THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

The interaction between diet breadth, geography and gene flow in herbivorous insects

Dean Robert Brookes

Bachelor of Environmental Sciences (Hons)

The University of Queensland

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2017

School of Biological Sciences

Abstract

The geographical distribution and life history of insects and their host plants determines when, and if, they will interact. The vast majority of herbivorous insects are host-plant specialists and their narrow host range limits their distribution and may also restrict gene flow between their populations. Generalist insect herbivores, by contrast, might be expected to have much higher gene flow across populations because the distribution of multiple host plant species is likely to be much more contiguous. Species distributions shift continually, so with enough time the populations of a specialist insect species should experience more events that fracture their distribution because of the changing distribution of their relatively fewer hosts. The changing distribution of insects and their host plants is thus important for mediating host-plant interactions and will also influence how and when adaptation to a host plant, or subset of host plants, occurs.

To investigate the influence of geography and host breadth on gene flow, and thus speciation, I explore the evolutionary history and population genetic structure of two herbivorous insects with different host plant relationships, the generalist bug *Nezara viridula* (Pentatomidae, Hemiptera), a global pest species, and the specialist thrips *Cycadothrips chadwicki* (Aeolothripidae, Thysanoptera), the pollinator in a brood-site mutualism with *Macrozamia* cycads. The results from these systems were then integrated with other similar data sets from phytophagous insects to understand better how gene flow and host specificity relate to one another, and to examine the role each has played (and may still play) in the evolution of insect-plant interactions more generally.

The pest bug *N. viridula* has two globally distributed mitochondrial lineages, the Asian (mtDNA lineage A) and the European (mtDNA lineage E), and the species status of these lineages has not been resolved. In Australia, I found that both of these *N. viridula* mitochondrial lineages were present. Microsatellite data show two genetically and geographically distinct populations, but these did not correspond directly to the two mitochondrial lineages. Together, mtDNA and microsatellite data shown that secondary contact has occurred in eastern Australia between the two lineages and has resulted in introgression of mtDNA lineage A into predominantly mtDNA lineage E populations in a narrow area, but without widespread introgression across the nuclear genome. Past research on this insect suggested that these lineages represented a single species because these mtDNA haplotypes were found together in single populations. The evidence presented here indicates that mating between lineages can occur, but that an analysis of mtDNA alone is not enough to understand the evolutionary history of these insects. Further, spatial and temporal genetics analyses of eastern Australian populations of *N. viridula*

show that they do not differ genetically across different host plants. Most host plants this insect feeds on in Australia have a narrower distribution than that of *N. viridula* and so gene flow is maintained between populations of *N. viridula* even in areas where most of its host plants are unavailable. Even substantial changes in the distribution of a number of potential host plants may not impact the population genetic structure of this insect significantly.

The Australian endemic thrips *C. chadwicki* is shown here to be a complex of at least five species, with some having diverged from one another about six Mya and others less than one Mya. Each thrips species is associated with a different subset of *Macrozamia* cycad species, with each found only on one or two *Macrozamia* species. One of these *Cycadothrips* species, *C. chadwicki* sp. 4, is found in southern Queensland where it is associated with *M. lucida* and *M. macleayi*. Population genetics analysis of this thrips across its geographic distribution shows high amounts of gene flow between populations, but also some weak spatial genetic structure. For *C. chadwicki*, more generally, these results support the presence of a species complex as no gene flow occurs over even short geographic distances between putative species in the *C. chadwicki* complex, even though a high amount of gene flow is observed within the single species *C. chadwicki* sp. 4. Geography appears important for speciation in *Cycadothrips* and *Macrozamia*, and so the conditions that lead to speciation may be the product of rare events that isolate both thrips and cycad populations. How much gene flow occurs between populations of *Macrozamia*, relative to the amount of gene flow among the associated *Cycadothrips* populations, requires further resolution.

The evolutionary history and population genetic structure of *N. viridula* and *C. chadwicki* fit with the general predictions of this study. An analysis of the population genetics literature involving phytophagous insects suggests that these results are consistent also for other herbivorous insect species, with relatively specialised phytophagous insects having lower amounts of gene flow between their populations. The approach developed in this thesis, as well as the insights from the study of *N. viridula* and *C. chadwicki*, will help understand the role that host breadth and geography play in the evolution of herbivorous insects.

Declaration by Author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer reviewed papers

Brookes, D. R., Hereward, J. P., Terry, L. I., & Walter, G. H. (2015). Evolutionary dynamics of a cycad obligate pollination mutualism–Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators. *Molecular phylogenetics and evolution*, 93, 83-93.

Conference abstracts

Brookes, D.R., Hereward, J.P., Wilson, L.J. and Walter, G.H., 2015. Gene flow and host use in the Green Vegetable Bug, *Nezara viridula*. Paper presented at the Australian Cotton Research Conference, Toowoomba, Queensland.

