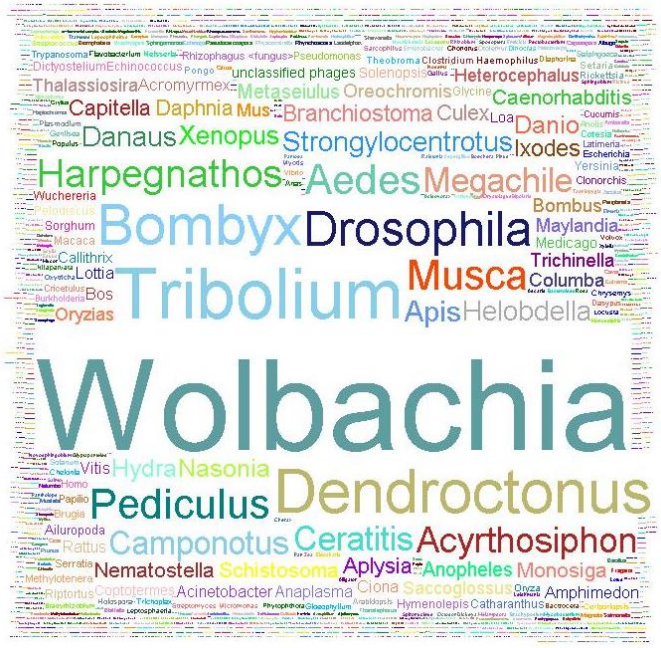


Figure 2.1. Word cloud representation of taxa detected for *Klapopteryx kuscheli*, size of name correlates to number of matching sequences

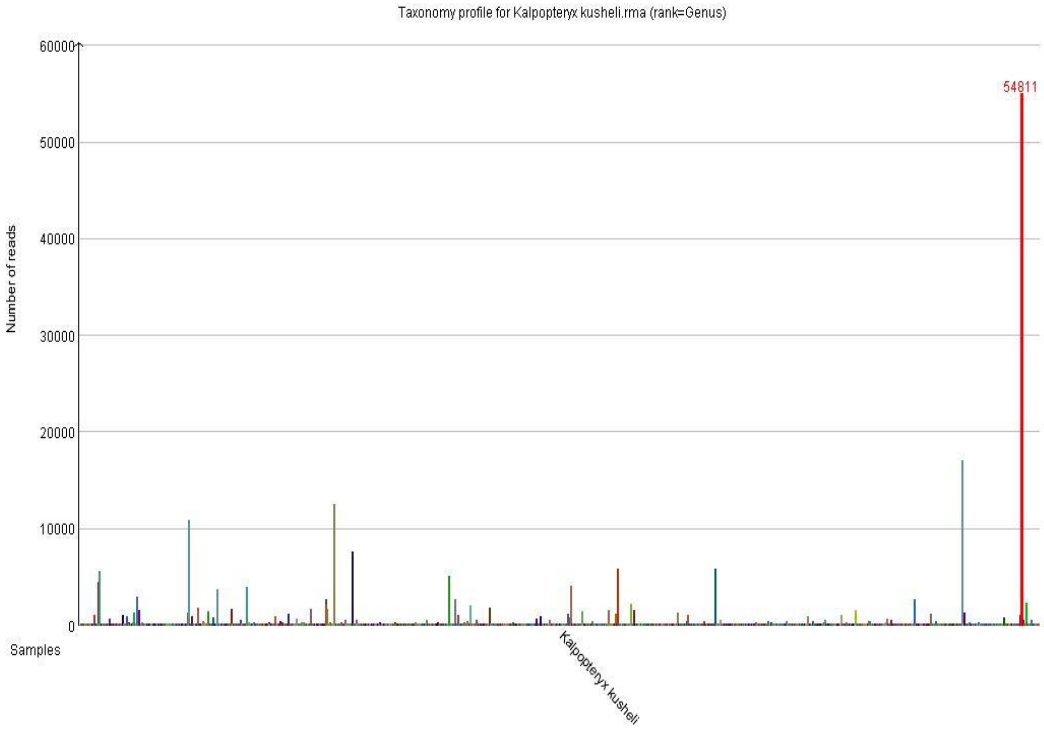


Figure 2.2. Bar graph representation of taxa detected for *Klapopteryx kuscheli*, *Wolbachia* highlighted

Aim 1

To investigate the prevalence of *Wolbachia* in a range of invertebrate hosts (Table 2.1) and to determine the level of coverage of any individuals positive for *Wolbachia*. Samples were

selected from a selection of previously collected datasets of New Zealand and foreign invertebrates and molluscs.

Aim 2

To place New Zealand *Wolbachia* diversity into the global picture and which *Wolbachia* super groups are present in New Zealand.

2.2 Methods

2.3.1 Bioinformatics

High through-put DNA sequences were generated from invertebrates, either whole body or tissue extracted DNA, using an Illumina HiSeq2000 (BGI, 2014). For each host approximately 1-4 Gigabytes of data was produced in each direction in the form of 100bp sequence reads (Table 2.1).

Table 2.1 HTS samples used with location, number of sequences matching *Wolbachia*, and the relative abundance of *Wolbachia* among genera detected (Rank) (Where available)

Class/Order	*HTS Specimen*	Location	Reads	Rank
Plecoptera	*Klapopteryx kuscheli*	Patagonia, Chile	54811	1
Orthoptera	*Hemiandrus 'bruce'*	South Island, New Zealand	17220	2
Orthoptera	*Macropathus sp.*	Masons Dry Cave, Hawks Bay	30817	1
Orthoptera	*Neonetus sp.1*	Mohi Bush, Hawks Bay	1363	14
Orthoptera	*Neonetus sp.2*	Hongi's Track, Rotorua	1346	16
Orthoptera	*Talitropsis edilloti*	Mohi Bush, Hawks Bay	2486	4
Orthoptera	*Miotopus sp.*	Waioeka Gorge, Gisborne	2384	9
Gastropoda	*Cominella virgate*	Spirits Bay, Aupouri Peninsula	0	na
Orthoptera	*Sigaus australis*	South Island	0	na
Orthoptera	*Hemiandrus crassidens*	South Island	20	na
Gastropoda	*Pelicaria vermis*	New Zealand	0	na
Gastropoda	*Alcithoe fusus*	New Zealand	7	na
Gastropoda	*Austrolittorina cincita*	New Zealand	24	na
Gastropoda	*Amalda australis*	New Zealand	17	na
Orthoptera	Cave weta (In process of identification)	Kapiti Island	47	na
Orthoptera	*Cratomelus armatus*	Oncol Park, Chile	0	na
Coleoptera	*Carcinops sp.*	Col d'Amieu, New Caledonia	24	na
Gastropoda	*Placostylus pophyrostomus*	Nekoro, New Caledonia	10	na
Gastropoda	*Cominella adspersa*	New Zealand	8	na
Orthoptera	*Noroplectron serratum*	Chatham Island	0	na
Orthoptera	*macropathus sp.*	German terrace mines, Westport	25	na

	Pharmacus chapman	Old Man Range, Otago	17	na
Orthoptera	*Isoplectron sp.*	Canterbury	154	na
Orthoptera	*Micropathus sp.*	Marakoopa, Tasmania	0	na
Orthoptera	CW1010	Balls Clearing	0	na
	Denniston unident	New Zealand	126	na
Orthoptera	*Hemiandrus focalis*	Lake Taupo	21	na
Orthoptera	ORT15	Solomons Island	10	na
	Onasandrus sp.	Mpunalanga Province, South Africa	17	na
Orthoptera	ORT145 Ground weta	cordillera la costa, Chile	103	na
Gastropoda	*Anotostoma sp.*	Queensland, Australia	12	na
Orthoptera	*Petalambon sp.*	Queensland, Australia	54	na
Orthoptera	*Exogryllus*	Queensland, Australia	36	na
Gastropoda	*Penion ormesi*	New Zealand	0	na
Gastropoda	*Amalda sp.*	New Zealand	0	na
Gastropoda	*Buccinulum vittatum*	New Zealand	0	na
Gastropoda	*Buccinulum vittatum littorinoides*	New Zealand	0	na
Gastropoda	*Amalda novaezelandiae*	Tuaranga	0	na
Gastropoda	*Amalda mucronata*	Island Bay, Wel	0	na
Gastropoda	*Amalda depressa*	Tuaranga	0	na
Gastropoda	*Penion benthicolus*	Chatham Rise	0	na
Gastropoda	*Penion chathamensis*	Chatham Rise	0	na
Gastropoda	*Penion cuvierianus*	Coromandel	0	na
Gastropoda	*Penion jeakingsi*	Tasman Bay	0	na
Gastropoda	*Penion sp.*	Three Kings Islands	0	na
Gastropoda	*Penion sulcatus*	Tuaranga	0	na
Gastropoda	*Buccinulum linea*	Castle Point	0	na
Gastropoda	*Buccinulum pallidum*	Stewart Island	0	na
Gastropoda	*Buccinulum pertinax finlayi*		0	na
Gastropoda	*Buccinulum robustrum*		0	na
Gastropoda	*Buccinulum fuscozonatum*		0	na
Gastropoda	*Austrofusus glans*	Island Bay, Wel	0	na
Gastropoda	*Aeneator elegans*	Chatham Rise	0	na
Gastropoda	*Aenetor benthicolus*	E of Cape Palliser	0	na
Gastropoda	*Aeneator recens*	SE of Cape Turnagain	0	na
Gastropoda	*Pararetifusus carinatus*	Chatham Rise	0	na
Gastropoda	*Alcithoe arabica*	Port William, StI	0	na
Orthoptera	*Transaevum laudatum*	Queensland, Australia	0	na
Orthoptera	*Cnemotettix sp.*	California, USA	0	na
Orthoptera	*Paterdecolyus genetrix*	Kyushu, Japan	0	na
Orthoptera	*Penalva flavocalceata*	Queensland, Australia	0	na
Gastropoda	*Penion cuvieranus*	Coromandel	0	na
Gastropoda	*Penion jeakingsi*	Tasman Bay	0	na
Gastropoda	*Penion mandarinus*	Eden-Gabo Island	0	na

Gastropoda	*Penion maximus*	New South Wales, Off Terrigal	0	na
Gastropoda	*Penion ormesi*	Cloudy Bay, off White Bluffs	0	na
Gastropoda	*Penion sulcatus*	Auckland?	0	na
Gastropoda	*Penion sp. 'West Coast'*	Kahurangi Point	0	na
Gastropoda	*Colus sp.*	Moray Firth, Scotland, UK	0	na
Gastropoda	*Buccinum undatum*	Gardskagi, Reykjanesskagi, Iceland	0	na
Gastropoda	*Volutopsius norwegicus*	Svalbard, Hornsund fjord, West Spitsbergen	0	na
Gastropoda	*Aeneator valedictus*		0	na
Gastropoda	*Aeneator otagoensis*	Tasman Bay	0	na
Gastropoda	*Potamopyrgus antipodarum*		0	na
Orthoptera	*Thoracic*		0	na
Orthoptera	*Ground*		0	na
Gastropoda	*Powelliphanta*		0	na
Gastropoda	*Lunella smaragdus*	Lottin Pt.	0	na
	Striracolpus pagoda	Tuaranga	0	na
Gastropoda	*Buccinulum linea*	Nelson	0	na
Orthoptera	*Crassidens*		0	na
Gastropoda	*Placostylus*		0	na
Orthoptera	*Tusk weta*		0	na
Gastropoda	*Struthiolaria papulosa*		0	na
Gastropoda	*Calliostoma simulans*	Western Chatham Rise	0	na
Gastropoda	*Notoacmea*	elongata	0	na
Coleoptera	*Hyperoides Oz*		0	na
Orthoptera	*Raspy cricket*		0	na
Orthoptera	*Lutosa BZ*		0	na
Orthoptera	*Hemiandrus*	Australia	0	na
Orthoptera	*Jerusalem cricket USA*		0	na
Orthoptera	*Hemiandrus fulica*		0	na
Gastropoda	*Haustrum haustorium*	East Cape	0	na
Gastropoda	*Cellana ornata*		0	na
Gastropoda	*Cellana denticulata*		0	na
Gastropoda	*Cookia sulcata*		0	na
Gastropoda	*Diloma aethiops*	Hicks/Onepoto Bays	0	na
Gastropoda	*Cantharidus tesselatus*	East Cape	0	na
Gastropoda	*Iredalina mirabilis*	Western Chatham Rise	0	na

HTS samples were processed through PAUDA (Huson & Xie, 2013)(Appendix A) and analysed using MEGAN5 (Huson, et al., 2011). This yielded taxonomic outputs showing all the species that matched sequences found within each HTS data set. MEGAN indicates the number of sequences associated with each species to determine the amount of each organism in the sample. If *Wolbachia* is in the top 16 of organisms detected (level genus)

then that invertebrate sample was treated as positive for an infection. Sequences from samples found to be positive for *Wolbachia* were converted from fastq format to fasta format using a script (Appendix B).

The *Wolbachia* sequences were then extracted from each HTS file (Appendix C).

The extracted *Wolbachia* sequences were mapped against published entire whole *Wolbachia* genome using the Geneious (version 6 http://www.geneious.com) (Kearse, et al., 2012) using the 'map to reference' function on medium sensitivity. *Wolbachia* endosymbionts of *Drosophila melanogaster* (NC002978, 1267782 bp) and *Culex quinquefasciatus Pel* (NC010981, 1482455 bp), were used as the reference in order to determine the coverage of the genome within the sample and what genes were being analysed.

To increase the coverage of the *Wolbachia* genome from each sample the whole HTS sequences were assembled into contigs of kmers ranging from 21 to 51. Due to the data size of the HTS samples (number of reads) it was not possible to produce the full range of contigs directly with the raw sample due to computational limitations; therefore, the samples were copied and treated with digital normalisation locally to reduce the size of the file. This first required the individual files to be interleaved into a single file (Appendix D), This sorts the files and allows the files to be interleaved (Appendix E). This aligns the lines from the two files so that the forward and reverse reads line up, adding 'inter' to the file name to signify the file has been interleaved.

Digital normalisation systematizes coverage in shotgun data sets, thereby decreasing sampling variation, discarding redundant data, and removing the majority of errors (Brown, et al., 2012). Digital normalization substantially reduces the size of shotgun data sets and decreases the memory and time requirements for sequence assembly, all without significantly impacting content of the generated contigs (Appendix F).

This produces five files e.g.

- inter_HTS_data_name.fq.keep
- inter_HTS_data_name.fq.keep.abundfilt
- inter_HTS_data_name.fq.keep.abundfilt.keep
- inter_HTS_data_name.fq.keep.abundfilt.keep.pe
- inter_HTS_data_name.fq.keep.abundfilt.keep.se.

Of these five files the .pe and .se files were used to generate an assembly using the Velvet assembler (Zerbino & Birney, 2008).

Normalized reads were assembled using Velvet to produce contigs using varying length kmers (Appendix G). These contigs were then compared to a *Wolbachia* database using BLASTN. Contigs matching *Wolbachia* with high sequence homology were extracted and added to the file containing the sequences extracted from the raw HTS as determined in MEGAN to be matching *Wolbachia*.

2.2.2 Phylogenetics

The combined dataset was then re-mapped using the Geneious (version 6 http://www.geneious.com) (Kearse, et al., 2012) map to reference function on medium sensitivity against the *Wolbachia* genomes.

Once sequences were aligned to the reference genome, the genes representing the MLST system (Baldo, et al., 2006) were identified by searching the gene annotation code matching the MLST region. The coverage at these loci was determined by identifying the primer locations sites on the reference gene, and viewing how much of the corresponding MLST region of the gene was covered by the HTS sequences. If there was full cover or only small gaps at either end of the MLST gene corresponding to the PCR products obtained available, the sequences were extracted and converted in to a single consensus sequence for Phylogenetic analysis. To determine the what super groups were detected, a MrBayes tree was run using the *Wolbachia* sequences combined with the example sequences used in Lo et al (2002) (Table 2.2).

2.2.3 Accuracy of PAUDA

PAUDA uses 'pDNA' to increase the speed of analysis and reduce computational resources, however as it reduces the protein 'alphabet' from 21 to 4 it can introduce higher chances of mismatching. This results in PAUDA not guaranteeing it will find the best match, to determine the accuracy of the PAUDA search, the putative *Wolbachia* sequences were used as input for BLASTN searches. The BLASTN output was separated in to two groups *Wolbachia* and other to determine the number of sequences that matched to *Wolbachia* between the two methods.

Table 2.2 GenBank samples used for super group identification.

GenBank sample (Host of *Wolbachia*)	GenBank ID
Diabrotica barberi clone	KC578107
Altica lythri isolate	KF163343.1
Pheidole vallicola	EU127749
Altica helianthemi	KF163366.1
Altica palustris	KF163363.1
Altica impressicollis	KF163368.1
Altica impressicollis	KF163367.1
Drosophila innubila	EU126333
Polistes dominulus	EU126353
Precis iphita	FJ392398.1
Jalmenus evagoras	FJ392417.1
Lissorhoptrus oryzophilus	DQ256473.1
Wolbachia sp.	AJ130717.1
Bombyx mandarina	KJ659910.1
Cydia fagiglandana	KJ140034
Bryobia kissophila	JN572863.1
Bryobia praetiosa	EU499322.1
Wolbachia pipientis	JN316217.1
Mesaphorura italica	AJ575103.1
Altica oleracea	KF163332.1
Melittobia digitata	EU170117.1
Altica oleracea	KF163325.1
Altica oleracea	KF163324.1
Serritermes serrifer	DQ837193.1
Cubitermes sp.	DQ127295.1

2.3 Results

Of the 100 High throughput samples 7 insects contained evidence for the presence of *Wolbachia* infection (Table 2.1) through the use of PAUDA (Huson & Xie, 2013) and MEGAN (Huson, et al., 2011). All samples tested had higher values when mapped against the *Drosophila melanogaster* (NC002978) endosymbiont than the *Culex quinquefasciatus Pel* (NC010981) endosymbiont and therefore this reference was used for all the subsequent observations described below.

The *Wolbachia* sequences were then mapped using the Geneious (version 6 http://www.geneious.com, (Kearse, et al., 2012) 'map to reference' against published whole genomes of *Wolbachia* endosymbionts of *Drosophila melanogaster* (NC002978) available via GenBank. The stonefly *Klapopteryx kuscheli* sample had the largest number

of *Wolbachia* sequences detected with 54,811 reads (using MEGAN). These *Wolbachia* DNA sequences covered 62% of the bacterial genome with pairwise similarity of 96% when mapped against *Wolbachia* infections of *Drosophila melanogaster* (NC002978).

Macropathus sp. and *Hemiandrus sp.* samples had similar levels of bacterial DNA (± <4%) when mapped against the endosymbiont *Wolbachia* genome from *Drosophila melanogaster* (NC002978): the sequences covered 30% and 33.6% of the *Wolbachia* genome respectively. That means the HTS data mapped to about 1/3 of the published *Wolbachia* genome (1267782BP). Their pairwise similarities were high: 90.2% and 92% respectively. Only one of the MLST genes (Table 2.4) had sufficient coverage to obtain sequence information from the HTS data for the infection of these two weta species. This was the ftsZ gene for the *Wolbachia* infecting *Macropathus sp.*

Table 2.3. Coverage of *Wolbachia* MLST regions from High throughput sequencing for each individual host positive for *Wolbachia*

Organism	gatB	coxA	hcpA	ftsZ	fpbA	wsp
Klapopteryx kuscheli	Missing 79bp section in middle	Full	Full	Full	Missing 92bp in middle	Site covered but large deletion gaps present
Macropathus sp.	Limited	No	No	Missing first 180bp	No	Missing 0-30 full after that
Hemiandrus 'bruce'	No	No	No	No	Limited	No
Talitropsis sp.	No	No	No	No	No	No
Miotopus sp.	No	No	No	No	No	No
Neonetus sp1	No	No	No	No	No	No
Neonetus sp2	No	No	No	No	No	No

Table 2.4. Analysis of MEGAN outputs rerun through BLASTN to determine accuracy of PAUDA results indicating number of reads associated to each group.

Host Organism	# *Wolbachia*	# *Vollenhovia*	# Other	# Not Associated	Total
Klapopteryx kuscheli	12561	1336	290	0	14187
Macropathus sp.	3747	108	81	3993	7929
Hemiandrus 'bruce'	4347	255	65	2	4669
Talitropsis edilloti	566	79	10	0	655
Miotopus sp.	544	29	8	0	581
Neonetus sp1	271	61	13	0	345
Neonetus sp2	141	29	4	178	352

DNA sequences from *Wolbachia* were detected in the DNA data sets from four other cave weta species but reads number were lower compared to the other three positive samples

(<2500 sequences) as determined by MEGAN (Table 2.1). This corresponded with low coverage <5% when mapped against both genomes, however the pairwise identical stayed above >90% for all samples excluding *Talitropsis edilloti*, who had higher coverage but significantly lower pairwise identical (Fig. 2.5).

The ftsZ region available for extraction from *Macropathus sp* was compared to corresponding regions from GenBank (Table 2.2). When the ftsZ fragment was analysed, this placed *Macropathus sp* within a group containing individuals limited to New Zealand including previously extracted sequences of *Hemiandrus maculifrons* GW816, *Hemiandrus* 'bruce' GW802, and *Ectopsocus* Pso1 (a Booklice) (Fig 2.4).

All five MLST genes were identified from the *K. kuscheli* alignment with three of the MLST genes having gapless coverage of the MLST region (coxA, hcpA, and ftsZ). One gene, fpbA, was missing a small section (92bp) that was complemented using PCR amplification of this region. This placed *K. kuscheli* separate from all but one sample on GenBank, a *Wolbachia* infected Collembola in France.

To determine the what super group, the individuals sequences a MrBayes tree was run using the example sequences used in Lo et al (2002) (Table 2.2). This showed that there were three different super groups detected. The clade containing *Macropathus sp* fell within super group B, the second clade detected in New Zealand was determined to fall within the super group A. The third super group was associated with *Klapopteryx kuscheli.* This sequences associated with super group E (Fig 2.3).

To determine the accuracy of the PAUDA (Huson & Xie, 2013), the results from PAUDA determined to *Wolbachia* (as shown by MEGAN) were extracted and processed with BLASTN. The resulting output was analysed (Appendix H) and the number of reads that were associated with each of the groups were determined. After the initial analysis there were a large number of reads associated with non-*Wolbachia* species, the majority of these reads were associating with the ant genus *Vollenhovia* (likely to be an undiagnosed *Wolbachia* infection). A third grouping was then established and the analysis of the BLASTN output was rerun producing three groups *Wolbachia, Vollenhovia,* and Other. The number of non-repeat reads was also compared to the number of reads in the output file from MEGAN (Huson, et al., 2011) to determine the proportion of reads that did not associate with any nucleotide sequences on GenBank. Table 2.4 indicates the number of reads from PAUDA that matched the correct target, *Wolbachia;* the ant genus *Vollenhovia;*

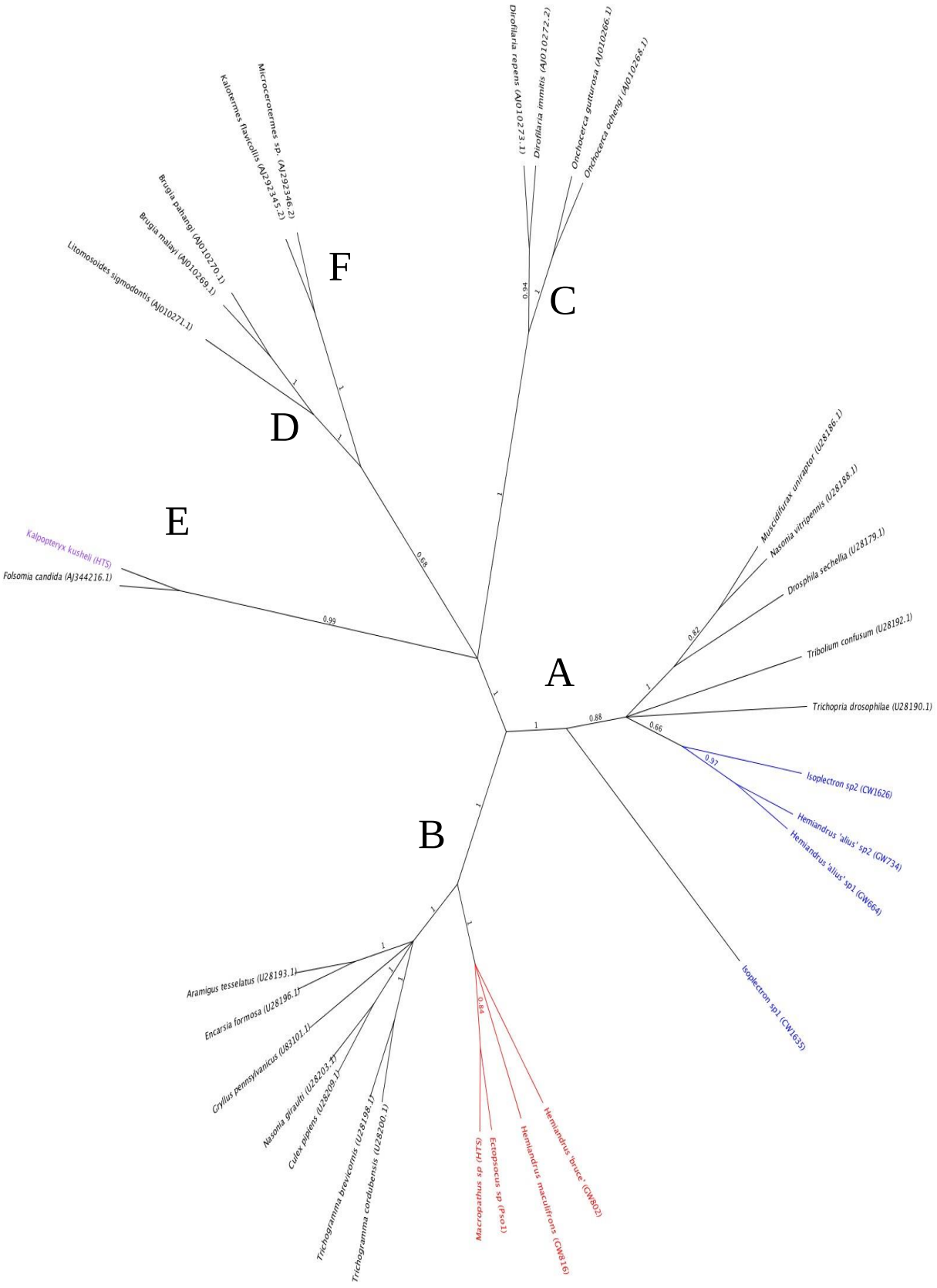


Figure 2.3 MrBayes tree of New Zealand and GenBank *Wolbachia* endosymbionts based on ftsZ sequences. Species names are those of the host. Super groups are indicated by the corresponding letter

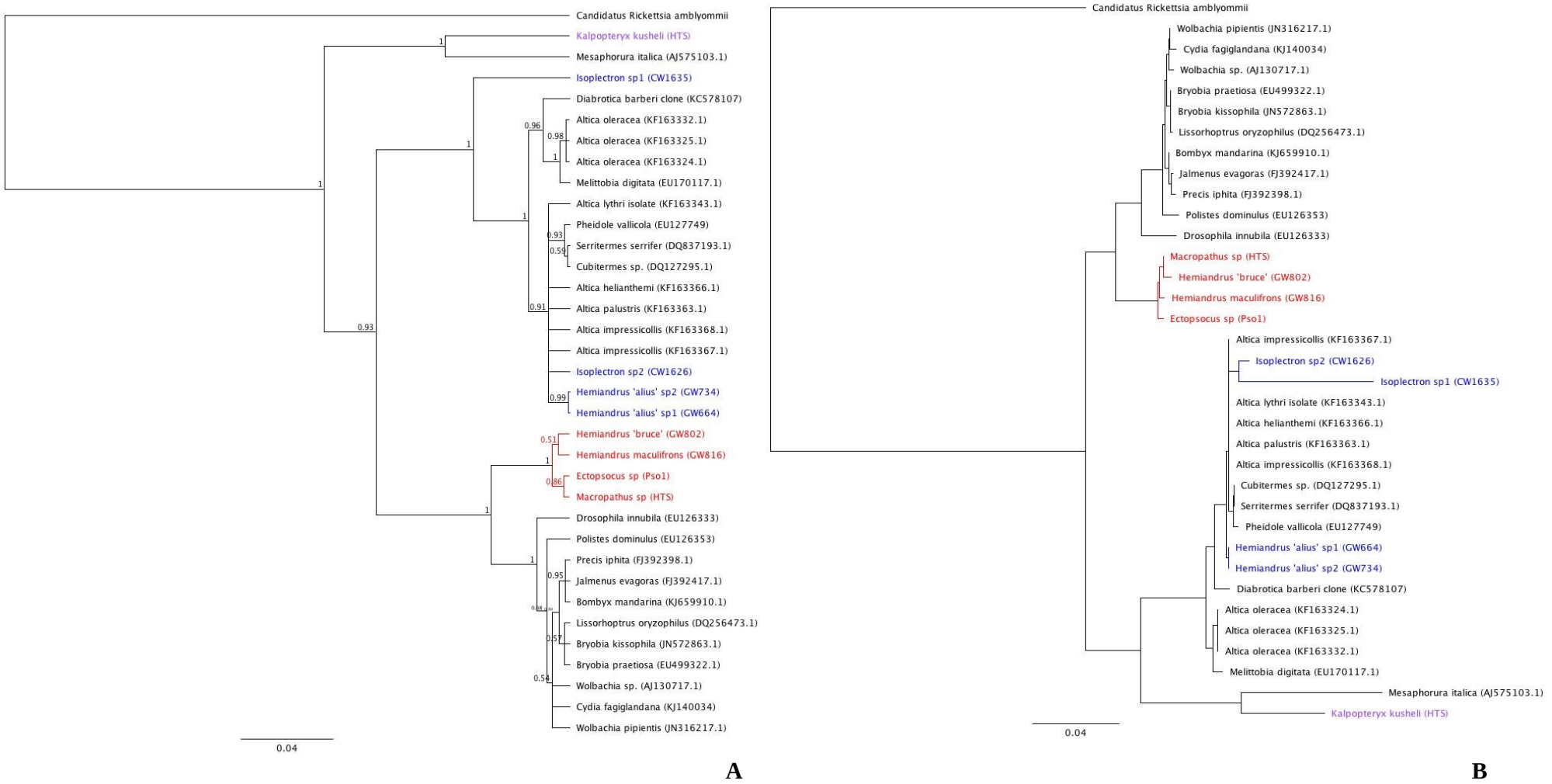


Figure 2.4 . MrBayes tree (A) and PHYML tree (B) of New Zealand and GenBank *Wolbachia* endosymbionts based on ftsZ sequences. Species names are those of the host. Strain one (super group A) is coloured blue, strain two (super group B) is coloured red.

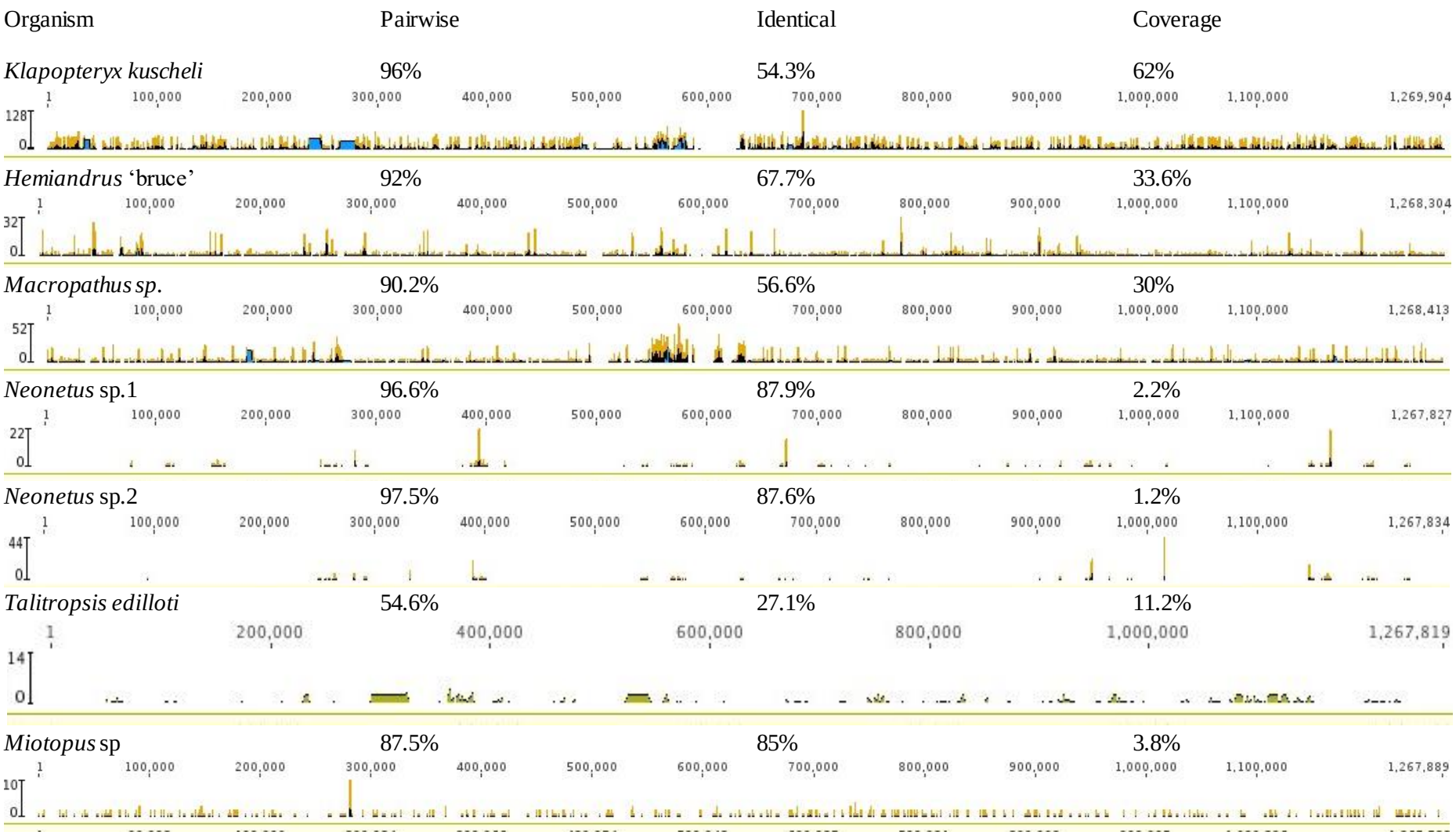


Figure 2.5. Graphical representation of relative coverage of *Wolbachia* endosymbionts of *Drosophila melanogaster* (NC002978)

and the number that matched other unrelated sequences, as well as indicating the number of reads that did not come back with a GenBank result. Manual analysis of the *Vollenhovia* associated sequences resulted in only a single sample of *Vollenhovia* in each of the BLAST searches with the rest of the results matching the expected *Wolbachia* genus

2.4 Discussion

Wolbachia was detected in seven individuals (*K. kuscheli, Macropathus* sp., *Hemiandrus* sp, *T. edilloti, Miotopus sp*, and two *Neonetus* sp) six of which are endemic to New Zealand. These are the first known cases of *Wolbachia* infection in native New Zealand invertebrates. The two Orthoptera species *Macropathus sp.* (Rhaphidophoridae) and *Hemiandrus sp* (Anostostomatidae) had a coverage of the *Wolbachia* genome of ~30% which was the largest of the New Zealand samples. *Klapopteryx kuscheli* had the largest total number of reads associated with *Wolbachia* as determined through MEGAN, this corresponded with the largest coverage of the *Wolbachia* genome with an increase to 62% coverage.

Wolbachia was also detected in a number of other cave weta (Rhaphidophoridae) species; however, the level of detection was low with around a tenth the number of reads detected by MEGAN compared to *Wolbachia* rich HTS. This corresponded with a much lower cover of the *Wolbachia* genome when mapped, 11.2% for *T. edilloti* and <5% for *Miotopus sp* and both *Neonetus* species. Although this result could indicate a false positive hit to *Wolbachia* this seems unlikely as *Wolbachia* represented the majority of prokaryote reads detected in the analysis. DNA analysis produced Pairwise % Identity and identical sites of the samples of ≥85% in *Miotopus sp* and both *Neonetus* species. *T. edilloti* had a lower pairwise of 54% and identical 27.1% which may account for the higher coverage. Another possibility is differences in the raw HTS output, as not all samples produced the same total number of sequences. MEGAN (Huson, et al., 2011) indicates the total number of reads input from the HTS file. This indicated that the three hosts (*K. kuscheli, Macropathus* sp., *Hemiandrus* sp.) (data not shown) had approximately twice the number of total reads analysed than (*T. edilloti, Miotopus sp* and both *Neonetus* species). Therefore, it correlates that the number of *Wolbachia* sequences detected will be higher if the total number of reads is higher.

To quantify how much confidence, we should have in the outputs it produced by PAUDA, the samples that suggested a *Wolbachia* infection was rerun through BLASTN using the reads identified as *Wolbachia*. The first thing noticed was a large number (1336) of reads from *Klapopteryx kuscheli* associating with *Vollenhovia*, a genus of ants. On further investigation it was found that the samples of *Vollenhovia* were tagged "PREDICTED" therefore likely to be determined by automated BLAST analysis and have yet been manually reviewed. Ants are a known host for *Wolbachia* therefore the matching sequences are likely to be from *Wolbachia* within the *Vollenhovia* sample that was sequenced and the PAUDA result indicating the sequence was from *Wolbachia* should be accepted. This indicates the potential issue of only looking at the top hits from BLAST outputs as can result in misleading results. This accounted for the majority of sequences within the data that matched *Wolbachia* in the PAUDA result but did not match to *Wolbachia* in the BLASTN rerun.

Two samples *Macropathus* sp and *Neonetus sp2* had another large grouping of sequences that did not match either *Wolbachia* or *Vollenhovia*. Both of these samples had a relatively high number of reads not matching any current nucleotide sequences on GenBank. As was indicated in Fig. 2.4 the *Macropathus sp* was located in a New Zealand isolated clade, this may indicate an ancestral infection. BLASTn requires a strong match to sequence to produce a result, this may result in changes in the *Wolbachia* genome that are no longer identified as *Wolbachia* in BLASTn searches but are still able to be detected in BLASTx searches potentially missing important information. This indicates that the PAUDA approach to mining large High-throughput datasets can be useful, however it should be one of a number of steps. It can be used to reduce the data set down to a subset of reads that are likely to be useful and allow you to run a more thorough search which would not be computationally efficient to do on the whole data set. However, using the raw output from PAUDA on its own may result in a misrepresentation of the data as a whole by introducing mismatched sequences rather than missing information.

Wolbachia is a maternally inherited cytoplasmic endoparasite that requires a host to survive and reproduce (Werren, 1997). *Wolbachia* can enter a host via vertical or horizontal transmission routes. Vertical transmission moves from mother to offspring, from generation to generation. Single strains of *Wolbachia* can be found across multiple host species, horizontal transmission involves the movement of *Wolbachia* from an infected host to a new uninfected host (of the same or different species) (Ahmed, et al., 2015; Heath, et

al., 1999). When the ftsZ sequence from NZ HTS data were compared to other sequences and GenBank it was apparent that they formed a unique clade (Fig. 2.4) representing New Zealand invertebrates, *Wolbachia* sequences from *K. kuscheli* were different and fell within a different super group (E). This suggests that the *Wolbachia* found within New Zealand insects had a common origin. As there was no match from GenBank the origin of this infection is not yet able to be determined, however the sister clade contained individuals from China, India, and Europe suggesting that the origin may be from the Asia region. Comparing the individuals in the New Zealand isolated clade to the super group tree (Lo, et al., 2002), placed these individuals in the B super group. Due to the recombination found within *Wolbachia* and the use of a single MLST gene it is to be noted that recombination could be the cause of the monophyly seen in this tree therefore further sequencing of the MLST gene will be required to confirm this observation.

As it has been shown that the clade containing *Wolbachia* from *Macropathus* sp also contained samples from *Hemiandrus sp.* and *Ectopsocus sp. Ectopsocus sp.* (book lice) is not closely related to either *Macropathus* sp or *Hemiandrus,* therefore the likelihood of vertical transmission being the mode of infection is very low. With vertical transmission *Wolbachia* would have been present since the last common ancestor of the host taxa. This is plausible in closely related species; however, in more distinctly related species the *Wolbachia* would have acquired significant differences that were not present in the samples analysed (Fig. 2.6). This suggests that horizontal transmission from one of these species (or another yet to be identified host) to the other species is the logical option.