Brookes, D.R., Hereward, J.P., Wilson, L.J. and Walter, G.H., 2016. Gene flow and host use, relative to cotton, in *Nezara viridula*. Paper presented at the World Cotton Research Conference 6, Goiânia - Goiás, Brazil.

Conference abstracts for both of the above references can be found in Appendix 2.

Publications included in thesis

Chapter 4: Brookes, D. R., Hereward, J. P., Terry, L. I., & Walter, G. H. (2015). Evolutionary dynamics of a cycad obligate pollination mutualism–Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators. *Molecular phylogenetics and evolution*, 93, 83-93.

Contributor	**Statement of contribution**
Dean Brookes	Designed the study (50%), developed molecular tools (60%), conducted sampling (95%), conducted molecular analysis (80%), performed statistical analysis of phylogenetic (100%) and microsatellite data (70%), wrote and edited the paper (60%).
James Hereward	Designed the study (10%), developed molecular tools (40%), conducted molecular analysis (20%), performed statistical analysis of microsatellite data (30%), wrote and edited the paper (15%).
Irene Terry	Designed the study (10%), conducted sampling (5%), wrote and edited the paper (10%).
Gimme Walter	Designed the study (20%), wrote and edited the paper (15%)

Contributions by others to this thesis

Professor Gimme Walter and Dr. James Hereward made significant contributions to the initial design of the project and provided guidance throughout. Dr. James Hereward contributed to the decision making regarding the development of each of the molecular tools that are included in this thesis. Dr. Lewis Wilson contributed to the structuring of the sampling program of *Nezara viridula* and provided numerous friendly and useful contacts toward that end, as well as discussions throughout. Irene Terry contributed to the published manuscript included as Chapter 4.

Statement of parts of the thesis submitted to qualify for the award of another degree

A significant amount of the work involved in the publication included in Chapter 4 was submitted for the qualification of an honours degree that was awarded by The University of Queensland in 2012:

> Brookes, D.R., Hereward, J.P., Terry, L.I. and Walter, G.H., 2015. Evolutionary dynamics of a cycad obligate pollination mutualism–Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators. *Molecular phylogenetics and evolution*, *93,* pp.83-93.

This publication is included here because my honours research was extended significantly in the early stages of my PhD program. The part of the research conducted during my PhD program provided the basis for the research carried out in Chapter 5. In particular, the research associated with this publication was the precursor to the research carried out in Chapter 4. The following are the specific contributions made to this publication before and after commencement of my PhD program:

> *Before PhD commencement:* As a part of my honours project insect collections were made from all but two of the sites that were represented in the final publication. PCR protocols for amplification of the 28S and COI genes were developed and insects from all but two sites were sequenced at these two gene regions. Primers were also developed for four of the microsatellite loci used in the final publication. These four microsatellite loci were used to genotype *Cycadothrips* from two populations. Following submission of my honours thesis, but before commencement of a PhD, the remaining six microsatellite loci were developed as outlined in the following publication:

Brookes, D.R. and Hereward, J.P., 2013. Characterisation of 18 microsatellite markers for the thrips *Cycadothrips chadwicki* (Thysanoptera: Aeolothripidae), an obligate mutualist pollinator of 11 *Macrozamia* cycad species. *Conservation Genetics Resources, 5*(4), pp.1167-1170.

During PhD: Thrips were collected from two further locations. *Cycadothrips* from six populations were genotyped at 10 microsatellite loci. Comprehensive phylogenetic, timing of divergence, and individual population assignment analyses were carried out. Synthesis of the relevant scientific information, writing of the manuscript, and submission for publication were then undertaken during the course of my PhD program.

Acknowledgements

I am grateful to each of my supervisors for their respective contributions to my personal development and to this thesis. I am particularly grateful to Prof. Gimme Walter whose dedication to his students has allowed me to improve personally and professionally in ways I could not have foreseen. Without the efforts of Dr. James Hereward I would not have the same confidence in the technical skills that I gained through the development of this thesis. Dr. Lewis Wilson helped me direct the research on *Nezara viridula*, especially during the difficult early period of the project, and has also provided useful direction throughout. I would also like to thank two examiners of this thesis for taking the time to provide comments that have improved this thesis.

I am indebted to the Cotton Research and Development Corporation (CRDC) for funding both my research on *N. viridula* and travel to World Cotton Research Conference 6 in Brazil. By working through the challenging aspects of this research, and my interactions with the cotton industry have helped my professional development immensely. Susan Maas in particular provided useful direction at critical points during this research.

Obtaining samples of *N. viridula* was perhaps the most difficult and time-consuming aspect of this research. Many researchers, agronomists, and growers across Australia have helped me locate these bugs with their regional expertise. The following people in particular contributed significant amounts of their time when they were under no obligation to do so. Tanya Smith shared her experience locating these bugs very early on in the project when I was having difficulty finding them. Geoff Hunter, Renee Anderson, and Heath McWhirter helped locate growers likely to have *N. viridula*, and so helped me to better understand their respective growing regions. Luke Halling, Sally Cowan, Dr. Mary Finlay-Doney, and Alan Boulton mailed *N. viridula* samples to me that I would not have otherwise obtained.

A number of people helped me develop skills and ideas throughout. Associate Professors Cynthia Riginos and Mike Furlong were my expert readers throughout my candidature. Their feedback at each thesis milestone was valuable and helped to shape the overall structure and direction of the research. Doctors Alicia Toon, Graham McCulloch, and Sharon van Brunschot each helped me develop skills in one area or another, and alongside Justin Cappadonna, Lauren Kaye, and Renee Rossini, provided useful discussions.

I have been fortunate to work with Professors Irene Terry and Bob Roemer each year when they come to Australia to work on *Macrozamia* and *Cycadothrips*. Their devotion and approach to their research is inspiring and I look forward to their arrival each year.

Professor Antônio Panizzi allowed me to visit his lab during my trip to Brazil. Discussions with Prof. Panizzi have allowed me to think more broadly about pentatomids. The hospitality of Prof. Panizzi and his students as well as the enthusiasm they have for their research was also a highlight of my stay.

Luke Halling, John Westaway, and Jane Ray each shared their expertise during my sampling trip to Kununurra and Darwin.

Finally, thank you to my friends and family for their support throughout.

Keywords

Phylogenetics, population genetics, biogeography, codiversification, gene flow, host breadth, insect, speciation.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060411, Population, Ecological and Evolutionary Genetics, 70%
ANZSRC code: 060302 Biogeography and Phylogeography, 15%
ANZSRC code: 070308 Crop and Pasture Protection (Pests, Diseases and Weeds), 15%

Fields of Research (FoR) Classification

FOR code: 0604 – Genetics, 70%
FOR code: 0603 – Evolutionary Biology, 15%
FOR code: 0703 – Crop and Pasture Protection, 15%

Table of Contents

List of Figures

S2.2. *Eurydema gebleri* and *Oechalia schellenbergii* (both also Pentatomidae) were included, with *O. schellenbergii* designated as the outgroup. The label ‘Other’ denotes a haplotype of *N. viridula*, common to individuals sampled from Greece, Guadeloupe, Galapagos, California, Japan, Italy and Brazil, from the Kavar (2006) dataset. Posterior probabilities are shown as node labels.

Figure S2.2. Tree representing the phylogeny of *Nezara viridula*, based on a MrBayes 3.2.6 analysis and using the 348 bp COI gene dataset (SF in the text) from all available individuals of both *N. viridula* and *N. antennata* (Table S2.2). Only unique sequences were used and if the number of samples that shared a sequence is greater than one then the number is shown in square brackets. The primary source of these sequences is this study and the DNA barcoding projects from various countries, specifically outlined in Table S2.2. *Eurydema gebleri* and *Oechalia schellenbergii* (both also Pentatomidae) were included, with *O. schellenbergii* designated as the outgroup. Posterior probabilities are shown as node labels.

Figure S2.3. Haplotype networks for the EF1α and Tubα1 genes sequenced for individuals of *Nezara viridula* collected in Australia, with haplotype colour matched to the mitochondrial lineage to which each individual belongs (see key). Haplotypes that correspond to the Asian mtDNA lineage are further subdivided into those from individuals collected from northern Australia and those from eastern Australia (Figure 2.2). A single sample represents one phase of the nuDNA of a single individual as estimated by DnaSP 5.1 (Librado & Rozas 2009).

Figure S2.4. The results of the STRUCTURE analyses for genotyped Australian *Nezara viridula* individuals (Table S2.1) for K=2 through K=4 and labelled according to their sampling location. Sample sites are organised into two groups, the northern populations (Kununurra and Darwin) and eastern populations (all others to the right) (Figure 2.2). The eastern sites are then organized in geographical sequence from north to south. K=2 was the most likely value of K as per the Evanno (2005) method.

Figure S2.5. Output from STRUCTURE Harvester showing the likelihood of K (a) and Delta K (b). This indicates that K=2 is the most likely value of K.

Figure 3.1. A map of genotyped *Nezara viridula* individuals grouped according to region and the season they were sampled. Populations represented by the STRUCTURE analysis boxes are labelled Y1 if collected during the 2014-2015 season, and Y2 if collected during the 2015-2016 season. A further label of ‘a’ or ‘b’ designates samples taken at different times within a season. Site names represent broad localities within regions. The scatter plot shows a Discriminatory Analysis of Principle Components (DAPC) with populations grouped according to region regardless of sampling time. The boxes represent a K=2 STRUCTURE analysis of the same genotyped individuals used in the DAPC.

Figure 3.2. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* individuals (Table 3.2), with individuals grouped according to sample location irrespective of time sampled. Geographic distance is used in Figure 3.1A and log transformed geographic distance is used in Figure 3.1B.