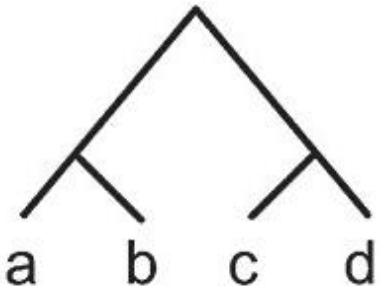
2.5 Conclusion

Through the use of HTS data *Wolbachia* was detected in seven individuals (*K. kuscheli*, *Macropathus sp*, *Hemiandrus* sp, *T. edilloti*, *Miotopus sp*, and two *Neonetus* sp), six of which are weta species endemic to New Zealand. This shows that the use of HTS data is a viable method for the detection of *Wolbachia* irrespective of the tissue originally used for the sequencing, however the level and/or rate of detection may be lower if somatic tissue is used. This may increase the detection of *Wolbachia* and add to the global knowledge of *Wolbachia* distribution. It does however highlight the necessity for additional conformation of the results, either by rerunning a subset through a stricter search algorithm or supplementing the information with directed molecular information such as PCR with

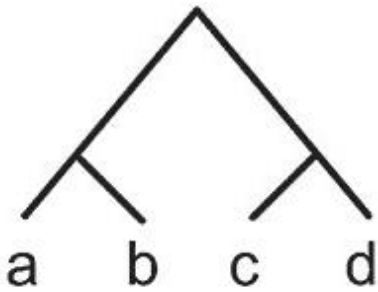
Wolbachia specific primers. Another consideration to be taken into account is the use of HTS does not always allow for the number of samples in each species that would be preferred for this type of approach and may result in a higher level of false negatives than other techniques, however if the samples are already available it remains a viable first step approach which would require limited input once the pipeline has been setup.

Predicted common ancestor-vertical transmission.

Host phylogeny



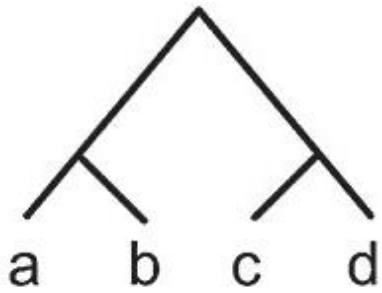
Wolbachia phylogeny



Observed

Either very low rate of molecular evolution or horizontal transfer

Host phylogeny



Wolbachia phylogeny

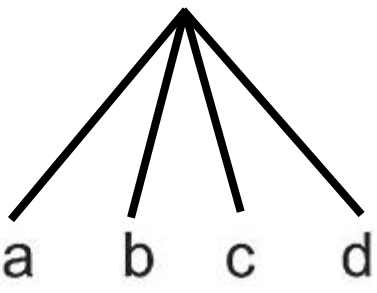


Figure 2.6 Example of phylogenetic trees based on either vertical or horizontal transfer.

2.6 References

Ahmed, M. Z. et al., 2015. The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. PLOS Pathogens, Volume 10(2).

Illumina, Inc, 2010. Technology Spotlight: Illumina Sequencing. [Online]
Available at: www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf
[Accessed 28 April 2015].

Altschul, S. et al., 1990. Basic local alignment search tool. Journal of Molecular Biology, Volume 215, pp. 403-410.

Baldo, L. et al., 2006. Multilocus Sequence Typing System for the Endosymbiont *Wolbachia* pipientis. Applied and Environmental Microbiology, Volume 76(11), pp. 7098–7110.

BGI, 2014. Illumina HiSeq2000. [Online]
Available at: http://www.genomics.cn/en/navigation/show_navigation?nid=2632
[Accessed 28 April 2015].

Brown, C. T. et al., 2012. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. [Online]
Available at: http://arxiv.org/pdf/1203.4802v2.pdf

Casiraghi, M. et al., 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology, Volume 151, pp. 4015-4022.

Handelsman, J., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology, Volume 5(10), pp. 245-249.

Heath, B., Butcher, R., Whitfield, W. & Hubbard, S., 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Current. Biology, Volume 9, pp. 313-316.

Huson, D. H. & Xie, C., 2013. A poor man's BLASTX—high-throughput metagenomic protein database search using PAUDA. Bioinformatics, Volume 30(1), pp. 38-39.

Huson, D. et al., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Research, Volume 21, pp. 1552-1560.

Jiggins, F. M., Schulenburg, J. H. v. D., Hurst, G. D. & Majerus, M. E., 2001. Recombination confounds interpretations of *Wolbachia* evolution. Proceedings of the Royal Society B, Volume 268(1474), pp. 1423-1427.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, Volume 25(12), pp. 1647-1649.

Lo, N. et al., 2002. How many Wolbachia supergroups exist?. Molecular Biology and Evolution, Volume 19, pp. 341-346.

Mardis, E. R., 2007. The impact of next-generation sequencing technology on genetics.

Shokralla, S., Spall, J. L., Gibson, J. F. & Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology, Volume 21, pp. 1794-1805.

Sogin, M. et al., 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". PNAS, Volume 103, pp. 12115-12120.

Werren, J. H., 1997. Biology of *Wolbachia*. Annual Review of Entomology, Volume 42, pp. 587-609.

Werren, J. H., Clark, M. E. & Baldo, L., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature, Volume 6, pp. 741-751.

Zerbino, D. R. & Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research, Volume 18(5), pp. 821-829.

Zhao, Y., Tang, H. & Ye, Y., 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics, Volume 20, pp. 125–126.

Chapter 3

The distribution and incidence of *Wolbachia* infections in New Zealand ground weta (*Hemiandrus* spp.) and cave weta (Rhaphidophoridae)

Chapter 3: The distribution and incidence of *Wolbachia* infections in New Zealand ground weta (*Hemiandrus* spp.), cave weta (Rhaphidophoridae), and South American stonefly (*Klapopteryx kuscheli).*

3.0 Abstract

Wolbachia is an endoparasite that moves through its host through vertical transmission down the maternal line. *Wolbachia* has also been shown to move horizontally through hosts of the same species or hosts of separate species. This movement allows for high propagation of *Wolbachia* through populations. To determine the distribution of *Wolbachia* through New Zealand, ground weta (*Hemiandrus* spp.), cave weta (Rhaphidophoridae), and South American stonefly (*Klapopteryx kuscheli)* were tested for *Wolbachia* using the MLST PCR protocol. *Wolbachia* was detected for the first time across both main islands of New Zealand and across multiple species or putative species of ground and cave weta.

Key words

Wolbachia, ground weta, cave weta, MLST, QGIS

Abbreviations

Multilocus sequence typing = MLST, polymerase chain reaction = PCR, Geographic Information System = GIS

3.1 Introduction

Wolbachia is a cytoplasmic endoparasite of arthropods best known for its ability to modify the reproductive strategies of its host. Examples of these modifications include, the dramatic generation of parthenogenesis in infected females, cytoplasmic incompatibility, feminization, and male killing. As *Wolbachia* is maternally inherited these modifications have evolved to help drive an increase in *Wolbachia* within an infected population (Werren, et al., 2008). *Wolbachia* is also able to move horizontally between species through the movement of tissue from an infected individual to an uninfected individual (Huson, et al., 2011). For example, parasitic wasps (Vavre, et al., 1999) and predation (Le Clec'H, et al., 2013), horizontal transmission has resulted in a global estimate of infection ranging from 40 to 66% of all arthropod species. However, we have yet to find any evidence of *Wolbachia* in certain groups of arthropod. An understanding of the incidence of *Wolbachia* infections will help determine the current distribution of the parasite within regions and the distribution of distinct *Wolbachia* lineages will help us understand its effect on local host species. This information will be useful in better understanding the mechanisms used by the bacteria to successfully move between different species.

Wolbachia has been documented for the first time in endemic New Zealand invertebrates through the use of HTS data (chapter 2). *Wolbachia* was detected in the New Zealand cave weta (*Macropathus* sp, *Talitropsis edilloti*, *Miotopus sp*, and two *Neonetus sp*) and a member of the New Zealand ground weta complex *Hemiandrus maculifrons*. *Wolbachia* was also detected in a Chilean stonefly, *Klapopteryx kuscheli,* a wide spread species within Chile. We do not know if these infections are modifying the host species via cytoplasmic interactions or male killing. Before determining the potential role for *Wolbachia* in altering their hosts, the distribution of *Wolbachia* within each of the three hosts must be investigated. Infection rates and the number of distinct genetic lineages of *Wolbachia* within each host will be investigated. The distribution pattern will show how much of an effect, *Wolbachia* could plausibly have had on the host. Cytoplasmic incompatibility would be expected to lead to closely related taxa being infected with different strains or some populations infected and some uninfected. This can introduce cryptic reproductive barriers between populations that can lead to speciation. This can become an issue when a species is under conservation efforts. Introduction of individuals from a source population can introduce negative consequences to the already small sink population if a *Wolbachia* infection has not identified.

The first family of New Zealand weta, Anostostomatidae, consists of tree, giant, tusked, and ground weta (Johns, 1997). Ground weta, *Hemiandrus*, is the most speciose genus with 11 described and 30 undescribed species (Johns, 2001). They are the only non-endemic genus of weta in New Zealand. *Hemiandrus* species are the smallest weta, their body size range between 12 – 45mm. *Hemiandrus maculifrons* is found in both the North and South Islands of New Zealand. However, this taxon is a species complex (Smith, 2016), consisting of at least three distinct lineages. The lineages sampled here are referred to as *Hemiandrus* 'south', 'central', and 'bruce' as determined by T. Smith (2016). We have detected *Wolbachia* in this species and we have suggested as a potential mechanism for instigating and/or propagating this separation.

Of the 18 genera of Rhaphidophoridae found in New Zealand *Pachyrhamma* is the best known with the greatest described species diversity. Species of *Pachyrhamma* are large as adults, often spanning 15cm. *Pachyrhamma* are the most recognised of New Zealand cave weta as several species regularly inhabit caves, tunnels, and long drops. However, many species are either rarely or never found in these environments. *Pachyrhamma* species are found in both North and South Island New Zealand. Of the 13 described species of *Pachyrhamma*, only one is known to be found exclusively in the South Island, *P. delli*, is restricted to Fiordland in the southern part of the South Island.

Klapopteryx is a genus in the order Plecoptera, commonly known as stoneflies. It resides within the family Austroperlidae, and as is usual with southern hemisphere genera of Plecoptera, it is endemic to Chile. Plecoptera have two major life stages, aquatic nymph and flying adult. *K. kuscheli* take around two years to develop to adulthood, with the mature nymph exceeding 3cm in length. *K. kuscheli* are widely distributed across Chile inhabiting Patagonian streams (Hollmann & Miserendino, 2008). Despite extensive geographic separation between populations mitochondrial diversity has been observed to be very shallow in this species (Personal communication), *Wolbachia* was previously detected using HTS. The *Wolbachia* infection of *K. kuscheli* could explain this observed low mitochondrial diversity and therefore further testing of this species across a number of populations will be undertaken.

One possible horizontal transmission mechanism of *Wolbachia* among carnivorous weta could be through their diet as all ground weta (*Hemiandrus* sp) are thought to be omnivorous scavengers or carnivores (Morgan-Richards, et al., 2008; Le Clec'H, et al.,

2013). However, cave weta feeding characteristics are poorly understood so it is not known if all cave weta eat other insects. However, there are wasps that parasitize cave weta eggs and these may provide an alternative mechanism of *Wolbachia* transfer. To explore this latter possibility, the weta parasitoid wasp *Archaeoteleia* was tested for *Wolbachia* infection.

To determine the distribution and diversity of *Wolbachia* within these groups, whole genomic DNA will be amplified (PCR) using the *Wolbachia*-specific multilocus sequence typing (MLST) scheme developed by (Baldo, et al., 2006) with the addition of *wsp* primers (Zhou, et al., 1998). The presence or absence of *Wolbachia*-specific DNA will be compared to the population structure of each of the target host species to determine whether *Wolbachia* is found in all populations or is localised in isolated populations. It is hoped that basic information on infect rates and distribution will shed light on any role played by *Wolbachia* in restricting gene flow among host populations.

3.2 Methods

Pairs of primers designed for internal fragments of five marker genes (*coxA*, *fbpA*, *gatB*, *hcpA*, and *ftsZ*) were used to amplify DNA. Primers for *wsp* were also included, as this locus has previously been shown to be highly polymorphic (Breeuwer & Jacobs, 1996; Watanabe, et al., 2013; Zhou, et al., 1998) and therefore potentially valuable for distinguishing between closely related *Wolbachia* lineages (Rokas, et al., 2002). As these primers have been designed to target non-eukaryotic genes, tailored specifically towards *Wolbachia*, PCR product using any of these six primer pairs provide strong evidence for the presence of *Wolbachia* DNA in a sample. Besides absence of *Wolbachia* in a sample, several technical issues could explain situations where the amplification reaction fails, including 1. Poor sample quality; 2. DNA at too low concentration; 3. Presence of PCR inhibitors; 4 Concentration of *Wolbachia* being low due to non ovary tissue used for extraction; and 5. DNA concentration too high. Therefore, positive control PCR reactions with universal insect mitochondria primers (LCO1490-HCO2198) that target host DNA were also performed to ensure DNA extractions were suitable templates for PCR amplification. Positive PCR controls for *Wolbachia* genes using DNA from the parasitic wasp *Nasonia*, which is known to be infected with *Wolbachia* were included in each PCR experiment.

DNA was extracted from 205 individual ground weta and tested for the presence of *Wolbachia* DNA using the MLST primer combinations (Table 3.1). These individuals represented 14 species/clades (*H.* 'alius', *H. bilobatus, H.* 'disparalis', *H. electra, H. elegans, H. focalis, H. furoviarius, H. horomaka, H. maculifrons, H. nitaweta, H. onokis, H. promontorius, H. subantarctic,* and *H. vianus*), with *H.* 'alius' being separated in two populations, North Island and South Island. The lineage *H. maculifrons* was separated into 3 taxa designated as 'central', 'south', and 'bruce'. Due to the differences between the groups being primarily genetic not all individuals have been designated, individuals not currently classified into one of the three groups will remain as the overarching *H. maculifrons*. Leg tissue was used as material for extraction, as the samples were previously extracted for the use of host CO1 sequencing, using a modified salting out method, incorporating an ice cold ethanol washing step before addition of room temperature ethanol and allowing the ethanol to evaporate leaving the DNA to be eluted in 50µl water (Miller, et al., 1988).

The primers were tested on 45 cave weta from the genera, *Pachyrhamma* (23), *Isoplectron* (6), and to be identified (16) (Table 3.1) Leg tissue or abdomen tissue was used as material for extraction, using a modified salting out method, incorporating an ice cold ethanol washing step before addition of room temperature ethanol and allowing the ethanol to evaporate leaving the DNA to be eluted in 50µl water (Miller, et al., 1988).

A total of 28 individuals of *Klapopteryx kuscheli* were tested for the presence of *Wolbachia* across 17 populations located in Chile. (Table 3.1). Nine parasitoid wasps, spanning (7) collection sites, of the genus *Archaeoteleia* (four species/taxa), were collected from pan traps and tested for *Wolbachia*. Ethanol preserved specimens of adult wasps were donated by John Early (Auckland War Memorial Museum). DNA was extracted from specimens using either abdomens, or whole individuals (Werren & Windsor, 2000), using either prepGEM Insect kit (ZyGEM, Hamilton, New Zealand) or GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich Co).

To increase the range of samples tested for *Wolbachia* a further 40 individuals consisting of 24 species were included. Samples were collected from southern North Island, from a range of urban and forest habitats. 16 exotic species were collected and 8 New Zealand native or endemic species of invertebrates (Table 3.1).

To determine the spatial distribution of *Wolbachia* infections in New Zealand the collection locations of all putative host individuals were recorded. The ground weta and cave weta collection locations were mapped using QGIS (QGIS Development Team, 2015). To visualise the distribution of *Wolbachia* in New Zealand, individual locations were coloured according to whether the insects collected there were infected with *Wolbachia* or not.

Table 3.1 List of specimens tested, their current identification, infection status, and location (Where available)

Code	Order	Identification	Present	Latitude	Longitude
FD1	Orthoptera	*H. nitaweta*	Y		
FD4	Orthoptera		N		
GW025	Orthoptera	*H. bilobatus*	N	-41.312818	174.780891
GW100	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW1010	Orthoptera	*H.* 'central'	N	-42.3811	172.3027
GW1011	Orthoptera	*H. maculifrons*	Y	-42.3811	172.3027
GW1013	Orthoptera	*H.* 'central'	Y	-42.3811	172.3027
GW1014	Orthoptera	*H.* 'central'	N	-42.3811	172.3027
GW1015	Orthoptera	*H.* 'central'	Y	-42.3811	172.3027
GW1017	Orthoptera	*H. maculifrons*	N	-42.3811	172.3027
GW1018	Orthoptera	*H. maculifrons*	N	-42.3811	172.3027
GW1019	Orthoptera	*H. maculifrons*	Y	-42.3811	172.3027
GW102	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW1020	Orthoptera	*H. maculifrons*	Y	-42.3811	172.3027
GW1021	Orthoptera	*H. maculifrons*	Y	-42.3811	172.3027
GW1022	Orthoptera	*H. maculifrons*	N	-42.3811	172.3027
GW1025	Orthoptera		N	-42.3811	172.3027
GW1026A	Orthoptera	*H.* 'bruce'	N	-41.758995	172.969655
GW1026B	Orthoptera	*H.* 'bruce'	N	-41.758995	172.969655
GW105	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW1069	Orthoptera	*H. focalis*	N	-46.565385	168.472198
GW108	Orthoptera	*H.* 'bruce'	N	-38.682034	176.068699
GW109	Orthoptera	*H.* 'bruce'	Y	-38.682034	176.068699
GW1096	Orthoptera	*H. furoviarius*	N	-44.259404	170.103591
GW110	Orthoptera	*H.* 'bruce'	N	-38.682034	176.068699
GW124	Orthoptera	*H. electra*	N	-42.351981	171.379809
GW129	Orthoptera	*H. horomaka*	N	-43.299577	171.749957
GW133	Orthoptera	*H. maculifrons*	N	-43.13831	171.74012
GW141	Orthoptera	*H. maculifrons*	N	-45.450659	167.57498
GW142	Orthoptera	*H. maculifrons*	N	-45.450659	167.57498
GW146	Orthoptera	*H. maculifrons*	N	-42.80157	171.57052
GW169	Orthoptera	*H.* 'bruce'	N	-37.9673	175.5714
GW172	Orthoptera	*H.* 'bruce'	Y	-37.9673	175.5714
GW195i	Orthoptera	*H.* 'bruce'	N	-38.76719	176.21866
GW195ii	Orthoptera	*H.* 'bruce'	Y	-38.76719	176.21866
GW196	Orthoptera	*H.* 'bruce'	Y	-38.76719	176.21866
GW198	Orthoptera	*H. maculifrons*	N	-45.59273	167.95144
GW202	Orthoptera	*H. maculifrons*	N	-45.59273	167.95144
GW21	Orthoptera	*H.* 'bruce'	Y	-38.2918	177.3848
GW219	Orthoptera	*H.* 'bruce'	N	-39.12406	175.39138
GW221	Orthoptera	*H.* 'bruce'	Y	-39.12406	175.39138
GW222	Orthoptera	*H.* 'bruce'	N	-39.17198	174.95486
GW223	Orthoptera	*H.* 'bruce'	Y	-39.17198	174.95486
GW224	Orthoptera	*H.* 'bruce'	Y	-39.17198	174.95486
GW225	Orthoptera	*H.* 'bruce'	Y	-39.17198	174.95486
GW229	Orthoptera	*H.* 'bruce'	N	-39.14865	173.93888
GW234	Orthoptera	*H.* 'bruce'	N	-36.752893	175.504443
GW237	Orthoptera	*H.* 'bruce'	N	-39.14865	173.93888
GW238	Orthoptera	*H.* 'bruce'	Y	-39.14865	173.93888
GW239	Orthoptera	*H.* 'bruce'	N	-39.14865	173.93888

GW247A	Orthoptera	*H.* 'bruce'	N	-36.752893	175.504443
GW259	Orthoptera	*H. maculifrons*	N	-43.906997	168.90806
GW261	Orthoptera	*H. maculifrons*	N	-43.906997	168.90806
GW27	Orthoptera		N	-42.736038	172.816929
GW32B	Orthoptera	*H.* 'south'	N	-46.572129	169.455619
GW33A	Orthoptera	*H. disparalis*	N	-41.835694	172.807297
GW35A	Orthoptera	*H.* 'SubA'	N	-48.029785	166.57985
GW36B	Orthoptera	*H. maculifrons*	N	-43.860175	169.451028
GW41	Orthoptera	*H. maculifrons*	Y	-45.529107	169.589632
GW41B	Orthoptera	*H. maculifrons*	Y	-45.529107	169.589632
GW44	Orthoptera	*H.* 'bruce'	N	-35.5074	173.4053
GW46	Orthoptera	*H.* 'alius' SI	N	-41.803274	172.845712
GW461	Orthoptera	*H. maculifrons*	Y	-42.33421	172.17611
GW462	Orthoptera	*H. maculifrons*	N	-42.33421	172.17611
GW462A	Orthoptera	*H. maculifrons*	Y	-42.33421	172.17611
GW463A	Orthoptera	*H. maculifrons*	Y	-42.33421	172.17611
GW463B	Orthoptera	*H. maculifrons*	N	-42.33421	172.17611
GW464	Orthoptera	*H.* 'bruce'	N	-35.31539	174.256
GW465	Orthoptera	*H.* 'central'	Y	-38.575053	177.102364
GW481	Orthoptera	*H. maculifrons*	N	-41.801818	172.851886
GW484	Orthoptera	*H. maculifrons*	Y	-41.801818	172.851886
GW49Bi	Orthoptera	*H.* 'bruce'	N	-35.187283	173.762883
GW49Bii	Orthoptera	*H.* 'bruce'	Y	-35.187283	173.762883
GW548	Orthoptera	*H.* 'bruce'	N	-38.51686	175.58072
GW549	Orthoptera	*H.* 'bruce'	Y	-38.51686	175.58072
GW550	Orthoptera	*H.* 'bruce'	Y	-38.51686	175.58072
GW551	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW552	Orthoptera	*H.* 'central'	Y	-39.67298	176.21216
GW553	Orthoptera	*H.* 'central'	Y	-39.67719	176.2501
GW554	Orthoptera	*H.* 'bruce'	Y	-39.727116	175.138788
GW557	Orthoptera	*H.* 'bruce'	Y	-39.727116	175.138788
GW558	Orthoptera	*H.* 'bruce'	Y	-39.727116	175.138788
GW559	Orthoptera	*H.* 'bruce'	Y	-39.727116	175.138788
GW560	Orthoptera	*H.* 'bruce'	N	-39.103194	175.378241
GW564	Orthoptera	*H. maculifrons*	N	-44.977308	168.017918
GW565	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW568	Orthoptera	*H.* 'central'	Y	-40.926518	172.857594
GW569	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW570	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW571	Orthoptera	*H.* 'central'	Y	-40.926518	172.857594
GW573	Orthoptera	*H.* 'central'	Y	-40.926518	172.857594
GW574	Orthoptera	*H.* 'central'	Y	-40.926518	172.857594
GW575	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW578	Orthoptera	*H. maculifrons*	Y	-40.926518	172.857594
GW586	Orthoptera	*H. promontorius*	N	-41.889423	173.623387
GW611	Orthoptera	*H.* 'central'	Y	-40.735401	175.380928
GW617	Orthoptera	*H.* 'bruce'	Y	-36.53724	174.710653
GW618	Orthoptera	*H.* 'bruce'	Y	-36.53724	174.710653
GW62	Orthoptera	*H. elegans*	N	-36.535849	175.401469
GW624	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW625	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW626	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW627	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW628	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW629	Orthoptera	*H.* 'bruce'	N	-38.84725	175.55739
GW630	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW631	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW632	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW633	Orthoptera	*H.* 'bruce'	Y	-38.84725	175.55739
GW64	Orthoptera	*H.* 'bruce'	Y	-35.165892	173.816495
GW664	Orthoptera	*H.* 'alius' NI	Y	-38.409659	177.414229
GW665	Orthoptera	*H.* 'bruce'	Y	-38.409659	177.414229
GW694	Orthoptera	*H.* 'south'	Y	-41.190025	172.747206

GW700	Orthoptera	*H.* 'bruce'	Y	-38.299747	177.333355
GW701	Orthoptera	*H.* 'bruce'	Y	-38.299747	177.333355
GW702	Orthoptera	*H.* 'bruce'	Y	-38.299747	177.333355
GW703	Orthoptera	*H.* 'bruce'	Y	-38.299747	177.333355
GW711	Orthoptera	*H.* 'central'	Y	-40.481816	175.636808
GW717	Orthoptera	*H. vianus*	N	-41.124813	174.055456
GW734	Orthoptera	*H.* 'alius' NI	Y	-38.757455	177.151378
GW735	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW736	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW737	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW738	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW74	Orthoptera	*H.* 'bruce'	Y	-36.9052	174.56966
GW740	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW741	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW742	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW745	Orthoptera	*H.* 'bruce'	Y	-38.76736	177.157729
GW750	Orthoptera	*H.* 'central'	Y	-38.766357	177.169402
GW753	Orthoptera	*H.* 'bruce'	Y	-38.757455	177.151378
GW758	Orthoptera	*H.* 'south'	Y	-41.765809	171.773694
GW759	Orthoptera	*H.* 'central'	Y	-40.469823	175.612302
GW760	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW761	Orthoptera	*H.* 'central'	Y	-40.469823	175.612302
GW765	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW767	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW769	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW772	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW773	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW775	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW777	Orthoptera	*H.* 'bruce'	Y	-39.14865	173.93888
GW778	Orthoptera	*H.* 'bruce'	Y	-39.14865	173.93888
GW797	Orthoptera	*H.* 'bruce'	N	-40.469823	175.612302
GW800	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW801	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW802	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW807	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW811	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW812	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW813	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW814	Orthoptera	*H.* 'bruce'	Y	-40.469823	175.612302
GW815	Orthoptera	*H.* 'central'	Y	-40.469823	175.612302
GW816	Orthoptera	*H. maculifrons*	Y	-40.469823	175.612302
GW817	Orthoptera	*H.* 'central'	Y	-40.469823	175.612302
GW834	Orthoptera	*H.* 'alius' SI	N	-42.647484	171.062828
GW836	Orthoptera	*H. maculifrons*	Y	-42.647484	171.062828
GW837	Orthoptera	*H. maculifrons*	Y	-42.647484	171.062828
GW872	Orthoptera	*H. maculifrons*	Y	-42.381032	172.40309
GW874	Orthoptera	*H. maculifrons*	N	-42.381032	172.40309
GW875	Orthoptera	*H. maculifrons*	Y	-42.381032	172.40309
GW878	Orthoptera	*H. maculifrons*	Y	-42.381032	172.40309
GW88Bi	Orthoptera	*H.* 'bruce'	N	-41.296882	173.573192
GW88Bii	Orthoptera	*H.* 'bruce'	N	-41.296882	173.573192
GW891B	Orthoptera	*H.* pallitarsis	N	-40.469823	175.612302
GW893	Orthoptera	*H. maculifrons*	Y	-39.324394	174.106418
GW896A	Orthoptera	*H.* 'alius' SI	N	-41.768026	171.778745
GW896B	Orthoptera	*H.* 'alius' SI	N	-41.768026	171.778745
GW897A	Orthoptera	*H. maculifrons*	Y	-41.296882	173.573192
GW898	Orthoptera	*H. maculifrons*	N	-42.946256	171.54695
GW899	Orthoptera	*H.* 'alius' SI	N	-41.768026	171.778745
GW90	Orthoptera	*H.* 'bruce'	N	-38.767644	177.111727
GW900	Orthoptera	*H.* 'central'	Y	-40.934636	172.972177
GW907	Orthoptera	*H. maculifrons*	Y	-41.296882	173.573192
GW908	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW909A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038

GW909B	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW909C	Orthoptera	*H. maculifrons*	N	-41.508874	173.933038
GW910A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW910B	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW911A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW911B	Orthoptera	*H. maculifrons*	N	-41.508874	173.933038
GW912	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW913A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW913B	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW913C	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW914A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW914B	Orthoptera	*H. maculifrons*	N	-41.508874	173.933038
GW914C	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW915A	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW915B	Orthoptera	*H. maculifrons*	Y	-41.508874	173.933038
GW916B	Orthoptera	*H.* 'central'	Y	-39.99521	176.098968
GW918	Orthoptera	*H.* 'alius' NI	N	-38.276729	177.341887
GW919	Orthoptera	*H.* 'bruce'	Y	-39.122346	174.124952
GW922A	Orthoptera	*H.* 'south'	Y	-42.474678	171.253426
GW922D	Orthoptera	*H.* 'south'	Y	-42.474678	171.253426
GW922E	Orthoptera	*H.* 'south'	Y	-42.474678	171.253426
GW922F	Orthoptera	*H.* 'south'	Y	-42.474678	171.253426
GW924A	Orthoptera	*H.* 'south'	N	-42.474678	171.253426
GW924B	Orthoptera	*H.* 'south'	N	-42.474678	171.253426
GW925A	Orthoptera	*H.* 'south'	N	-42.474678	171.253426
GW925B	Orthoptera	*H.* 'south'	N	-42.474678	171.253426
GW937	Orthoptera	*H.* 'bruce'	N	-37.900601	176.203608
GW94A	Orthoptera	*H.* 'alius' NI	Y	-38.568594	177.102913
GW958	Orthoptera	*H. onokis*	N	-41.505733	173.79801
CW346	Orthoptera	*Pachyrhamma*	N	-38.731823	176.704302
CW368	Orthoptera	*Pachyrhamma*	N	-41.968909	172.69001
CW671	Orthoptera		N	-36.971179	176.081525
CW676	Orthoptera		N	-36.956187	174.598709
CW677	Orthoptera		N	-36.956187	174.598709
CW978	Orthoptera	*Pachyrhamma*	N	-45.450237	167.573875
CW680	Orthoptera		N	-36.956187	174.598709
CW688	Orthoptera		N	-41.01868	172.902778
CW746	Orthoptera	*Pachyrhamma*	N	-41.394465	174.045869
CW747	Orthoptera	*Pachyrhamma*	N	-41.394465	174.045869
CW766	Orthoptera	*Pachyrhamma*	N	-36.503015	175.425756
CW1006	Orthoptera		N	-39.270076	176.497241
CW1047	Orthoptera	*Pachyrhamma*	N	-40.926518	172.857594
CW1051	Orthoptera		N	-37.284607	176.25091
CW1625	Orthoptera		N	-35.183648	173.310813
CW1635	Orthoptera	*Isoplectron*	Y	-39.660058	177.030674
CW1636	Orthoptera		N	-38.997697	176.286275
CW1639	Orthoptera		N	-35.183648	173.310813
CW1652	Orthoptera		N	-40.847829	174.913451
CW1656	Orthoptera		N	-40.847829	174.913451
CW1688	Orthoptera		N	-41.089269	174.781222
CW1872	Orthoptera	*Pachyrhamma*	N	-39.878272	176.099311
CW108	Orthoptera	*Pachyrhamma*	Y	-38.262039	175.112886
CW156	Orthoptera	*Isoplectron*	N	-40.864726	175.862665
CW1626	Orthoptera	*Isoplectron*	Y	-39.660058	177.030674
CW1827	Orthoptera	*Pachyrhamma*	Y	-38.263806	175.125766
CW1871	Orthoptera	*Pachyrhamma*	Y	-38.263806	175.125766
CW1887	Orthoptera		N	-39.878272	176.099311
CW1914	Orthoptera	*Pachyrhamma*	Y	-38.263806	175.125766
CW192	Orthoptera	*Isoplectron*	N	-45.045763	168.547902
CW1961	Orthoptera	*Isoplectron*	Y	-39.660058	177.030674
CW1974	Orthoptera	*Pachyrhamma*	Y	-38.093873	177.291171
CW1978	Orthoptera	*Pachyrhamma*	Y	-38.703876	176.024104
CW1979	Orthoptera	*Pachyrhamma*	Y	-38.703876	176.024104

CW1980	Orthoptera	*Pachyrhamma*	N	-38.703876	176.024104
CW198A	Orthoptera	*Pachyrhamma*	Y	-38.768179	177.11207
CW2657	Orthoptera	*Isoplectron*	N	-39.366642	176.515815
CW318	Orthoptera	*Pachyrhamma*	Y	-38.260403	175.101611
CW418	Orthoptera	*Pachyrhamma*	Y	-35.719893	174.350604
CW494	Orthoptera	*Pachyrhamma*	Y	-36.752826	175.504037
CW495	Orthoptera		Y	-36.752826	175.504037
CW53	Orthoptera	*Pachyrhamma*	N	-39.222994	176.378145
CW682	Orthoptera		N	-36.949039	174.533247
CW69	Orthoptera	*Pachyrhamma*	Y	-39.173645	175.3943
CW96	Orthoptera	*Pachyrhamma*	N	-38.723252	176.701555
bk21	Plecoptera	*Klapopteryx kuscheli*	N		
bk01	Plecoptera	*Klapopteryx kuscheli*	N		
bk48	Plecoptera	*Klapopteryx kuscheli*	N		
bk02	Plecoptera	*Klapopteryx kuscheli*	N		
tz02	Plecoptera	*Klapopteryx kuscheli*	N		
md01	Plecoptera	*Klapopteryx kuscheli*	N		
br10	Plecoptera	*Klapopteryx kuscheli*	N		
rv01	Plecoptera	*Klapopteryx kuscheli*	N		
sc01	Plecoptera	*Klapopteryx kuscheli*	N		
rv10	Plecoptera	*Klapopteryx kuscheli*	N		
ma04	Plecoptera	*Klapopteryx kuscheli*	N		
ch40	Plecoptera	*Klapopteryx kuscheli*	Y		
el03	Plecoptera	*Klapopteryx kuscheli*	Y		
lm02	Plecoptera	*Klapopteryx kuscheli*	N		
ac02	Plecoptera	*Klapopteryx kuscheli*	N		
po01	Plecoptera	*Klapopteryx kuscheli*	N		
po02	Plecoptera	*Klapopteryx kuscheli*	N		
ch61	Plecoptera	*Klapopteryx kuscheli*	N		
bk6o	Plecoptera	*Klapopteryx kuscheli*	N		
ch71	Plecoptera	*Klapopteryx kuscheli*	N		
Tq01	Plecoptera	*Klapopteryx kuscheli*	Y		
Bk44	Plecoptera	*Klapopteryx kuscheli*	Y		
Bk63	Plecoptera	*Klapopteryx kuscheli*	N		
Bk64	Plecoptera	*Klapopteryx kuscheli*	N		
TR01	Plecoptera	*Klapopteryx kuscheli*	N		
MA01	Plecoptera	*Klapopteryx kuscheli*	N		
LP02	Plecoptera	*Klapopteryx kuscheli*	N		
PZ02	Plecoptera	*Klapopteryx kuscheli*	N		
Arc1	Hymenoptera	*Archaeoteleia gilbertae*	Y	-41.350393	173.59871
Arc2	Hymenoptera	*Archaeoteleia karere*	N	-38.728376	174.963112
Arc3	Hymenoptera	*Archaeoteleia onamata*	N	-40.891285	172.98522
Arc4	Hymenoptera	*Archaeoteleia karere*	Y	-38.728376	174.963112
Arc5	Hymenoptera	*Archaeoteleia onamata*	N	-40.891285	172.98522
Arc6	Hymenoptera	*Archaeoteleia onamata*	N	-40.891285	172.98522
Arc7	Hymenoptera	*Archaeoteleia karere*	Y	-38.728376	174.963112
Arc8	Hymenoptera	*Archaeoteleia* 'gilberti'	N	-41.386308	173.210385
Arc9	Hymenoptera	*Archaeoteleia* 'gilberti'	Y	-40.891285	172.98522
Vir01	Lepidoptera	*Aenetus virescens*	N	-40.382621	175.619402
Vul01	Hymenoptera	*Vespula vulgaris*	N	-40.382621	175.619402
Vul02	Hymenoptera	*Vespula vulgaris*	N	-40.382621	175.619402
Vul03	Hymenoptera	*Vespula vulgaris*	N	-40.382621	175.619402
Ger01	Hymenoptera	*Vespula germanica*	N	-40.389769	175.623049
Ger02	Hymenoptera	*Vespula germanica*	N	-40.389769	175.623049
Sco01	Hemiptera	*Scolypopa australis*	N	-40.382621	175.619402
Sco02	Hemiptera	*Scolypopa australis*	N	-40.382621	175.619402
Sten01	Plecoptera	*Stenoperla sp.*	N	-40.382621	175.619402
Cha01	Coleoptera	*Halmus chalybus*	N	-40.382621	175.619402
Cha02	Coleoptera	*Halmus chalybus*	N	-40.382621	175.619402
Pro01	Hymenoptera	*Proctotrupoidea sp.*	N	-40.382621	175.619402
Dom01	Diptera	*Musca domestica*	N	-40.382621	175.619402
Dan01	Lepidoptera	*Danaus plexippus*	N	-40.382621	175.619402
Pso01	Psocoptera	*Ectopsocus sp.*	Y	-40.382621	175.619402

Pso02	Psocoptera	*Ectopsocus sp.*	Y	-40.382621	175.619402
Tip01	Diptera	*Tipulidae*	N	-40.382621	175.619402
Ruf01	Diptera	*Chrysomya rufifacies*	N	-40.382621	175.619402
Can01	*Diptera*	*Fannia canicularis*	N	-40.382621	175.619402
Chl01	*Diptera*	*Chlorops sp.*	Y	-40.382621	175.619402
Dro01	*Diptera*	*Drosophila sp.*	N	-40.382621	175.619402
Lep01	Diptera	*Leptotarsus sp.*	N	-40.382621	175.619402
Lep02	Diptera	*Leptotarsus sp.*	N	-40.382621	175.619402
Lept01	Diptera	*Leptotarsus sp.*	N	-40.382621	175.619402
Lept02	Diptera	*Leptotarsus sp.*	N	-40.382621	175.619402
Mel01	Hymenoptera	*Apsis mellifera*	Y	-40.382621	175.619402
Mel02	Hymenoptera	*Apsis mellifera*	Y	-40.382621	175.619402
Tri01	Diptera	*Trigonospila brevifacies*	N	-40.382621	175.619402
Col01	Ephemeroptera	*Coloburiscus humeralis*	N	-40.382621	175.619402
Col02	Ephemeroptera	*Coloburiscus humeralis*	N	-40.382621	175.619402
Aot01	Trichoptera	*Aoteapsyche sp.*	N	-40.382621	175.619402
Aot02	Trichoptera	*Aoteapsyche sp.*	N	-40.382621	175.619402
Sip01	Hemiptera	*Siphanta acuta*	N	-40.382621	175.619402
Sip02	Hemiptera	*Siphanta acuta*	N	-40.382621	175.619402
Div01	Megaloptera	*Archichauliodes sp.*	N	-40.382621	175.619402
Div02	Megaloptera	*Archichauliodes sp.*	N	-40.382621	175.619402
Lig93	Isopoda	*Ligia novaezealandiae*	N	-37.630044	176.429154
Ony366	Euonychophora	*Peripatus morgani*	N		
Ony365	Euonychophora	*Peripatus morgani*	N		
Ony368	Euonychophora	*Peripatus morgani*	N		

3.3 Results

An individual insect was considered positive for a *Wolbachia* infection if DNA amplification produced a product for at least one of the MLST or wsp gene fragments. Positive results for two independent lineages of both cave weta and ground weta were detected through PCR. *Wolbachia* was detected in the *Hemiandrus maculifrons* complex and *Hemiandrus* 'alius' clade of ground weta and the genus *Pachyrhamma* and *Isoplectron* for the cave weta (Rhaphidophoridae) (Table 3.1).

The *Hemiandrus maculifrons* complex consists of three genetically distinct clades currently designated as *H.* 'bruce', *H.* 'central', and *H.* 'south'. *Wolbachia* was detected in all three clades. Infections rates varied from 56% - 90% of individuals tested (Table 3.2). These rates were 65 of 89 for *H.* 'bruce', 18 of 20 for *H.* 'central', 30 of 49 for *H.* 'south', and 14 of 25 for yet to be determined *Hemiandrus maculifrons* (Table 3.2) were positive for at least one of the MLST primers. *Hemiandrus* 'alius' is a clade of ground weta that is yet to be formally described but genetic and morphological evidence supports it as a separate species (Smith, 2016). *Hemiandrus* 'alius' is currently separated in to two groups, North Island and South Island individuals. *Wolbachia* was detected in three of four individuals tested from North Island, but not present in any of the six individuals from the South Island (Table 3.2). In addition to the two ground weta species mentioned a further 12 species were

tested (Table 3.1), 1 other sample of the further 12 individuals was positive for infection, this was a single *Hemiandrus nitaweta* individual.