Figure 3.3. Within-region analyses of genotyped *Nezara viridula* individuals using Discriminatory Analysis of Principle Coordinates (DAPC) and K=2 STRUCTURE analyses. *Nezara viridula* individuals are grouped within regions according to the season in which they were sampled. Populations sampled during the 2014-2015 season are labelled Y1, and those from the 2015-2016

season labelled Y2. Labels with an 'a' or 'b' indicate population samples taken at different times within a season. See Figure 3.2 for these same analyses but conducted with all populations, rather than within regions, and for a map of the sample locations. The boxes of STRUCTURE analyses here are coloured the same as in Figure 3.2.

Figure S3.1. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* collected in late 2014 and early 2015 (see Table 3.1). Geographic distance is used in Figure S3.1A and log transformed geographic distance is used in Figure S3.1B.

Figure S3.2. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* individuals collected in late 2015, early 2016, and from northern Queensland in 2014 (see Table 3.1). Geographic distance is used in Figure S3.1A and log transformed geographic distance is used in Figure S3.1B.

Figure 4.1. The distribution of all 13 species of *Macrozamia* cycad that are pollinated by *Cycadothrips* is represented in these maps of Australia (see key and inset), developed from records obtained through the Atlas of Living Australia (www.ala.org.au) and Queensland Herbarium (Brisbane). Populations of these cycads are patchy, so do not extend over the entire geographical range depicted for each species. The distributions of *M. macleayi* and *M. lucida* overlap in some areas so their population boundaries are not as discrete as depicted. Sites at which *C. chadwicki* were collected for genetic analysis are indicated by black dots. Note that *C. emmaliami* and *C. cardiacensis* were not sampled.

Figure 4.2. Map of eastern Australia to show the association of the 59 COI haplotypes (networks left of vertical line) and four 28S haplotypes (networks right of vertical line) of *Cycadothrips chadwicki* relative to their host *Macrozamia* species. For *C. albrechti,* both the five COI and one 28S haplotypes are shown in dark grey. Each *Macrozamia* species is indicated by a different colour (same as that used in the key to host species and their distribution in Figure 4.1) with respect to the sample locality on the map (see below) and in relation to the thrips collected from them (in the networks). The COI haplotypes of *C. chadwicki* fall into one of three main clades, each of which has its network plotted alongside the map. For both genes, the small black dots represent haplotypes for which samples were not obtained, and the size of each haplotype circle corresponds to the number of individuals with that haplotype (see key in bottom right hand corner of figure). All three *C. chadwicki* networks are joined by a central node, dotted lines with a number represent the number of nodes to this central node. In the map, shading on the distribution of each *Macrozamia* species represents the plants with which the three main thrips clades (north (cross hatch), central (dashed lines) and southern (dotted), are associated. The letters A, B and C associated with different *M. macleayi* populations on the map are cross-referenced to both haplotype networks. For the extent of divergence between these clades also see Figure 4.3.

Figure 4.3. Chronogram of *C. chadwicki* samples by host plant. The tree is a maximum clade credibility tree based on COI and 28S data (BEAST 1.8.0) and adapted from FigTree 1.4 outputs. The root is calculated from the uncorrelated molecular clock. Coloured blocks represent the *Macrozamia* species from which each individual was sampled as per the key to host species (and these colours are the same as those used in Figures 4.1 and 4.2). Haplotypes shared across *Macrozamia* species indicate presence and absence so are not coloured proportionately. The letters on the *M. macleayi* blocks correspond to the population from which these thrips were sampled (Figure 4.2). The scale bar represents the number of years (in millions) represented by each branch. The posterior probabilities are displayed as node labels and the 95% Highest Posterior Densities (95% HPD) are shown as branch labels. The values for nodes (i) through (ix) are listed in Table S2.

Figure 4.4. Map showing the distribution of the *Macrozamia* populations represented in Figure S4.5. The distribution of each *Macrozamia* species is represented by a different colour (see key on right of figure). The distribution of *M. lucida* overlaps with much of the distribution of *M. macleayi* and hybridization is thought to occur in some locations. Each box shown represents the individuals included in the K3 STRUCTURE analysis and they are presented separately to show the origin of the insects relative to a particular host plant species (Figure S4.5).

Figure S4.1. Chronogram from a BEAST analysis using a COI substitution rate of 0.0145 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

Figure S4.2. Chronogram from a BEAST analysis using a COI substitution rate of 0.0112 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

Figure S4.3. Chronogram from a BEAST analysis using a COI substitution rate of 0.0178 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

Figure S4.4. STRUCTURE HARVESTER results showing mean maximum likelihood (left) and Delta K (right). Delta K indicates a hierarchical structure with K=2 as the most likely value of K, although K=3 also has considerable support.

Figure S4.5. This Principle Coordinate Analysis (PCoA) shows ordination of genetic identity between different populations of *Cycadothrips* individuals from the central clade (Figures 4.2 and 4.3) and the *Macrozamia* species from which they were collected. The same populations are included in this analysis as those that were used in the STRUCTURE analysis displayed in Figures 4.4 and S4.6. Thrips from different host species are coded by colour and different populations of those host species are represented by different symbols (see key above the plots). In PCoA 1 (top), the first and second axes explain 41.4% and 15.7% (12.5% third) of the variation respectively. In PCoA 2 (middle), the first and second axes explain 34.1% and 17.1% (14.7% third) of the variation respectively. In PCoA 3 (bottom), the first and second axes explain 23.8% and 18.9% (16.2% third) of the variation respectively.