Table 3.2 Level of infection in different species tested

Species	Total	Positive	% Infected
Bruce	89	65	73
Central	20	18	90
South	49	30	61
Maculifrons	25	14	56
Pachyrhamma	23	12	52
Isoplectron	6	3	50
Cave weta	16	1	6
Klapopteryx	28	4	14
Archaeoteleia	9	4	44

A total of 45 individuals of cave weta were tested for the presence of *Wolbachia* from the genera *Pachyrhamma* and *Isoplectron*. Three of the six *Isoplectron* tested were positive for *Wolbachia*. For *Pachyrhamma* 12 of the 23 individuals tested for *Wolbachia* were positive for at least one of the *Wolbachia* specific primers. Thus infection rate estimates within *Pachyrhamma* is 52% and infection rate within *Isoplectron* estimated at 50%. *Isoplectron* individuals that were positive for infection showed a stronger response than individuals identified as *Pachyrhamma* indicating a potentially higher within individual infection within *Isoplectron* individuals. One further individual of the 'to be confirmed' cave weta was positive for *Wolbachia* (CW495).

In addition to the New Zealand individuals tested for *Wolbachia* infections, Chilean stoneflies, *Klapopteryx kuscheli*, were also tested for the parasite. Testing 28 individuals across 17 populations resulted in 4 positive results, each individual was collected from a different population.

As a potential vector for *Wolbachia* horizontal transmission the parasitoid wasp *Archaeoteleia* was tested for infection. Of the nine tested, four individuals were positive for an infection, two from *A.* 'gilberti' and an individual from *A. onamata* and *A. kawere* (Table 3.1).

High throughput sequencing samples previously analysed in chapter two and positive for *Wolbachia* infections: *Klapopteryx kuscheli*, *Hemiandrus* 'bruce', *Macropathus* sp, *Neonetus* sp1, *Neonetus* sp2, *Talitropsis sedilloti*, and *Miotopus* sp. This increases the

detection of Stonefly individuals by one, ground weta ('bruce') by one, and increases the number of cave weta species positive for *Wolbachia* by five.

Of the further 40 individuals from twenty-four species collected from North Island New Zealand and tested for *Wolbachia* using MLST primers. *Wolbachia* infection was identified in three species; *Ectopsocus* sp. (indigenous booklice), *Chlorops* sp. (exotic frit fly), and two individuals of *Apis mellifera* (Western honey bee native to Europe, Asia and Africa).

The spatial distribution of *Wolbachia* infections of weta were visualised for the New Zealand sampling (Fig. 3.1-3.3). Many locations were found to have ground weta individuals both positive and negative for *Wolbachia,* this suggests that where *Wolbachia* is present it is not at saturation. *Wolbachia* was detected throughout the North Island and northern South Island (Fig. 3.1). However, *Wolbachia* was noticeably absent from the southern half of the South Island apart from two individuals that were positive for *Wolbachia* at Lake Onslow Rd, Otago (Fig 3.1). The distribution of *Wolbachia* infections in cave weta was localised to central and northern North Island (Fig. 3.3).

3.4 Discussion

Detecting the presence of *Wolbachia* DNA in insect genomic DNA extractions is used to infer active infections by this bacterial parasite (Baldo, et al., 2006). However, amplification of one or more *Wolbachia* specific markers might result from horizontal transfer of DNA sequences from parasites into host genomes in the evolutionary past. Thus sequencing of the amplified products is important to establish that the infection is active. In the case of New Zealand weta HTS data, chapter two provided evidence that *Wolbachia* DNA was from a current infection. Failure to detect *Wolbachia* DNA via amplification might result from low copy numbers in the tissue used to extract host DNA. Thus negative results must be considered preliminary. DNA from all potential host weta was extracted from the same tissue (femur muscle) in collaboration with Smith (2016) where *Wolbachia* is not expected to be found at high densities. However, within this study infection rates detected can be compared as the host DNA was of similar quality and quantity and a single protocol was followed.

Wolbachia is a widespread bacterial endosymbiont estimated to infect 66% of all arthropods (Hilgenboecker, et al., 2008). It has been shown that *Wolbachia* appear to

Figure 3.1 Distribution of ground weta specimens. Blue represents positive, Green represents negative, and teal represents both positive and negative individuals at same location (Individual Codes)

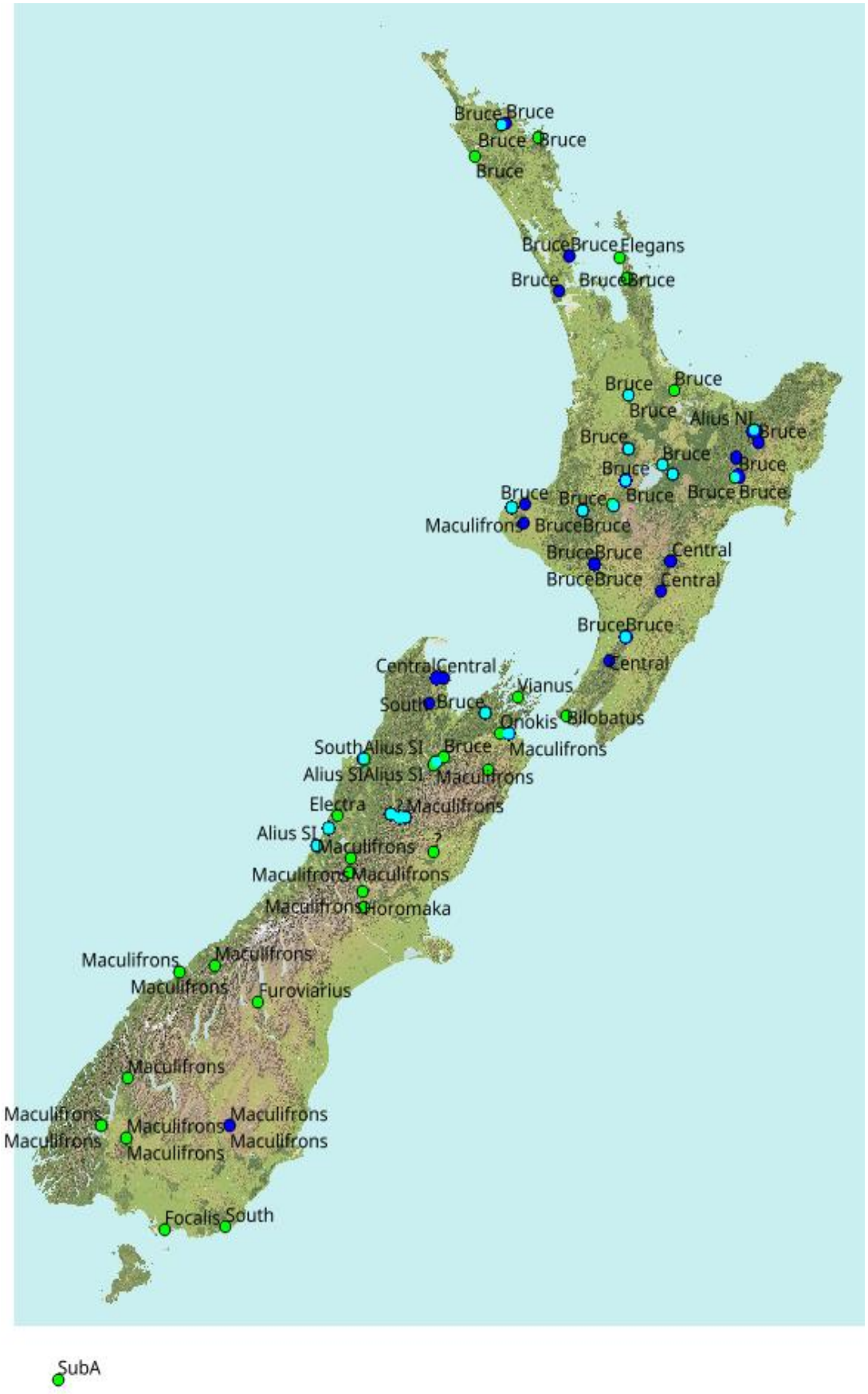


Figure 3.2 Distribution of ground weta specimens. Blue represents positive, Green represents negative, and teal represents both positive and negative individuals at same location (Identifications)

Figure 3.3 Distribution of cave weta specimens. Blue represents positive, Green represents negative, and teal represents both positive and negative individuals at same location

exhibit a most or few infection pattern (Hilgenboecker, et al., 2008). This suggests that if *Wolbachia* is detected in a species it will either infect very few numbers <10% or the majority of individuals >90%. This may rely heavily on the reproductive modification method employed by *Wolbachia* with lower infection rates (5-50%) observed in male-killing *Wolbachia* (Hurst & Jiggins, 2000). *Wolbachia* has been considered to primarily reside in the reproductive tissues (Werren, et al., 2008), however it is becoming more apparent that *Wolbachia* is residing in somatic tissue (Chen, et al., 2005; Dobson, et al., 1999) although the presence and bacterial load within the somatic tissue can differ between hosts (Cheng, et al., 2000). The distribution of *Wolbachia* can be varied between host species (Hilgenboecker, et al., 2008), however *Wolbachia* distribution can also differ between infected populations of the same species (Zhang, et al., 2013). Distribution frequencies were observed to differ between populations of the same species, ranging from 30% infection to 100% infection, the difference of infection could be due to a multitude of effects but there appears to be a geological correlation as indicated by Zhang, et al (2013).

Wolbachia was detected in all three clades of the ground weta complex *Hemiandrus maculifrons*. Infections were detected in the majority of individuals tested, with 73% of *H.* 'bruce', 90% of *H.* 'central', 61% of *H.* 'south', and 56% of un-categorised *H. maculifrons*. This accounts for 69% of all *H. maculifrons* individuals tested being positive for *Wolbachia* infections. *H.* 'bruce' and *H.* 'central' both fall within or close to the high level pattern of infection as suggested by Hilgenboecker, et al (2008). *H.* 'south' and uncatagorised had a much lower level of infection well below the >90% of the high infection but much higher than the <10% seen in the lower level infections. This could be due to a combination of low individual numbers and not all MLST regions being able to be tested due to time constraints, resulting in an underestimation of the infection status of these 'species'. *Wolbachia* has been shown to exhibit different reproductive modifications in different hosts despite being the same strain of *Wolbachia* (Werren, et al., 2008). As observed by Hurst and Jiggins (2000) the reproduction modification can alter the observered infection, however as these 'species' are still closely related it would be unlikely to see different modifications in these species expecially if the *Wolbachia* was a contributing factor in their speciation.

The transmission method of *Wolbachia* in these species is not known. Determining the transmission pathways is important in identifying the potential effects of the parasite on the infected host. Transmission method and host fitness will determine the rate of infection

within a population. Complete infection of all individuals has been observed in some host species (Dedeine, et al., 2001; Dedeine, et al., 2005), therefore it is yet not possible to distinguish whether the 69% infection rate observed is representative of the species as a whole. While all clades were positive for *Wolbachia* not all samples had the same amplification of *Wolbachia* specific markers. The clade *H.* 'bruce' consistently produced strong amplification with the MLST primer set, as indicated by the brightness of the band in gel electrophoresis. With all individuals undergoing the same DNA extraction method this may indicate a higher infection level within the *H.* 'bruce' line of ground weta, providing a potential mechanism for the reproduction isolation seen in the *H.* 'bruce' lineage from the surrounding *H.* 'central' clade, alternatively the *Wolbachia* sequence in this species may have happened to bind more strongly to the primers aka primer binding efficiency was higher for this host species, sequencing will be required to determine if there are differences between the *Wolbachia* found within and between each of the clades.

Wolbachia was also detected in the *Hemiandrus* 'alius' clade of ground weta. It was detected in the North Island individuals, with three of the four individuals tested being positive. As *Wolbachia* was only detected in the North Island individuals this might suggest the infection occurred after the separation of the two populations. This may have allowed for the North Island 'alius' to remain differentiated from other North Island Ground weta species in the surrounding area. However as is apparent with the infection rates in *Hemiandrus*, a larger sample size is need to confirm this.

Wolbachia was detected in three clades of cave weta, *Pachyrhamma*, *Isoplectron*, and *Macropathus*. *Wolbachia* was detected in 52% of all *Pachyrhamma* with 12 of 23 samples positive. The *Pachyrhamma* individuals that were found to be positive for *Wolbachia* produced identifiable but weak bands with PCR, this may indicate that *Pachyrhamma* is exhibiting a limited infection, possibly suggesting that *Wolbachia* is present in such a level to be detected but not enough to have a measurable effect on their hosts, (Breeuwer & Werren, 1993; Hoffmann, et al., 1996) in a sense hitching a ride through the host generation. *Wolbachia* propagate through a population by altering the host reproduction system to its own end, the limited level of infection could be a sign of a new infection making its way into the population or an old infection on its way out. Breeuwer and Warren (1993) found that it was the bacterial density that determined the number of compatibility issues, altering the level of bacterial load through antibiotics altered the ability of the females to successfully mate with infected males. Further sequencing of *Wolbachia* may elude to the

state of infection within these individuals. Of course, with using host extractions as a source of testing it is possible that interference from host DNA (Nuclear and mitochondrial) to reduce the ability of the primers to find a match and produce limited product (Cogswell, et al., 1996).

Isoplectron was also determined to be a host to an infection with a 50% infection rate (3 of 6), with the three individuals producing strong definitive products, suggesting a strong infection. Further samples will need to be tested to determine the level of infection at both the population and genera level. Sequencing of the three positive samples, as well as any individuals in future screening, will be required to investigate any effect on the host genetics at either the population level or potential speciation effects. *Wolbachia* was detected in *Macropathus* through HTS. This individual was previously miss identified as *Pachyrhamma waitomoensis,* and therefore no further individuals of this genus were tested in this study, however inclusion of this genus in further surveys is recommended. Of the further 16 cave weta tested there was a single positive result detected (CW495). As these individuals are in the process of identification it is not currently known whether *Wolbachia* would be expected in these individuals based on the results already obtained however it may be useful in further analysis of these hosts once they obtain formal identification.

Klapopteryx kuscheli is a stonefly found across Chile. Individuals of *K. kuscheli* from 17 populations were tested, resulting in a positive result for individuals from four populations. In conjunction with the HTS data analysed, results in five positive results. This infection rate of 14% is the lowest of the genera tested (not including the un-identified cave weta), although a limited number of individuals per population were tested. *K. kuscheli* has been shown to have limited mitochondrial genetic diversity (Unpublished), therefore the presence of *Wolbachia,* which can have dramatic reduction in the hosts genetic diversity, is still a viable possibility with the five positive samples indicating that *Wolbachia* is both present in Chile and specifically in *K. kuscheli.* As mentioned earlier *Wolbachia* has been shown to be a most or few pattern, therefore it is possible that *Wolbachia* is falls with the 'few' category in this species and the low mitochondrial diversity is due to other effects (environmental, previous *Wolbachia* infection) rather than the current infection. As was shown from the HTS data (chapter 2) the ftsZ sequence from *K. kuscheli* was quite different from the majority of sequences on GenBank therefore making it possible that the primers used were not very specific to this strain. This may also account for no positive indication for the ftsZ region (data not shown) in any of the samples tested. However, due to the wide

spread nature of the species, it will require a much larger and more focused study to determine the true infection rate in *K. kuscheli* and potential effects of said infection that was not possible in this study.

Archaeoteleia is a primitive parasitoid wasp known for its parasitism of New Zealand cave weta. Individuals were tested to determine if there was the potential for being a vector for horizontal transfer of *Wolbachia* between individuals and species. *Wolbachia* was found in 4 individuals consisting of three lineages. This holds potential for movement of *Wolbachia* between cave weta and other weta genera presuming that *Archaeoteleia* also parasitize or come into contact in a meaningful way with ground weta. It has been shown that the insertion of infected tissue can result in the transmission of infection in to the new individual (Watanabe, et al., 2013). This allows the possibility of failed parasitic events resulting in movement of *Wolbachia* from the parasitoid to the host. Ground weta hunt invertebrate prey which could also facilitate infection of *Wolbachia* through consumption of infected prey (Le Clec'H, et al., 2013). As has been demonstrated by Ahmed, et al (2015) *Wolbachia* can also be moved from infected eggs to uninfected eggs via the mouth and ovipositor. When parasitoids visit their prey it does not always result in death of the prey as they may feed or oviposition check rather than lay eggs. Of the individuals that had been visited by parasitoids and emerged as whiteflies, Ahmed, et al (2015) found that 93.8% bacame infected after surviving the parasitoid penetration. With such high infection rates in surviving individuals, parasotids hold another avenue for movement of *Wolbachia* between unrelated species.

Wolbachia was detected in both North and South Islands of New Zealand (Fig. 3.1-3.3). In the ground weta *Hemiandrus, Wolbachia* was distributed throughout the North Island and northern South Island (Fig. 3.1-3.2). In the cave weta *Wolbachia* was only detected in the North Island, primarily the centre North Island (Fig. 3.3). Relative to the ground weta, the number of individuals tested was limited therefore the respective distribution of positive results would also be lower. A larger dataset may show a greater level of infection similar to that in the ground weta. The two likely explanations for this distribution are that the infection occurred in the North Island and are spreading though the North Island and in to the South Island with the two outliers potentially translocated from further north. This hypothesis will explain the presence of *Wolbachia* in the North Island group of *H.* 'alius' but the absence of *Wolbachia* in the South Island *H.* 'alius'. The second explanation is that current sampling is not detecting the full level of infection and the distribution being

observed is a subsection of the underlying infection in New Zealand. As this is the first known focused search for *Wolbachia* in New Zealand there is no guarantee that the species that have been tested in this study are the primary species effected by the *Wolbachia* infection and what has been detected maybe the after effects of interactions with the yet to be detected primary host.

3.5 Conclusion

High infections rates of the different ground weta lineages did not suggest that this parasite is involved in creating reproductive barriers between ground weta species. No definitive pattern in *Wolbachia* distribution has yet been determined in New Zealand. It was present across all lineages of *Hemiandrus maculifrons* and spanning both main islands. Further surveying will elude to the prevalence of the disparity of infections between islands, requiring both increased numbers in the species tested as well as increasing the species count of actively infected taxa, especially increasing the number of individuals collected from the South Island. *Wolbachia* was detected in multiple species supporting the hypothesis of large amounts of horizontal transfer within New Zealand.

3.6 References

Ahmed, M. Z. et al., 2015. The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. PLOS Pathogens, Volume 10(2).

Baldo, L. et al., 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia* pipientis. Applied and Environmental Microbiology, Volume 76(11), pp. 7098–7110.

Breeuwer, J. A. J. & Jacobs, G., 1996. *Wolbachia*: intracellular manipulators of mite reproduction. Experimental & Applied Acarology, Volume 20(8), pp. 421-434.

Breeuwer, J. A. J. & Werren, J. H., 1993. Cytoplasmic Incompatibility and Bacterial Density in *Nasonia vitripenni*. Genetics, Volume 135, pp. 565-574.

Cheng, Q. et al., 2000. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. Medical and Veterinary Entomology, Volume 14(1), pp 44-50.

Chen, W.-J.et al., 2005. Using in situ hybridization to detect endosymbiont *Wolbachia* in dissected tissues of mosquito host. Journal of Medical Entomology, Volume 42(2), pp. 120-124.

Cogswell, F. B. et al., 1996. Host DNA Can Interfere with Detection of *Borrelia burgdorferi*. Journal of Clinical Microbiology, Volume 34(4), pp. 980–982.

Dedeine, F., Boulétreau, M. & Vavre, F., 2005. *Wolbachia* requirement for oogenesis: occurrence within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation in *A. tabida*. Heredity, Volume 95, pp. 394–400.

Dedeine, F. et al., 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. PNAS, Volume 98(11), pp. 6247–6252.

Dobson, S. L. et al., 1999. *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. Insect Biochemistry and Molecular Biology, Volume 29(2), pp. 153–160.

Hilgenboecker, K. et al., 2008. How many species are infected with *Wolbachia*? – a statistical analysis of current data. FEMS Microbiology Letters, 281, pp. 215–220.

Hoffmann, A. A., Clancy, D. & Duncan, J., 1996. Naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity, Volume 76, pp. 1-8.

Hollmann, M. & Miserendino, M., 2008. Life history and emergence patterns of stonefly species in mountain streams of the Futaleufú basin, Patagonia (Argentina). International Journal of Limnology, Volume 44(2), pp. 135-144.

Hurst, G. D. & Jiggins, F. M., 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. emerging infectious diseases, Volume 6, pp. 329–336.

Huson, D. et al., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Research, Volume 21, pp. 1552-1560.

Johns, P. M., 1997. The Gondwanaland Weta: Family Anostostomatidae (Formerly in Stenopelmatidae, Henicidae or Mimnermidae): Nomenclatural Problems, World Checklist, New Genera and Species. Journal of Orthoptera Research, Volume 6, pp. 125-138.

Johns, P. M., 2001. Distribution and conservation status of ground weta, *Hemiandrus* species (Orthoptera: Anostostomatidae). Science for conservation, Volume 180, pp. 5-25.

Le Clec'H, W. et al., 2013. Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial isopods. PLoS ONE, 8(4).

Miller, S. A., Dykes, D. D. & Polesky, H. F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, Volume 16, p. 1215.

Morgan-Richards, M., Trewick, S. A. & Dunavan, S., 2008. When is it coevolution? The case of ground wētā and fleshy fruits in New Zealand. New Zealand Journal of Ecology, Volume 32, pp. 108-112.

QGIS Development Team, 2015. QGIS Geographic Information System.

Rokas, A. et al., 2002. The incidence and diversity of *Wolbachia* in gallwasps (Hymenoptera; Cynipidae) on oak. Molecular Ecology, Volume 11(9), pp. 1815-1829.

Smith, T., 2016. Evolution of diversity: analysis of species and speciation in *Hemiandrus* ground wētā. PhD Thesis Massey University.

Vavre, F. et al., 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. Molecular Biology and Evolution, Volume 16(12), pp. 1711-1723.

Watanabe, M., Kageyama, D. & Miura, K., 2013. Transfer of a parthenogenesis-inducing *Wolbachia* endosymbiont derived from Trichogramma dendrolimi into Trichogramma evanescens. Journal of Invertebrate Pathology, Volume 112(1), pp. 83-87.

Werren, J. H., Baldo, L. & Clark, M. E., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature, Volume 6, pp. 741-751.

Werren, J. H. & Windsor, D. M., 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium?. Proceedings of the Royal Society of London B, pp. 1277-1285.

Zhang, Y.-K.et al., 2013. Diversity of *Wolbachia* in natural populations of spider mites (genus *Tetranychus*): evidence for complex infection history and disequilibrium distribution. Microbial Ecology, Volume 65, p. 731–739.

Zhou, W., Rousset, F. & O'Neill, S., 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. Proceedings of the Royal Society of London B, Volume 265(1395), pp. 509-515.

Chapter 4

Distribution of *Wolbachia* strains and genetic diversity of their hosts in New Zealand

Chapter 4: Distribution of *Wolbachia* strains and genetic diversity of their hosts in New Zealand

4.0 Abstract

Wolbachia infections have the ability to alter their hosts reproductive mechanisms. This can result in a reduced genetic diversity of their hosts. *Wolbachia* moves vertically in a similar process as mitochondria therefore comparing the diversity of the *Wolbachia* infection to that of the hosts mitochondria can provide initial suggestions to any effect the *Wolbachia* detected is having on their hosts. Two strains of *Wolbachia* was detected in the New Zealand samples consisting of super groups A and B. Both strains were detected in the ground weta species complex *Hemiandrus maculifrons* however it is not yet possible with the current resolution determine if the observed diversity of *Hemiandrus maculifrons* complex is a result of the host parasite interaction.

Key words

Co-speciation, genetic diversity, horizontal transfer

Abbreviations

Cytoplasmic incompatibility = CI, High throughput sequencing = HTS Multilocus sequence typing = MLST, Geographic Information System = GIS

4.1 Introduction

Mechanisms that select for particular genotypes or individuals within a population have the potential to reduce the genetic diversity of a population. *Wolbachia* is a bacterial endoparasite that can alter a host's reproduction mechanisms. By altering its hosts reproduction, the bacterium increases the chance of being transferred into the next generation (Werren, et al., 2008). As *Wolbachia* spreads through a population, the majority of hosts surviving and/or reproducing will be infected, resulting in a population arising from a limited number of originally infected individuals. In cases where *Wolbachia* produces parthenogenesis in its host, a new host population can be created from a single mother. When *Wolbachia* causes cytoplasmic incompatibility (CI) the host population can be replaced by a single or a limited number of mothers and their offspring. Both of these scenarios result in greatly reduced genetic diversity in the host via population bottleneck.

DNA barcoding has become a staple in the field of molecular biology and molecular ecology wherever extensive reference collections exist (Valentini, et al., 2009; Hebert, et al., 2003). DNA barcoding uses a short fragment of DNA sequence to compare and contrast similarity of individuals/populations and match against databases of known species. In eukaryotes the established gene fragment shown to generally have sufficient within species conservation and high between species diversity is the mitochondrial cytochrome oxidase 1 gene (COI) (Hebert, et al., 2003). Part of COI has become the universal gene fragment for barcoding animals, however there are many cases where greater resolution between individuals of the same species is desired, or distinguishing between non eukaryote species is the objective. This is where custom primers are needed. In *Wolbachia* the established protocol is the multilocus sequence typing system (MLST) established by Baldo et al (2006). This uses the combination of five genes (gatB, coxB, hcpA, ftsZ, and fpbA) with the addition of the *Wolbachia* surface protein (WSP) gene (Zhou, et al., 1998) to distinguish between strains of *Wolbachia*. Due to the transmission mechanism of *Wolbachia* through the maternal line of the host it is possible to compare the COI diversity of the host and the MLST diversity of the *Wolbachia* parasite to determine if the *Wolbachia* has had an effect on the diversity of the host.

Genetic and morphological data have revealed that *Hemiandrus maculifrons* comprises a complex of three genetically distinct clades, which will soon be given species names (Smith, 2016). The lineages sampled here are referred to as *Hemiandrus H. 's*outh', *H.*

*'c*entral', and *H.* 'bruce' as determined by T. Smith (2016). These putative species have been shown to have distinct spatial distributions (Appendix G) with regions of overlapping territories. These regions of overlap should allow for interaction and gene flow between the *H. maculifrons* complex unless there is strong reproductive barriers.

Wolbachia has been detected for the first time in native New Zealand invertebrates, with several examples in New Zealand Orthoptera of the family Anostostomatidae (*Hemiandrus*; ground weta) (chapters 2 and 3). *Wolbachia* was detected in a number of hosts in two clades, *Hemiandrus maculifrons* and *Hemiandrus '*alius*'*. The presence of *Wolbachia* within this species complex provides a potential mechanism for reproductive isolation of the three clades even though they are widely sympatric. *Hemiandrus* 'alius' is a clade of ground weta with two distinct groups, divided spatially and genetically, one located in the North Island and the other in the South Island of New Zealand. *Wolbachia* was detected in only the North Island population and again the presence of *Wolbachia* holds potential for a possible mechanism for isolation.

In addition to the *Wolbachia* infection of ground weta, chapter 2 presented the first support for *Wolbachia* infection in cave weta (Rhaphidophoridae), with five individuals providing differing levels of coverage. *Macropathus* sp had a similar strength response as *Hemiandrus maculifrons*, whereas *Talitropsis sedilloti*, *Miotopus* sp, and both *Neonetus* species had a positive, but limited, response. Chapter three two new genera (*Pachyrhamma* and *Isoplectron*). Cave weta are a very diverse family of weta with 18 genera found only in New Zealand. The presence of *Wolbachia* provides a potential mechanism for the speciation of cave weta and reproductive isolation of species in sympatry.

Archaeoteleia is a genus of parasitoid wasp known to parasitize New Zealand Orthoptera. *Wolbachia* has been detected in these wasps which therefore have potential to be a vector for the horizontal transfer of *Wolbachia* (chapter three). DNA from individual hosts that gave positive signal for W*olbachia* (chapter three) are included in the analysis to determine if there is similarity between sequences found within the host weta samples to determine the likelihood of *Archaeoteleia* being a vector.

There are two distinct strains of *Wolbachia* infecting New Zealand invertebrates (Chapter two). Evidence is based on a limited number of individuals and therefore further sequencing is required to determine the distribution of these two strains both spatially and through the different host species. To determine the effect of *Wolbachia* on the genetic diversity of the

hosts, the ftsZ fragment of the *Wolbachia* multilocus sequence typing (MLST) system was sequenced. Phylogenetic analysis of the sequence data could be compared to the phylogenetic trees produced from the mtDNA COI region of the host to test the hypothesis of cophylogeny. To determine the potential for *Wolbachia* to have had an effect on the host it has infected, first cophylogeny must be determined. Once determined specific experiments can be undertaken to determine the extent the infection has had on the host.

4.2 Methods

A large number of PCR sequences of *Wolbachia* MLST genes were produced (chapter three). 86 samples that produced a positive result for the ftsZ region (Baldo, et al., 2006) were sequenced by Macrogen Inc. (Korea). Sequences were then uploaded into Geneious (version 6 http://www.geneious.com,) (Kearse, et al., 2012) and added to the sequences collected previously (chapter two).

Wolbachia sequences were aligned in Geneious and trimmed to produce an alignment of 95bp to 485bp. Sequence alignments were subjected to Bayesian phylogenetic analysis (MrBayes) (Fig. 4.1) (Huelsenbeck & Ronquist, 2001) (Ronquist & Huelsenbeck, 2003). This enabled me to identify the *Wolbachia* strain hosted by each weta. COI sequences available for *H. maculifrons* were uploaded in to Geneious and used to produce a phylogenetic tree (PHYML) (Guindon, et al., 2010). This allowed each individual to be categorised in to one of the three clades of *H. maculifrons* (Smith, 2016).

The *Wolbachia* and *H. maculifrons* trees were then compared (Fig. 4.2), to reveal which *H. maculifrons* individuals had both COI data and *Wolbachia* infection data and which corresponding strain was found within the host.

Sequences from clade A and clade B (as identified by MrBayes) were uploaded into PopART (PopART, 2015) and a minimum spanning network (Bandelt, et al., 1999) (epsilon 0) was completed (Fig. 4.3).

Strain information was added to the QGIS dataset (QGIS Development Team, 2015)(chapter two) to display the distribution of the hosts found carrying each strain and determine if hosts of differing strains were likely to be found in the same area or in separate areas. A third functional group was also included to represent double infected host

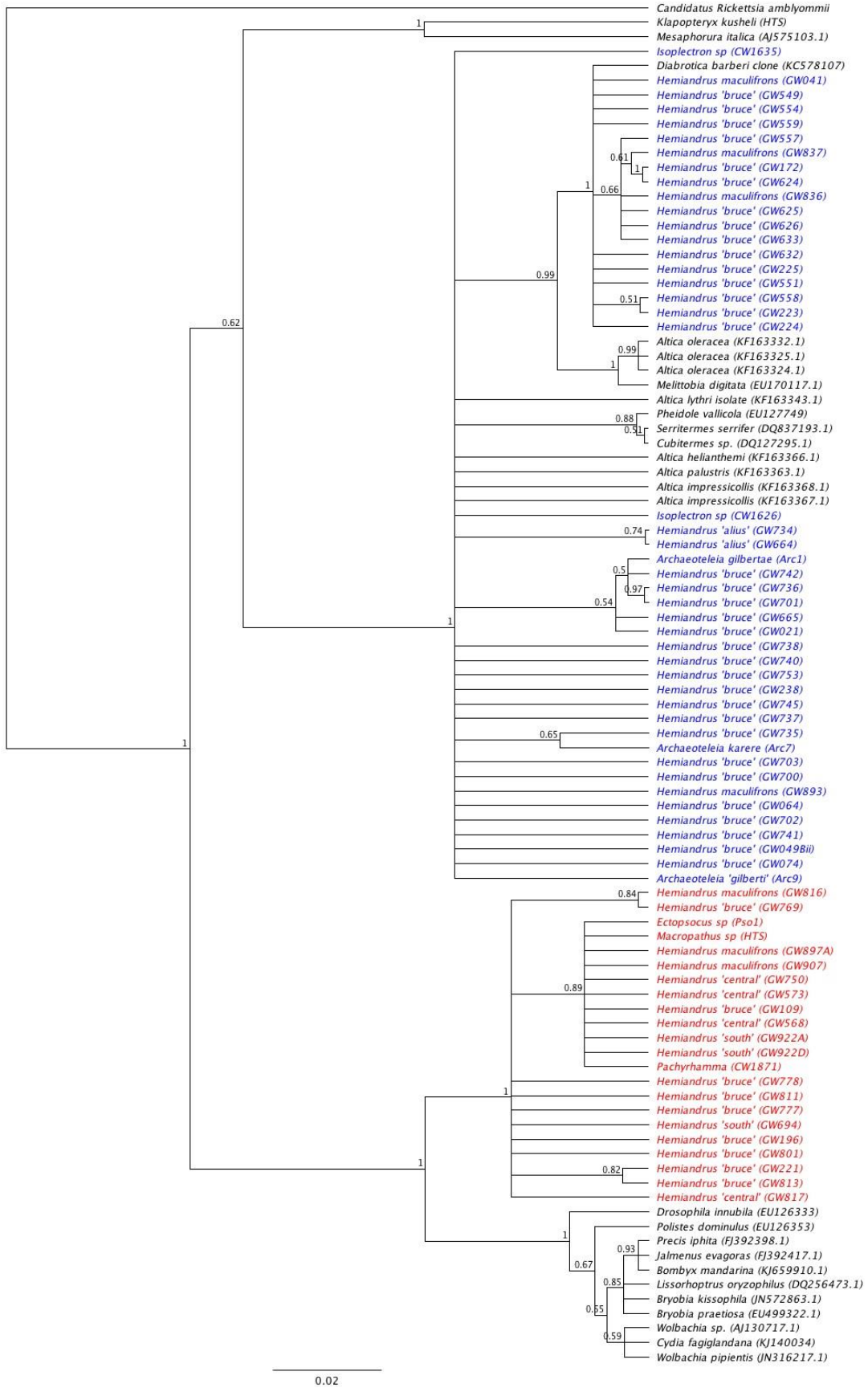


Figure 4.1 MrBayes tree of New Zealand and GenBank *Wolbachia endosymbionts based on* ftsZ sequences. *Species names are those of the host.* Strain one (*super group* A) is coloured blue, strain two (*super group* B) is coloured red.

individuals and reveal where both strains were detected and whether the double infected hosts were isolated or primarily found associated with one or the other strains.

4.3 Results

Sequencing of the ftsZ region of *Wolbachia* infections yielded a total of 86 sequences consisting of 77 *Maculifrons*, 1 *Pachyrhamma*, 2 'alius', 2 *Isoplectron*, 3 *Archaeoteleia*, and 1 *Ectopsocus* (Table 4.1).

Strain one (Fig. 4.1) matched infections from GenBank and had GenBank *Wolbachia* sequences interweaved within New Zealand samples (Table 4.1). The lack of monophyly in the occurrence of this strain suggested that it was not an ancestral infection that has moved vertically through the host species lineage, more closely resembling a relatively recent invasion that has transferred horizontally through New Zealand. Strain one was determined to be within super group A (Lo, et al., 2002) (chapter 2) and will be referred to Strain A. Strain two (Fig. 4.1) showed no strong match to any *Wolbachia* sequences currently available on GenBank producing a New Zealand monophyletic clade. The closest match to strain two sequences were the sister clade consisting of infections from China, India, and Europe suggesting that the origin may be from the Asia region. Strain two was determined to be in super group B (Lo, et al., 2002) and will be referred to as strain B. A number of hosts analysed appeared to exhibit sequence data from both strains GW550, GW552, GW569, GW628, and GW631, and GW919, this suggests that these hosts or their recent ancestors may have obtained a secondary infection.

Table 4.1 Specimen list indicating location name and GPS locations, and strain found

Code	Order	Identification	Latitude	Longitude	Clade
GW021	Orthoptera	*H.* 'bruce'	-38.2918	177.3848	1
GW041	Orthoptera	*H. maculifrons*	-45.529107	169.58963	1
GW049Bii	Orthoptera	*H.* 'bruce'	-35.187283	173.76288	1
GW064	Orthoptera	*H.* 'bruce'	-35.165892	173.8165	1
GW074	Orthoptera	*H.* 'bruce'	-36.9052	174.56966	1
GW100	Orthoptera	*H. maculifrons*	-40.926518	172.85759	2
GW109	Orthoptera	*H.* 'bruce'	-38.682034	176.0687	2
GW172	Orthoptera	*H.* 'bruce'	-37.9673	175.5714	1
GW196	Orthoptera	*H.* 'bruce'	-38.76719	176.21866	2
GW221	Orthoptera	*H.* 'bruce'	-39.12406	175.39138	2
GW223	Orthoptera	*H.* 'bruce'	-39.17198	174.95486	1

GW224	Orthoptera	*H.* 'bruce'	-39.17198	174.95486	1
GW225	Orthoptera	*H.* 'bruce'	-39.17198	174.95486	1
GW238	Orthoptera	*H.* 'bruce'	-39.14865	173.93888	1
GW465	Orthoptera	*H. 'c*entral'	-38.575053	177.10236	2
GW549	Orthoptera	*H.* 'bruce'	-38.51686	175.58072	1
GW550	Orthoptera	*H.* 'bruce'	-38.51686	175.58072	3
GW551	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW552	Orthoptera	*H. 'c*entral'	-39.67298	176.21216	3
GW553	Orthoptera	*H. 'c*entral'	-39.67719	176.2501	1
GW554	Orthoptera	*H.* 'bruce'	-39.727116	175.13879	1
GW557	Orthoptera	*H.* 'bruce'	-39.727116	175.13879	1
GW558	Orthoptera	*H.* 'bruce'	-39.727116	175.13879	1
GW559	Orthoptera	*H.* 'bruce'	-39.727116	175.13879	1
GW565	Orthoptera	*H. maculifrons*	-40.926518	172.85759	2
GW568	Orthoptera	*H. 'c*entral'	-40.926518	172.85759	2
GW569	Orthoptera	*H. maculifrons*	-40.926518	172.85759	3
GW571	Orthoptera	*H. 'c*entral'	-40.926518	172.85759	2
GW573	Orthoptera	*H. 'c*entral'	-40.926518	172.85759	2
GW574	Orthoptera	*H. 'c*entral'	-40.926518	172.85759	2
GW575	Orthoptera	*H. maculifrons*	-40.926518	172.85759	2
GW611	Orthoptera	*H. 'c*entral'	-40.735401	175.38093	2
GW624	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW625	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW626	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW628	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	3
GW631	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	3
GW632	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW633	Orthoptera	*H.* 'bruce'	-38.84725	175.55739	1
GW664	Orthoptera	Alius NI	-38.409659	177.41423	1
GW665	Orthoptera	*H.* 'bruce'	-38.409659	177.41423	1
GW694	Orthoptera	*H. 's*outh'	-41.190025	172.74721	2
GW700	Orthoptera	*H.* 'bruce'	-38.299747	177.33336	1
GW701	Orthoptera	*H.* 'bruce'	-38.299747	177.33336	1
GW702	Orthoptera	*H.* 'bruce'	-38.299747	177.33336	1
GW703	Orthoptera	*H.* 'bruce'	-38.299747	177.33336	1
GW711	Orthoptera	*H. 'c*entral'	-40.481816	175.63681	2
GW734	Orthoptera	Alius NI	-38.757455	177.15138	1
GW735	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW736	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW737	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW738	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW740	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW741	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW742	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW745	Orthoptera	*H.* 'bruce'	-38.76736	177.15773	1
GW750	Orthoptera	*H. 'c*entral'	-38.766357	177.1694	2

GW753	Orthoptera	*H.* 'bruce'	-38.757455	177.15138	1
GW759	Orthoptera	*H. 'c*entral'	-40.469823	175.6123	2
GW765	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW769	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW777	Orthoptera	*H.* 'bruce'	-39.14865	173.93888	2
GW778	Orthoptera	*H.* 'bruce'	-39.14865	173.93888	2
GW801	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW802	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW811	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW813	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW814	Orthoptera	*H.* 'bruce'	-40.469823	175.6123	2
GW815	Orthoptera	*H. 'c*entral'	-40.469823	175.6123	2
GW816	Orthoptera	*H. maculifrons*	-40.469823	175.6123	2
GW817	Orthoptera	*H. 'c*entral'	-40.469823	175.6123	2
GW836	Orthoptera	*H. maculifrons*	-42.647484	171.06283	1
GW837	Orthoptera	*H. maculifrons*	-42.647484	171.06283	1
GW893	Orthoptera	*H. maculifrons*	-39.324394	174.10642	1
GW897A	Orthoptera	*H. maculifrons*	-41.296882	173.57319	2
GW907	Orthoptera	*H. maculifrons*	-41.296882	173.57319	2
GW919	Orthoptera	*H.* 'bruce'	-39.122346	174.12495	3
GW922A	Orthoptera	*H. 's*outh'	-42.474678	171.25343	2
GW922D	Orthoptera	*H. 's*outh'	-42.474678	171.25343	2
CW1635	Orthoptera	*Isoplectron*	-39.660058	177.03067	1
CW1626	Orthoptera	*Isoplectron*	-39.660058	177.03067	1
CW1871	Orthoptera	*Pachyrhamma*	-38.263806	175.12577	2
Arc1	Hymenoptera	*Archaeoteleia* gilbertae	-41.350393	173.59871	1
Arc7	Hymenoptera	*Archaeoteleia* karere	-38.728376	174.96311	1
Arc9	Hymenoptera	*Archaeoteleia* 'gilberti'	-40.891285	172.98522	1
Pso01	Psocoptera	*Ectopsocus* sp.	-40.382621	175.6194	2

To determine the level of support between Strain A's three main groups, 342bp of ftsZ were loaded into PopART (PopART, 2015). The center grouping contained both New Zealand and GenBank *Wolbachia* sequences whereas the two outer groupings contained only New Zealand *Wolbachia* sequences. There was limited differences between the groupings within strain A however there was the distinction between the two groupings of New Zealand *Wolbachia* sequences observed in the MrBayes analysis (Fig. 4.3). To determine the differentiation between the two clades within strain B, the corresponding sequences were loaded in to PopART (PopART, 2015) and a minimum spanning network was run with 298bp. This showed that there was a minimum of eight differences between *Wolbachia* sequences in the New Zealand and GenBank clades (Fig. 4.3).