Figure S4.6. Results of the STRUCTURE algorithm illustrate the posterior probability of an individual's assignment to one of K clusters (given on the left of each diagram) based on eight microsatellite loci screened across 287 *C. chadwicki* individuals from eight populations and four *Macrozamia* species. Analyses were performed using the admixture model with allele frequencies correlated, a burn-in of 100,000 iterations and a further 1,000,000 iterations after burn-in. Each vertical bar represents one individual, comprised of one or more colours, representing the posterior probability that each individual belongs to each of the sampled populations. The K-value associated with each iteration represents the number of hypothetical populations tested for that STRUCTURE run.

Figure 5.1. A topographic map of south-east Queensland showing the locations in which individuals of *Cycadothrips chadwicki* sp. 4 were collected (Table 5.1) together with the results of a K=4 STRUCTURE analysis (Figure S5.2). The latter is displayed in coloured boxes associated with each collection site. The results of a pairwise exact G-test to assess genic population differentiation (Table S5.1) are represented as lines connecting populations. The green lines represent pairwise comparisons that were not significantly different across populations and the yellow lines represent ones that were

moderately or nearly significant (p = >0.01 and <0.06). Pairwise interactions of populations sampled in 2011 are not shown. Boxes alongside one another represent two independent collections from the same locations but made in different years (Table 5.1). All thrips were collected in 2015 except for those labelled with the numbers 11 (2011) and 13 (2013) in their site codes.

Figure 5.2. Discriminatory Analysis of Principle Components (DAPC) using microsatellite data of all genotyped *Cycadothrips chadwicki* sp. 4 (Table 5.1). Individuals are coloured to indicate the northern and southern groups, as well as site BSF, all of which were identified by STRUCTURE analysis and exact G-tests (Figure 5.1). Site LSF was grouped separately because it a geographical outlier, and to observe its relationship to other populations. The number of discriminatory axes that were retained from the analysis was 3.

Figure 5.3. The results of two isolation by distance analyses (IBD) using Mantel tests performed on the *Cycadothrips chadwicki* sp. 4 genotype data. Analysis A includes all populations that were sampled during 2015 and analysis B excludes the single population that was a genetic outlier with respect to its geographical location, BSF (Figure 5.1). In both cases significant isolation by distance was detected with low to moderately high R^2 values.

Figure S5.1. The output from STRUCTURE harvester showing Delta K (top) and mean of the likelihood of K (bottom). K=4 had the highest value of Delta K (16.5) but the value for K=2 was also high (14.4).

Figure S5.2. The results of the STRUCTURE analyses for K=2 through K=5. K=5 is a major cluster (11/20). SRUCTURE harvester indicated that K=4 was the most likely value of K (Figure S5.1).

Figure S5.3. The cycad populations sampled and represented in Figure 5.1 are mapped onto this map with satellite imagery (ArcMap 10.2 basemap) to visualize the connectivity of the populations with remaining vegetation (following clearing of the landscape). Red dots are sites where individuals of *Cycadothrips chadwicki* sp. 4 were sampled from cycads for use in this study (Table 5.1). The following represent host records for the three *Macrozamia* species sampled from (HERBRECS database); teal squares – *M. macleayi*, blue triangles – *M. lucida*, and green triangles – *M. longispina*. Other populations are known for which host records do not exist, but these represent the majority of sites.

Figure 6.1. IBD relationships (from IBDWS) for *Nezara viridula* are shown. All Australian *N. viridula* populations are included in the top figure (a) (see Chapter 2) and the eastern Australian populations are included below (b) (this is Figure 3.2b - see Chapter 3). IBD R^2 values are 0.54 (a) and 0.39 (b).

Figure 6.2. IBD relationships (from IBDWS) for *Cycadothrips chadwicki sensu lato.* All central clade *C. chadwicki* species are included in the top figure (a) (those from Chapter 4) and individuals of *C. chadwicki* sp. 4 are included alone in the figure below (b) (this is Figure 5.3a - see Chapter 5). IBD R^2 values are 0.66 (a) and 0.12 (b).

Figure S6.1. IBD relationship (from IBDWS) for *Bombus morio* (Francisco *et al.* 2016) prior to the populations IVIT and TSAM being removed in the top figure (a) and after their removal below (b).

Figure S6.2. IBD relationships (from IBDWS) for *Graellsia isabellae* (Marí-Mena *et al.* 2016) in the top figure (a), and for *Rosalia alpina* (Drag *et al.* 2015) below (b).

List of Tables

Table S2.5. Allelic richness of *Nezara viridula* populations from northern Australia (141 individuals from northern NT and northern WA) and those populations from eastern Australia (383 individuals from all other regions but excluding populations LHR and GRU (Figure 2.2)). The proportion of the total number of alleles for each group is show in brackets.

Table 3.2. Groupings of sample sites for the genetic analysis of *Nezara viridula*. Population names represent multiple samples taken from nearby sites at similar times. The sites and host plants that correspond to each population name can be found in Table S3.1. Population names with an asterisk are spatial samples only. Regional samples include total the following number of sites and individuals – (i) northern QLD, two populations and 83 individuals, (ii) central Queensland, five populations and 146 individuals, (iii) southern QLD, four populations and 129 individuals, (iv) central NSW, six populations and 157 individuals, and (v) southern NSW, four populations and 144 individuals. The column 'n' shows the number of *N. viridula* genotyped from each population. Sample labels with state abbreviations are as follows: NSW – New South Wales, and QLD – Queensland.

Table 3.3. The results of an analysis of *Nezara viridula* microsatellite genotypes using the software Bottleneck 1.2.02 to perform Wilcoxon tests for heterozygosity excess (one-sided p) and for both heterozygosity excess and heterozygosity deficiency (two-sided p). Time sampled is divided into three categories, early (E – August, September, and October), middle (M – November, December, January), and late (L – February, March, April). In this regard, populations TOW and BOW are excluded because of regional climatic differences and BRZ is excluded because the samples that constitute that population do not represent a single sample time. Bottleneck parameters were a variance of 0.36 and using both the two-phase mutation model (TPM) and single step mutation model (SMM). The TPM used a single-step mutation proportion of 95%.

Table 3.4. AMOVA were conducted to assess that partitioning of *Nezara viridula* genetic differences, primarily among populations and among two seasons. Global locus by locus AMOVA results are presented for an F_{ST}-like analysis conducted in Arlequin (Excoffier & Lischer 2010). AMOVA a) includes all populations of *N. viridula* (Table 3.2), AMOVA b) includes populations from regions that were sampled in both the 2014-2015 and the 2015-2016 seasons (so excluding TOW and BOW) and with populations separated into groups according to season.

Table S3.1. Collection information related to the *Nezara viridula* genotyped, including locality, date, and host plants. The site code links the information in this table to raw data files. Samples at all sites were collected by D. Brookes. Entries with multiple host plants represent sites where insects were collected from multiple host plants growing immediately next to one another, and so plants could not be sampled independently of one another. Insects from site codes with an asterisk, totalling 314 individuals, were genotyped previously and the results included in analyses presented in Chapter 2. Insects from all other site codes, totalling 240 individuals, were genotyped for the aims of Chapter 3. The column 'n' shows the number of individuals genotyped at the microsatellite loci as described in Appendices 2.2 and 2.3.

Table 3.1. Statistics of the microsatellite loci included in the population genetic analysis of *Nezara viridula.* Only samples collected from the 17 populations with 19 or more individuals were used to generate these statistics (see Table 3.1), with the exception of the total number of alleles (∑Na).

Table 3.2. Pairwise exact G-test results for genic differentiation across temporally separated populations of *Nezara viridula* collected in the 2014-2015 and 2015-26 seasons (Table 3.2), with lines

separating the regions. Non-significant pairwise associations are highlighted green. Population GR-Y2 is excluded for having too few sampled individuals.

Table 4.1. Statistics of the microsatellite loci included in the population genetic analysis of *Cycadothrips chadwicki.*

Table 4.2. All known *Cycadothrips* species, including the cryptic species revealed in the current study and the evidence used to designate them as such. The *Cycadothrips* populations not sampled are indicated as such in the left hand column.

Table S4.1. Host plant species, their pollinators and sampling localities for *Cycadothrips* samples included in this study. Those species pollinated by both thrips and *Tranes* weevils are indicated. The exact localities of many *Macrozamia* species are considered sensitive information in Queensland and so have not been specified precisely. *Macrozamia macleayi* is associated with different thrips clades in different areas (see results), so it is labelled A, B and C for ease of cross-referencing to the text and Figure 4.2. NP = National Park; SF = State Forest. The QDPC and Genbank code identifies the host population that each of the individuals is associated with.

Table S4.2. Posterior probability, date estimates and 95% HPD for nodes (i) through (ix) for the chronograms shown in Figures 4.3, S4.1, S4.2 and S4.3.

Table S4.3. Pairwise F_{ST}s calculated using microsatellite data in Genepop 4.2 between central clade thrips populations as shown in Figures 4.4 and 4.5.

Table 5.1. Summary of sample information for the 583 *Cycadothrips chadwicki* included in this study. Some entries represent two independent collections from the same location made in different years (see text). *Macrozamia* cycads are protected species in Queensland and so specific site information has been withheld, but see map (Figure 5.1) for the distribution of the sites sampled. Pairwise geographic distance can be found in Table S5.2.

Table 5.2. Summary statistics for the microsatellite loci genotyped for *Cycadothrips chadwicki* sp. 4 individuals (Table 5.1). All loci are tetra-nucleotides (Brookes & Hereward 2013).