Wolbachia propagates through a population by influencing its host to interact with other individuals with the same strain of infection (Werren, et al., 2008; Werren, et al., 1995). The inclusion of the infection of multiple strains results in three functional reproductive

groups. The distribution of the three functional groups of *Wolbachia* varied among the limited number of samples. Strain A was detected in the majority of both the North and South Islands. Strain B was predominantly in hosts from the southern half of the North Island and the northern South Island. The distribution of the doubly infected hosts was more restricted with five of these individuals located in the central north island (GW550, GW552, GW628, GW631, and GW919) with the other (GW569) collected in the northern South Island (Fig. 4.4).

Phylogenies of ground weta and *Wolbachia* infections were compared. Hosts were coded according to whether they were infected with *Wolbachia* or not (Fig. 4.1) coloured with their corresponding *Wolbachia* strain. Some *H. maculifrons* individuals did not have a positive result or successful sequencing for ftsZ, and some *Wolbachia* sequences have not had their hosts COI regions sequenced yet so this tree does not contain all infected hosts. The clades within *H. maculifrons* matched either Strain A, Strain B, or had no corresponding *Wolbachia* detection (for the ftsZ region), with all but one clade (the clade containing individuals GW238, GW777, GW778, and GW893) containing a single *Wolbachia* strain.

4.4 Discussion

Wolbachia is a bacterial endoparasite that is able to modify its hosts reproductive mechanisms to increase its own propagation through a population. *Wolbachia* is distributed through the host's generations via the maternal line (Werren, 1997). This allows genetic testing of *Wolbachia* to determine its distribution pathway through the host population/species in a manner analogous to DNA barcoding using COI applied to its insect host. Due to the similarity in mode of inheritance of *Wolbachia* MLST regions to host COI (coxA is the *Wolbachia* equivalent of COI) (Smith, et al., 2012) it is possible to compare genealogies of hosts and *Wolbachia* to determine the likelihood of the infection of *Wolbachia* affecting the host genetics.

After the initial assessment of the *Wolbachia* sequences it was determined that two distinct clades were present, however after closer analysis it was observed that in many key sites on the DNA sequence that the strains differed, sequences from a number of hosts showed above background level signal for the nucleotide of the other strain with some SNPs being almost equal in identification strength. If the secondary peak detected in the chromatogram was varied or different between these hosts, it could be considered as random or an artefact of read error from the sequencing run. However, as the majority of alternative SNP calls matched one or other of the *Wolbachia* strain sequences indicating that these hosts were infected with of both *Wolbachia* strains. It is beneficial for *Wolbachia* inheritance for host individuals that mate to have the same infection, this results in a functional third clade consisting of the double infection (Fig. 4.1). Six individual ground weta that were identified as infected with both strains of *Wolbachia*; four *H.* 'bruce' (GW550, GW628, GW631, GW919), one *H. 'c*entral' (GW552), and one other unidentified *H. maculifrons* complex weta (GW569). The *H.* 'bruce' and *H.* 'central' samples came from southern North Island while GW569 was collected from the northern South Island. This indicates that the double infection occurs where the hosts of both clades are present in close proximity although it is not yet possible to determine whether it is direct interactions between the differently infected hosts or a third vector that facilitates the double infection.

The presence of *Wolbachia* in a population can provide ways for reproductive isolation of hosts from others of the same reproductive population. The first expectation when *Wolbachia* is detected is that infected hosts would be reproductively isolated from non-infected individuals but able to reproduce with fellow infected individuals. Cytoplasmic incompatibility is the most common *Wolbachia* reproductive mechanism. However, detection of *Wolbachia* does not exclude the possibility that the infection is with *Wolbachia* that produces a parthenogenetic outcome in the host as this results in reproductive isolation of hosts. This isolation can also occur between cytoplasmic incompatibility modified hosts if they contain different *Wolbachia*. Within the New Zealand insect samples tested, two distinct clades of *Wolbachia* were detected. One clade was found to be similar to *Wolbachia* sequences on GenBank and did not form any distinctive New Zealand clades (Fig. 4.1). This clade is situated in the A super group (shown in Chapter 2) and will be called from here on strain A. The intermixing DNA sequencing types from New Zealand hosts and data from GenBank, would indicate a recent infection to the country and possibly multiple cases of infection or horizontal transfer of *Wolbachia* rather than the typical vertical transmission

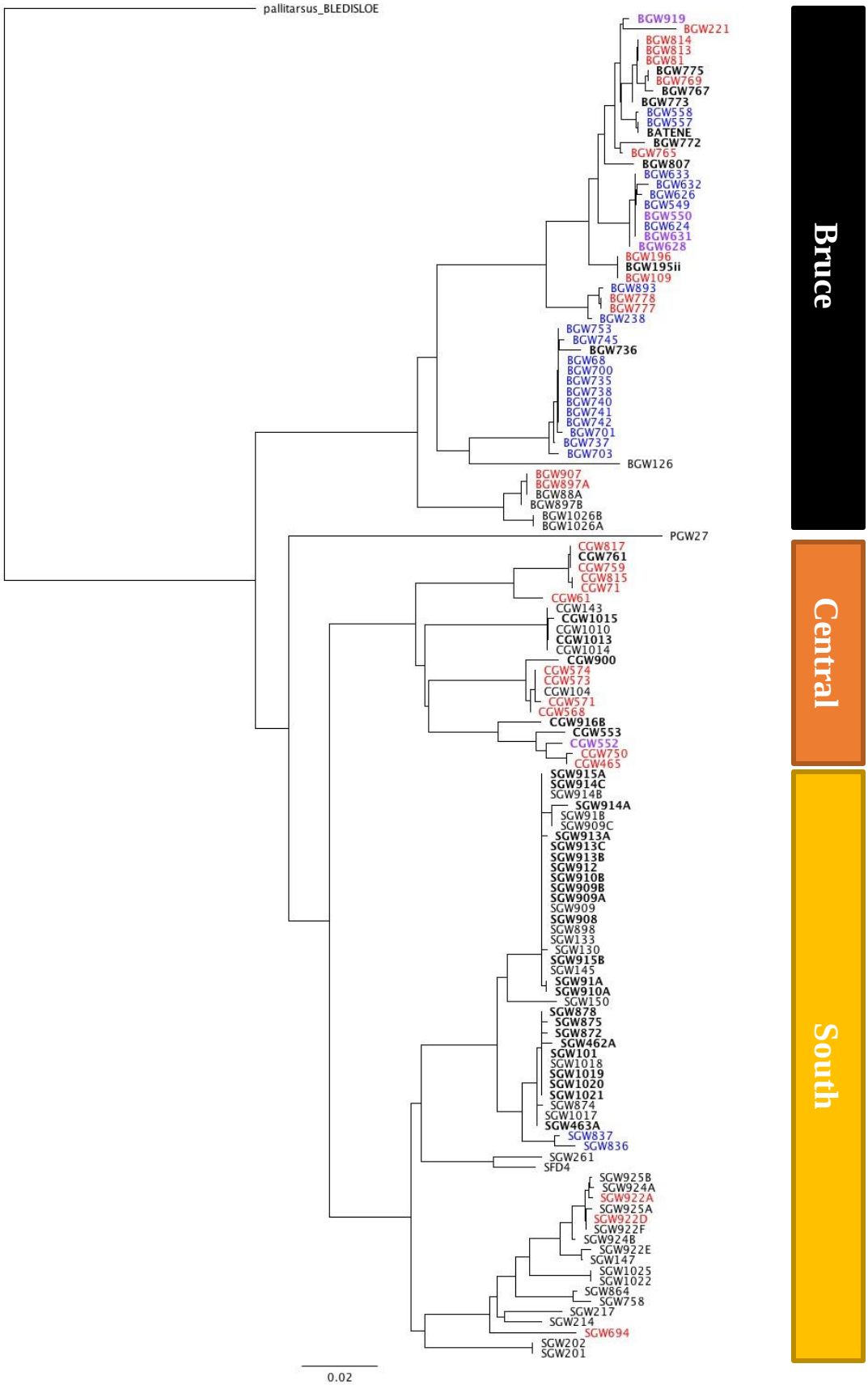


Figure 4.2 PHYML tree of H. maculifrons indicating which individuals were infected and had the ftsZ regions sequenced. Strain one is coloured blue, strain two is coloured red, and double infection is coloured Purple. Samples in bold indicate *Wolbachia* was detected but not sequenced

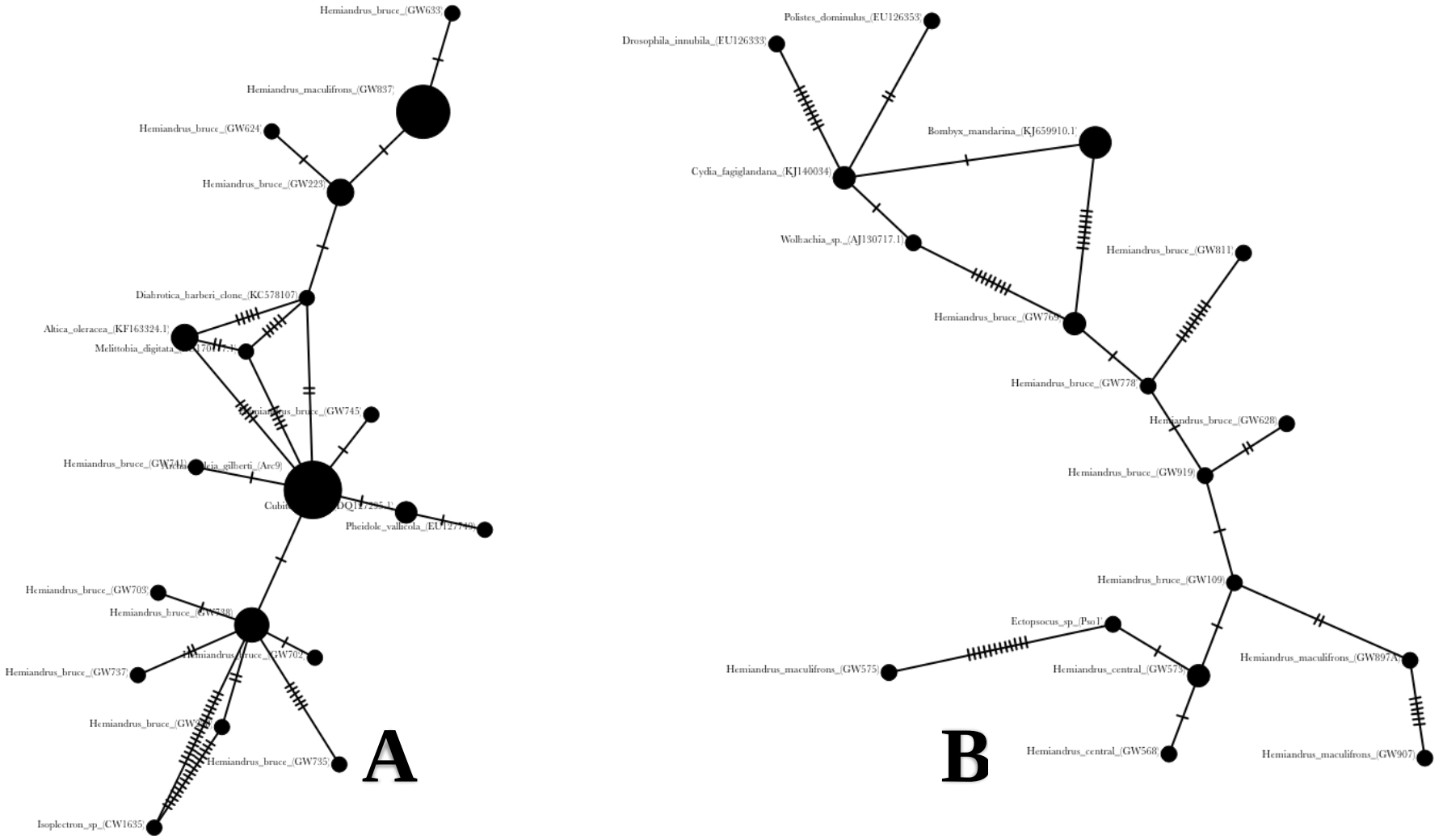


Figure 4.3 Minimum spanning network of strain A and *strain B* indicating number of differences between New Zealand and GenBank *Wolbachia* sequences

from mother to daughters. The tree showed low within clade resolution however the two clades of strain A is well resolved with a posterior probability of '1'. This provides support for the idea of two separate invasion events of strain A *Wolbachia* into New Zealand. However due to the similarity of the sequences the confidence in this hypothesis is reduced and further sequencing of the MLST regions will be needed to determine if the samples produce a monophyletic clade with further MLST fragments. Comparing the *Wolbachia* sequences from strain A in a medium spanning network showed three groupings of *Wolbachia*, two consisting of only New Zealand *Wolbachia* sequences and a third mixed grouping. These groups had limited sequence differences between the three groups, however the separation of the three groups provides additional to the support to two separate invasion events of *Wolbachia*.

'Strain A' included *Wolbachia* sequences from two *H. 's*outh' weta (GW836 & GW837) and another sample (GW41) was collected in the central South Island so is likely to fit within the *H. 's*outh' clade but has yet to have its CO1 sequenced. All other host samples containing 'strain A' were determined to be *H.* 'bruce'. This may suggest that *Wolbachia* entered the *H.* 'bruce' clade and spread before infecting the other clades. However, as the majority of sequenced hosts were *H.* 'bruce' the low presence of *H.* 'central' and *H.* 'south' hosts in strain A may be due to the saturation of *H.* 'bruce' samples and a focused sequencing project on the *H.* 'central' and *H.* 'south' clades will be required to tease out the distribution of strain A among the other *H. maculifrons* complex weta.

The second *Wolbachia* clade, produced a monophyletic clade of New Zealand samples when analysed with data from other parts of the world (Fig 4.1). This clade is situated within the *Wolbachia* B super group and is here referred to as strain B. Not only did they produce a single clade, the clade was also separate from other samples from New Zealand and any *Wolbachia* sequences currently on GenBank with a posterior probability of '1'. This genetic differentiation from other sequences in the tree suggests that the *Wolbachia* inhabiting these samples originated from an ancestral New Zealand infection. Among sequence alignment of 298bp of the MLST gene (ftsZ)(PopART) there was a minimum of eight sequence differences between any New Zealand and GenBank *Wolbachia* sequence sample (Fig. 4.3). As the clade contains *Wolbachia* sequences from multiple species it is not possible to determine if the infection of the species tested are from an ancestral infection or are new infections by the ancestral *Wolbachia*. As with strain A further testing of the MLST regions will be required to support the idea that these *Wolbachia* originated from an

ancestral New Zealand Isolated individual, due to the observed horizontal recombination of MLST fragments between strains of *Wolbachia* it is possible, albeit unlikely that the use of a single MLST fragment could create an artificial monophyletic clade. The distribution of strain B through the *H. maculifrons* complex was extensive with at least 11 confirmed *H. 'c*entral' hosts and 3 confirmed *H. 's*outh' hosts in addition to the 14 confirmed *H.* 'bruce' hosts. The suggested ancestral infection of strain B *Wolbachia* in New Zealand could account for the more even distribution among the ground weta however increased sequencing of the two under represented clades will answer this.

Comparing the variation among host COI sequences with the variation in the *Wolbachia* sequences obtained for the ftsZ locus indicates some correlation where clades within *H. maculifrons* match up with strains of *Wolbachia*. The observation that the hosts within each clade are likely to all exhibit the same strain of *Wolbachia,* and the shallowness of variation within each of the clades relative to the diversity between the clades, provides support for the idea of *Wolbachia* having an effect on the host's genetic diversity by influencing diversification between clades differently infected. It was observed that only one of the shallow clades exhibited both strains of *Wolbachia*. However, this is a small subset of individuals collected from this host species complex and in some instances a number of individuals were collected from the same site at the same time. This may result in individuals from a single site being closely related (due to sampling which would account for the low within clade diversity) and happen to have the same *Wolbachia* strain due to being spatially close to each other. An increase in both the number of *Wolbachia* sequences (both ftsZ and the other MLST regions) and *H. maculifrons* sequences will be required to determine in this patterning observed is due to the effect of *Wolbachia* on its host or is just a sampling artefact.

In addition to the *H. maculifrons* sequences, two infected individuals of North Island *H. 'alius'* were infected with strain A *Wolbachia*. Geographically these hosts were located within a grouping of 'clade A *H.* 'bruce'', this fits within the suggested idea that strain A *Wolbachia* has infected *H.* 'bruce' and spread before horizontally to other weta species. One suggested effect of the infection on *H. 'alius'* would be the separation from South Island *H.* 'alius' however the lack of differentiation between *H. 'alius' Wolbachia* and *H. maculifrons* complex *Wolbachia,* would suggest that the acquisition of *Wolbachia* was quite recent and unlikely to have had the time required to produce tangible effects on the host. The second suggested effect of the infection was to keep the *H.* 'alius' hosts isolated

from the other *Hemiandrous* species, again however as the samples tested were located geographically with hosts infected with the same strain of *Wolbachia* this effect would not occur.

Wolbachia was detected in a number of cave weta species however only a three individuals successfully sequenced the ftsZ region used in this analysis, two *Isoplectron* (GW1626, GW1635) and one *Pachyrhamma* (GW1871). Addition of the *Macropathus* HTS sample resulted in four sequences being added to the analysis (Fig. 4.1). This resulted in two individuals falling in to strain A, the two *Isoplectron*, and two individuals, into strain B, *Macropathus* and *Pachyrhamma*. The two *Isoplectron* hosts were collected from the same location separate from the other strain A hosts (Fig. 4.4) however our distribution is in no way complete. As with the two *Isoplectron* samples from strain A, the two hosts from strain B were also collected in the same area (Fig. 4.4), however their origin is different as they are different genera of cave weta. This shows that both strains of *Wolbachia* have managed to infect the Rhaphidophoridae family of weta. This spatial distribution of *Wolbachia* super groups is further supported by these, albeit small sample size, of cave weta, as well as the suggested role of horizontal transmission of *Wolbachia* among different species of nearby invertebrates.

In addition to the weta, which was the main focus of the genetic testing, a small number of extra New Zealand insects had their *Wolbachia* sequences added to the analysis. The first was an individual of the Psocoptera order, *Ectopsocus* sp., the inclusion does not change the overall picture very much but does provide another example of horizontal transfer of *Wolbachia* between unrelated New Zealand organisms. The second organism tested was *Archaeoteleia*, a parasitic wasp known for parasitizing New Zealand cave weta. Nine individuals were tested and four produced positive results, with three sequencing the required ftsZ region for analysis. All *Archaeoteleia Wolbachia* sequenced fell within one of the two groups of strain A. This provides a possible vector for the rapid movement of *Wolbachia* through the weta species despite a suspected more recent introduction of strain A *Wolbachia* into New Zealand. It also provides a potential vector for the movement of strain A *Wolbachia* into hosts already infected with strain B. As the primary hosts for these parasitic wasps is suspected to be cave weta further analysis of the species of cave weta will help identify the full effect *Archaeoteleia* is having on the movement of *Wolbachia* through New Zealand.