Table 5.3. Locus by locus AMOVA results for four different analyses conducted in Arlequin (Excoffier & Lischer 2010), each with F_{ST}-like and R_{ST}-like analyses. AMOVA a) includes all populations. AMOVA b) includes two host plant groups, population pairs of *M. lucida* or *M. macleayi*. AMOVA c) includes only the four populations that were sampled in both 2011 and 2015, and so year sampled is included. AMOVA d) includes only the two groups of populations that had no significant pairwise differences, the northern and southern groups (Figure 5.1).

Table 5.4. The results of analyses using the software Bottleneck 1.2.02 to perform Wilcoxon tests for heterozygosity excess (one-sided p) and for both heterozygosity excess and heterozygosity deficiency (two-sided p) in *Cycadothrips chadwicki* sp. 4 genotype data. These values were computed using a two-phase mutation model (TPM) with 0.36 variance and using three different single-step mutation proportions; 70%, 80%, and 95%.

Table S5.1. Pairwise exact G-test for genic differentiation. For samples taken during 2015, non-significant pairwise associations are highlighted green, moderately or nearly significant associations (p

= >0.01 and <0.06) are highlighted in yellow, and both are displayed in Figure 5.1 using lines of the same colour to connect populations, so as to display the relationship between populations graphically. Highly significant associations are indicated with an asterisk. Samples found below the horizontal line are from the temporal dataset only and are not shown in Figure 5.1.

Table S5.2. Pairwise F_{ST}s are presented below the diagonal and above it is the pairwise geographic distance between populations in kilometres values (km). All thrips were collected in 2015 except for those labelled with the numbers 11 (2011) and 13 (2013) in their site codes.

Table 6.1. Results from ANOVA analyses of the relationship between host specificity indices at different taxonomic levels (of host plant), and various independent variables, including IBD Slope, IBD R^2, mean pairwise F_{ST} (labelled as mean F_{ST}), and highest pairwise F_{ST} (labelled as highest F_{ST}). Mean pairwise geographical distance (labelled as geo. distance) was included alongside each other independent variable in each ANOVA, as was the interaction between them. Degrees of freedom were one in all cases, and the sum of squares, F value, and p value are reported for the ANOVAs. Significance levels for p values are '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.10 ' '. Significant and near significant values are shown in bold. Eta squared is used as a measure of effect size.

Table 6.2. For studies that had mitochondrial sequence data together with microsatellite data the following information is shown - species name and host breadth (in square brackets as either M – monophagous, O – oligophagous, and P – polyphagous), highest pairwise F_{ST}, highest pairwise distance between mtDNA haplotypes (%), and gene region sequenced. Highest mtDNA pairwise distance was obtained manually using sequences from Genbank that were associated with each study. Sequences were aligned using MAFFT and then the highest distance between any two haplotypes was calculated. Mean pairwise FST and Highest pairwise FST are shown without transformations.

Table S6.1. Details of the studies included in the analysis presented here that were found using the Web of Science database. Study species, study authors, Digital Object Identifier (DOI), and notes relevant to how the data from each study was incorporated into the data set presented here.

Table S6.2. Columns G, F, O, and M, are weighted host specificity values for Genus, Family, Order, and the Mean of these three weighted host specificity values respectively. The first time the weighted host specificity of each insect passes 80% at a taxonomic level it is considered a specialist at that level. Host use is designated into monophagous (host specific at the level of Genus), oligophagous (host specific at the level of Family), or polyphagous (not host specific at the levels Genus or Family) according to weighted host specificity.

Table S6.3. The species from each study are classified into monophagous, oligophagous, or polyphagous and comparison made between them. Mean geographic distance, and its standard deviation, are shown for the included studies. Mean distance is calculated a second time after excluding studies that exist outside the standard deviation of geographic distance. The Order to which insects from each host breadth category belong is shown, with the number of species from each Order shown in brackets.

List of Included Abbreviations

28S (gene) – 28S ribosomal RNA
AGRF – Australian Genome Research Facility
bp – Base pairs
CABI – Centre for Agriculture and Bioscience International
COI (gene) – Cytochrome C oxidase subunit 1
COII (gene) – Cytochrome C oxidase subunit 2
Cytb (gene) – Cytochrome b
DAPC – Discriminatory analysis of principle components
DNA – Deoxyribonucleic acid
DOI – Document online identifier
EF1α (gene) – Elongation factor 1 alpha
GVB – Green vegetable bug
HPD – Highest posterior density
HWE – Hardy-Weinberg equilibrium
IBD – Isolation by distance
mtDNA – Mitochondrial DNA
NAQS – North Australian Quarantine Strategy
NSW – New South Wales
NT – Northern Territory
nuDNA – Nuclear DNA
PCoA – Principle coordinates analysis
PCR – Polymerase chain reaction
QDPC – Queensland Department of Agriculture, Fisheries and Forestry Insect Collection
QLD – Queensland
RNA – Ribonucleic acid
Tubα1 (gene) – Tubulin alpha 1
WA – Western Australia