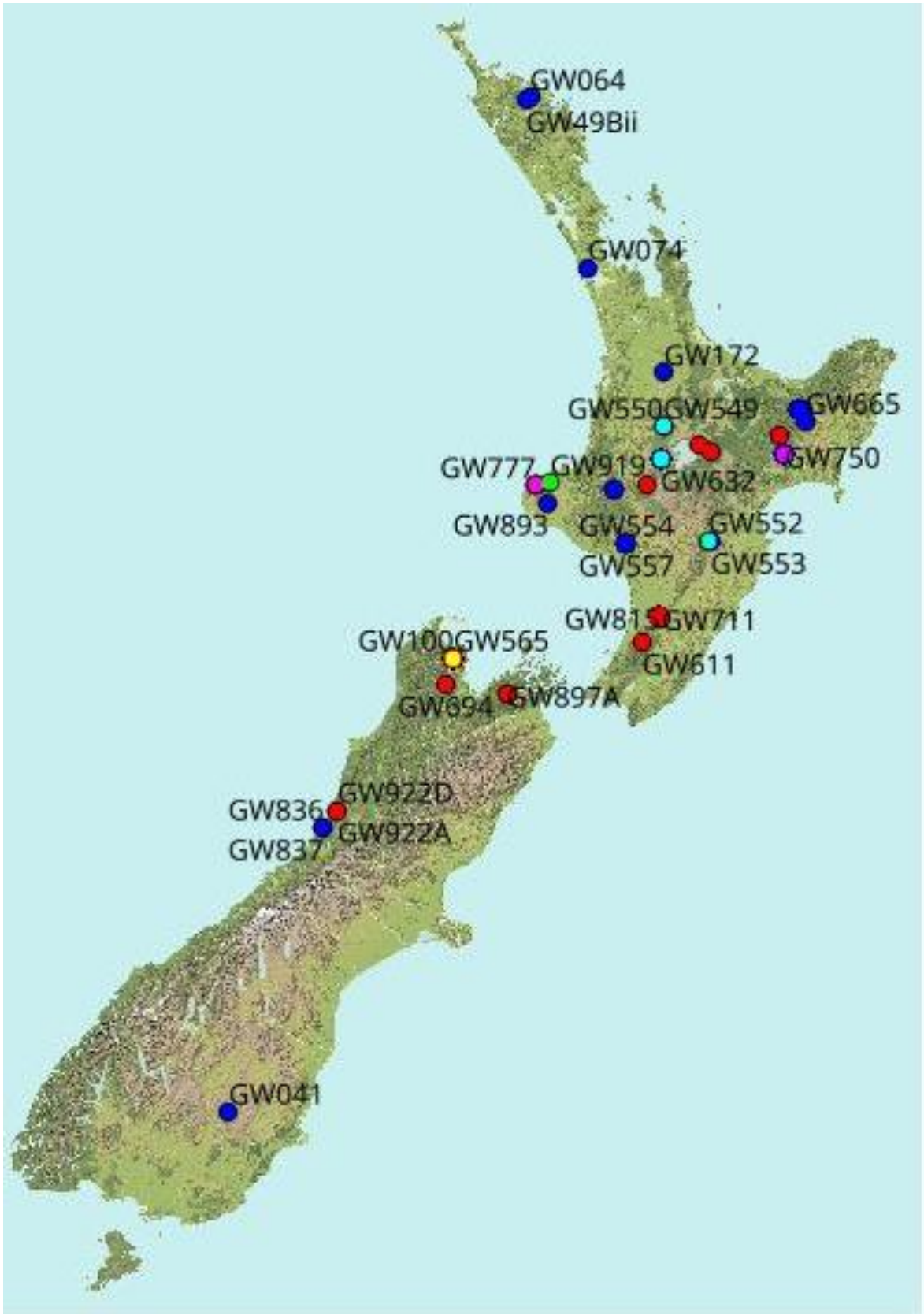


Figure 4.4 QGIS map showing locations of *H. maculifrons* indicating strain, Strain one blue, Strain two red, double infection green. Colours in between indicate locations where multiple strains are found (Teal = 1&3, Yellow = 2&3, and Purple = 1&2).

The distribution of the three groups varied across New Zealand with the newer invasion of *Wolbachia* (strain A) being present along the majority of the two main islands (Fig. 4.4). Strain B had a far more constrained distribution centralized around the southern North Island and the northern South Island (Fig. 4.4). This may be due to the limited number of individuals tested in the far north and south resulting an under representation of strain B in this analysis. If strain A is a more recent invasion and remains more spatially diverse it may indicate a more transferable strain of *Wolbachia*. This could be an effect of the *Wolbachia* itself, or due to the infection of a mobile vector such as *Archaeoteleia* (Vavre, et al., 1999; Ahmed, et al., 2015) which was shown to have exclusively strain A *Wolbachia* in the hosts tested in this small sample size. The third group consisting of the double infected hosts showed a similar pattern to that of strain B. This is unsurprising if we accept the likely cause of the double infection resulting from close proximity between hosts of both strain A and strain B. As strain B are the more restricted strain this would be the limiting factor on the range of double infected hosts, at least in the short term. If group three remains reproductively successful and isolated from the other two it is possible that movement of the hosts will result in a third distribution pattern unique to this group.

4.5 Conclusion

Through the use of molecular sequencing of the ftsZ gene of the MLST protocol, it was determined that New Zealand was host to two strains of *Wolbachia*, one from super group A, which was closely related to *Wolbachia* sequences found on GenBank. This clade from our analysis appears to consist of two separate groups of related *Wolbachia* potentially resulting from two separate invasions events possibly at different time points. The second clade was from super group B; this clade contains sequences exclusively from New Zealand hosts with a sister clade found in the Asia regions (a possible transmission route). In addition to these two clades there have been a number of hosts detected to have evidence of both clades, as *Wolbachia* introduces incompatibilities between different strains of *Wolbachia* in these hosts this double infection will introduce a third reproductively functional clade of *Wolbachia* in New Zealand invertebrates. Another test of this data was to determine if the *Wolbachia* had resulted in any noticeable changes to the host genetics and diversity. It was noticed there was a correlation between the clades formed within *H. maculifrons* and the *Wolbachia* found within those clades. This provides the first indication for the potential effect of *Wolbachia* on the genetic diversity of endemic New Zealand invertebrates.

4.6 References

Ahmed, M. Z. et al., 2015. The intracellular bacterium *Wolbachia* uses parasitoid wasps as phoretic vectors for efficient horizontal transmission. PLOS Pathogens, 10(2).

Baldo, L. et al., 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia* pipientis. Applied and Environmental Microbiology, Volume 76(11), pp. 7098–7110.

Bandelt, H., Forster, P. & Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, Volume 16(1), pp. 37-48.

Guindon, S. et al., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.. Systematic Biology, Volume 59(3), pp. 307-321.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & de Waard, J. R., 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B, Volume 270, pp. 313–321.

Hebert, P. D. N, Ratnasingham, S. & de Waard, J. R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London B, Volume 270, pp. S96-S99.

Huelsenbeck, J. P. & Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics, Volume 17, pp. 754-755.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, Volume 25(12), pp. 1647-1649.

Lo, N. et al., 2002. How many Wolbachia supergroups exist?. Molecular Biology and Evolution, Volume 19, pp. 341-346.

PopART., 2015. PopART, [Online] Available at http://popart.otago.ac.nz/index.shtml

QGIS Development Team, 2015. QGIS Geographic Information System.

Ronquist, F. & Huelsenbeck, J. P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, Volume 19, pp. 1572-1574.

Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, et al. (2012) Wolbachia and DNA barcoding insects: patterns, potential, and problems. PLoS ONE Volume 7(5)

Smith, T., 2016. Evolution of diversity: analysis of species and speciation in *Hemiandrus* ground wētā. PhD Thesis Massey University.

Valentini, A., Pompanon, F. & Taberlet, P., 2009. DNA barcoding for ecologists. Trends in Ecology & Evolution, Volume 25(2), pp. 110-117.

Vavre, F. et al., 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. Molecular Biology and Evolution, Volume 16(12), pp. 1711-1723.

Werren, J. H., Baldo, L. & Clark, M. E., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature, Volume 6, pp. 741-751.

Werren, J. H., 1997. Biology of *Wolbachia*. Annual Review of Entomology, Volume 42, pp. 587-609.

Zhou, W., Rousset, F. & O'Neill, S., 1998. Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. Proceedings of the Royal Society of London B, Volume 265(1395), pp. 509-515.

Appendix 1

Python coding for Chapter 2

Appendix 1: Python coding for Chapter 2

Appendix A.

```
from onding import FastqGeneralIterator #Biopython 1.51 or later

#############################################################
#
# Change the following settings to suit your needs
#

input_forward_filename = "corrected_FCC3PBYACXX-SZAXPI036570-34_L2_1.fq"
input_reverse_filename = "corrected_FCC3PBYACXX-SZAXPI036570-34_L2_2.fq"

output_paired_forward_filename = "out_forward_pairs.fastq"
output_paired_reverse_filename = "out_reverse_pairs.fastq"
output_orphan_filename = "out_unpaired_orphans.fastq"

f_suffix = "/1"
r_suffix = "/2"

#############################################################

if f_suffix:
    f_suffix_crop = -len(f_suffix)
    def f_name(title):
        """Remove the suffix from a forward read name."""
        name = title.split()[0]
        assert name.endswith(f_suffix), name
        return name[:f_suffix_crop]
else:
    def f_name(title):
        return title.split()[0]

if r_suffix:
    r_suffix_crop = -len(r_suffix)
    def r_name(title):
        """Remove the suffix from a reverse read name."""
        name = title.split()[0]
        assert name.endswith(r_suffix), name
        return name[:r_suffix_crop]
else:
    def r_name(title):
        return title.split()[0]

print "Scaning reverse file to build list of names..."
reverse_ids = set()
paired_ids = set()
for title, seq, qual in FastqGeneralIterator(open(input_reverse_filename)):
```

```
    reverse_ids.add(r_name(title))

print "Processing forward file..."
forward_handle = open(output_paired_forward_filename, "w")
orphan_handle = open(output_orphan_filename, "w")
for title, seq, qual in FastqGeneralIterator(open(input_forward_filename)):
    name = f_name(title)
    if name in reverse_ids:
        #Paired
        paired_ids.add(name)
        reverse_ids.remove(name) #frees a little memory
        forward_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
    else:
        #Orphan
        orphan_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
forward_handle.close()
del reverse_ids #frees memory, although we won't need more now

print "Processing reverse file..."
reverse_handle = open(output_paired_reverse_filename, "w")
for title, seq, qual in FastqGeneralIterator(open(input_reverse_filename)):
    name = r_name(title)
    if name in paired_ids:
        #Paired
        reverse_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
    else:
        #Orphan
        orphan_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
orphan_handle.close()
reverse_handle.close()
print "Done
```

Appendix B.

```
from Bio import SeqIO
import sys

try:
    infile = open(sys.argv[1])

except IOError:
    print "usage: python change_format.py infile.fq"
    sys.exit()

outfile_name = sys.argv[1]+".fas"
outfile = open(outfile_name,'w')

for rec in SeqIO.parse(infile,'fastq'):
    SeqIO.write(rec,outfile,'fasta')
```

```
infile.close()
outfile.close()
```

Appendix C.

```
f2 = open('Neonetus1.txt','r')
f1 = open('corrected_FCC3PBYACXX-SZAXPI036570-34_L2_1.fq.fas','r')
f3 = open('Neonetus1_1.fa','w')

AI_DICT = {}
for line in f2:
   AI_DICT[line[:-1]] = 1

skip = 0
for line in f1:
   if line[0] == '>':
      _splitline = line.split('|')
      accessorIDWithArrow = _splitline[0]
      accessorID = accessorIDWithArrow[1:-1]
      # print accessorID
      if accessorID in AI_DICT:
         f3.write(line)
         skip = 0
      else:
         skip = 1
   else:
      if not skip:
         f3.write(line)

f1.close()
f2.close()
f3.close()
```

Appendix D.

```
from Bio.SeqIO.QualityIO import FastqGeneralIterator #Biopython 1.51 or later

#############################################################
#
# Change the following settings to suit your needs
#

input_forward_filename                                                    =
"corrected_120223_I238_FCC0HH4ACXX_L5_SZAXPI005944-102_1.fq"
input_reverse_filename                                                    =
"corrected_120223_I238_FCC0HH4ACXX_L5_SZAXPI005944-102_2.fq"

output_paired_forward_filename = "out_forward_pairs.fastq"
output_paired_reverse_filename = "out_reverse_pairs.fastq"
output_orphan_filename = "out_unpaired_orphans.fastq"
```

```
f_suffix = "/1"
r_suffix = "/2"

############################################################

if f_suffix:
    f_suffix_crop = -len(f_suffix)
    def f_name(title):
        """Remove the suffix from a forward read name."""
        name = title.split()[0]
        assert name.endswith(f_suffix), name
        return name[:f_suffix_crop]
else:
    def f_name(title):
        return title.split()[0]

if r_suffix:
    r_suffix_crop = -len(r_suffix)
    def r_name(title):
        """Remove the suffix from a reverse read name."""
        name = title.split()[0]
        assert name.endswith(r_suffix), name
        return name[:r_suffix_crop]
else:
    def r_name(title):
        return title.split()[0]

print "Scaning reverse file to build list of names..."
reverse_ids = set()
paired_ids = set()
for title, seq, qual in FastqGeneralIterator(open(input_reverse_filename)):
    reverse_ids.add(r_name(title))

print "Processing forward file..."
forward_handle = open(output_paired_forward_filename, "w")
orphan_handle = open(output_orphan_filename, "w")
for title, seq, qual in FastqGeneralIterator(open(input_forward_filename)):
    name = f_name(title)
    if name in reverse_ids:
        #Paired
        paired_ids.add(name)
        reverse_ids.remove(name) #frees a little memory
        forward_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
    else:
        #Orphan
        orphan_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
forward_handle.close()
del reverse_ids #frees memory, although we won't need more now
```

```
print "Processing reverse file..."
reverse_handle = open(output_paired_reverse_filename, "w")
for title, seq, qual in FastqGeneralIterator(open(input_reverse_filename)):
    name = r_name(title)
    if name in paired_ids:
        #Paired
        reverse_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
    else:
        #Orphan
        orphan_handle.write("@%s\n%s\n+\n%s\n" % (title, seq, qual))
orphan_handle.close()
reverse_handle.close()
print "Done"
```

Appendix E.

```
#!/usr/bin/python
# encoding:utf8
# author: SÃŠbastien Boisvert
# part of Ray distribution
"""This script takes two fastq files and interleaves them

Usage:
    interleave-fasta.py fasta_file1 fasta_file2
"""

# Importing modules
import sys

# Main
if __name__ == '__main__':
    try:
        file1 = "out_forward_pairs.fastq"
        file2 = "out_reverse_pairs.fastq"
    except:
        print __doc__
        sys.exit(1)

    with open(file1) as f1:
        with open(file2) as f2:
            while True:
                line = f1.readline()
                if line.strip() == "":
                    break
                print line.strip()

                for i in xrange(3):
                    print f1.readline().strip()

                for i in xrange(4):
```

```
print f2.readline().strip()
```

Appendix F.

```
!/bin/bash
~/khmer/scripts/normalize-by-median.py -C 20 -k 17 -p -N 4 -x 2e9 inter_neonetus1.fq
~/khmer/scripts/load-into-counting.py -k 17 -N 4 -x 2e9 -T 4 table.kh inter_neonetus1.fq.keep
~/khmer/scripts/filter-abund.py -C 2 -T 4 table.kh inter_neonetus1.fq.keep
~/khmer/scripts/normalize-by-median.py -C 5 -k 17 -N 4 -x 2e9 inter_neonetus1.fq.keep.abundfilt
~/khmer/scripts/extract-paired-reads.py inter_neonetus1.fq.keep.abundfilt.keep
```

Appendix G.

```
#!/usr/bin/bash

#run velvet with multiple kmers

BASEOUT=/data/bjbridge/round3/data/neonetus1/


for i in $(seq 21 2 51); do
        echo "velveth ${BASEOUT}$i $i -fastq -short ${BASEOUT}inter_neonetus1.fq.keep.abundfilt.keep.se -short $ BASEOUT}inter_neonetus1.fq.keep.abundfilt.keep.pe";

        velveth ${BASEOUT}$i $i -fastq -short ${BASEOUT}inter_neonetus1*se -short ${BASEOUT}inter_neonetus1*.pe;

        echo "velvetg ${BASEOUT}$i -exp_cov auto -cov_cutoff auto -min_contig_lgth 200";

        velvetg ${BASEOUT}$i -exp_cov auto -cov_cutoff auto -min_contig_lgth 200;
done
```

Appendix H.

```
outfile = open("Neonetus2_stat.txt", "w")
f1 = open("Neonetus2_1.blastn.annot.txt", 'r')
f2 = open("Neonetus2_2.blastn.annot.txt", 'r')
total = sum(1 for line in open('Neonetus2.txt'))
Wolbachia=[]
NWolbachia=[]
Wol_list=[]
Double=[]
Vollenhovia=[]
for line in f1:
   split = line.split('\t')
   if split[0] in Wol_list:
      Double.append(split[0])
```

```
    else:
        Wol_list.append(split[0])
        if 'Wolbachia' in split[-1]:
            Wolbachia.append(split[-1])
        elif 'PREDICTED: Vollenhovia emeryi' in split[-1]:
            Vollenhovia.append(split[0])
        else:
            NWolbachia.append(split[-1])
for line in f2:
    split = line.split('\t')
    if split[0] in Wol_list:
        Double.append(split[0])
    else:
        Wol_list.append(split[0])
        if 'Wolbachia' in split[-1]:
            Wolbachia.append(split[-1])
        elif 'PREDICTED: Vollenhovia emeryi' in split[-1]:
            Vollenhovia.append(split[0])
        else:
            NWolbachia.append(split[-1])

Wol = len(Wolbachia)
NWol = len(NWolbachia)
Vol = len(Vollenhovia)
UnID = total - Wol - NWol - Vol
out = "Sequences identified as Wolbachia %i\nSeqyences identifyed as
Vollenhovia %i\nSequences not identified as Wolbachia %i\nSequences not
identified %i\n" %(Wol, Vol, NWol, UnID)
outfile.write(out)
for n in NWolbachia:
    outfile.write(n)

outfile.close()
f1.close()
f2.close()
```

Appendix 2

Hemiandrus maculifrons geographical distribution map for chapter 4

Appendix 2 - *Hemiandrus maculifrons* geographical distribution map for chapter 4

Appendix G

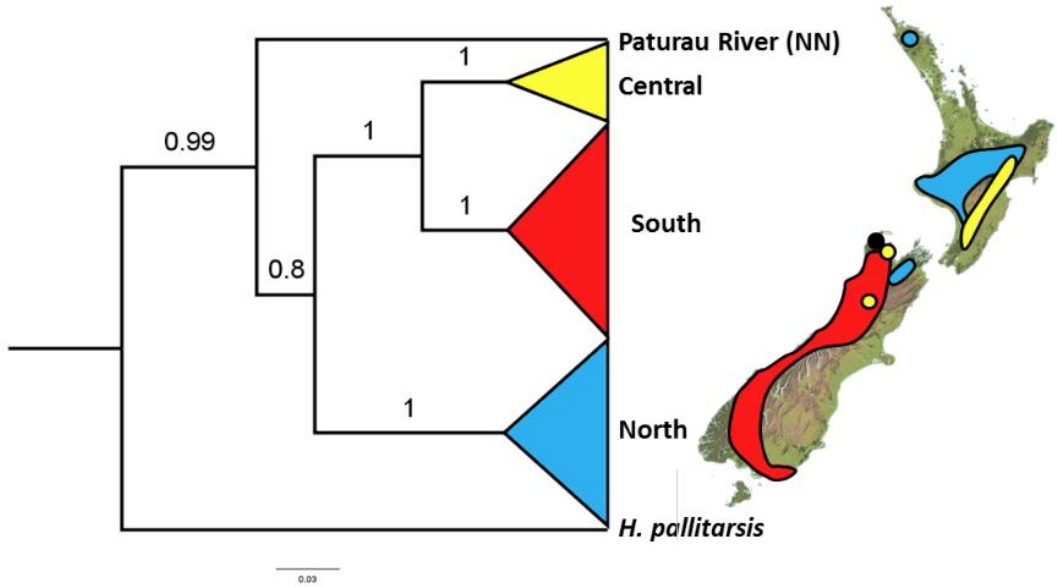


Figure 4.1. Bayesian mtDNA COI phylogeny from chapter 1 showing the relationships among species of the *Hemiandrus maculifrons* cryptic species complex, and their minimum distribution in New Zealand.

Geographical distribution of the three clades constituting the *Hemiandrus maculifrons* weta complex as extracted from (Smith, 2016)