Chapter One – Introduction

1.1 General background

The vast majority of herbivorous insects are host-plant specialists (Joshi & Thompson 1995) and their reliance on one plant resource may limit their distribution and their population connectivity (Devictor *et al.* 2008; Groot *et al.* 2011; Östergård & Ehrlén 2005). Host plant specialisation may also favour insect speciation through host shifting (Jousselin *et al.* 2008; Kawakita & Kato 2009; Machado *et al.* 2005; Smith *et al.* 2008a), or co-diversification with their hosts in allopatry (Althoff *et al.* 2012; Smith *et al.* 2008b), all ultimately driven by changes in the geographical distribution of the plant and insect. Conversely, generalist insects may have populations that are more highly connected over greater geographic distances through their potential association with a greater number of hosts. However, the specific host associations of generalists may differ spatially and temporally (Velasco & Walter 1992). As a result the processes that may be responsible for speciation in generalist insects are also less clear, and will be less dependent on any one host plant when compared with a specialist. Most speciation occurs following the geographical separation of populations (Mayr 1942; Mayr 1970; Futuyma *et al.* 1980; Fitzpatrick *et al.* 2009). This means that host breadth, and how this affects the distribution of insect populations and their movement among hosts, is crucial for understanding how insect-plant interactions evolve.

Theoretical proposals attempt to explain the evolution of insect host plant interactions and insect host specificity is often central to the explanation. However empirical evidence for the various theories is often mixed. The coevolutionary process is perhaps the best example to illustrate this. Here, reciprocal adaptation between herbivorous insects and their host plants is said to drive the evolution of highly specific insect-plant interactions, with this also considered to potentially lead to speciation (Ehrlich & Raven 1964; Thompson 2005; Thompson & Cunningham 2002). Coevolution is said to be particularly important in obligate pollination mutualisms, in which the plant is reliant on a (usually) single specialist insect for pollination and the insect, in turn, relies on the plant for larval and adult resources. However, support is poor for a general role for coevolution in these systems (Althoff *et al.* 2014; Hembry *et al.* 2014; Nuismer *et al.* 2010; Yoder & Nuismer 2010). This realisation has been achieved through thorough investigation of the exact species associations of obligate pollination mutualisms, as well the specific biology, ecology, and evolutionary history of the interacting organisms. Only then have accurate inferences about the processes that have driven their evolution been made (Althoff *et al.* 2014; Ehrlich & Raven 1964; Hembry *et al.* 2014; Kawakita 2010; Machado *et al.* 2005; Molbo *et al.* 2003).

Another theoretical approach that relates to host plant specificity and focuses on the interaction itself is the investigation of fitness trade-offs that are said to exist between generalist and specialist insects (Egas *et al.* 2004; Joshi & Thompson 1995; McPeek 1996; Wilson & Yoshimura 1994). Generalist species are said to be able to use multiple host plants at the cost of the effective use of any one host, while specialist species are said to use a single host at the expense of host availability (Slobodkin & Sanders 1968; Wilson & Yoshimura 1994) effectively. Host plant specialisation has also been considered as an evolutionary ‘dead end’ by some (Kelley & Farrell 1998; Tripp & Manos 2008). Butterflies from the subfamily Nymphalinae show that host use can transition from specialist to generalist, and so specialization is not a dead end for all groups of organisms (Janz *et al.* 2006; Slove & Janz 2011). The evolution of insect-plant interactions is clearly nuanced, and investigation into the role of specific aspects of the biology of organisms, such as the role of feeding mode (Ali & Agrawal 2012), have proved insightful. It is however intuitive that host plant specificity does play a general role in the evolution of insect-plant interactions.

For the investigation of specific insect-plant interactions, an approach that begins with a focus on the organisms concerned is perhaps justified (Walter 2003). This requires knowledge of the nature of the interaction between insects and each of their host plant species, as patterns of multiple host use can be complex (Kishimoto-Yamada *et al.* 2013). Generalist insects utilize host plants for a variety of purposes (Hereward *et al.* 2013; Panizzi 2000; Velasco *et al.* 1995). For example, the bug *Nezara viridula* uses only some of the plants it feeds on for reproduction, and merely persists on others (Panizzi 1997, 2000; Velasco *et al.* 1995). Whereas generalists are associated with multiple hosts, a core of primary host species may be of critical importance (Oleksyn *et al.* 1998; Panizzi 2000; Rajapakse & Walter 2007; Wallingford *et al.* 2013; Walter 2003). Even insects that are perhaps the most specialised, those found in obligate pollination mutualisms, have been shown to have more subtle patterns of host association than was historically assumed, and this has led to a revised understanding of their evolutionary history and the processes that contribute most to their species diversity (Althoff *et al.* 2012; Kawakita 2010; Machado *et al.* 2005).

Understanding the patterns of host association in insect-plant interactions should begin with an accurate assessment of the species diversity of insects relative to their host plants. Species complexes are frequently found within generalist taxa (Hebert *et al.* 2004; Rafter *et al.* 2013; Smith *et al.* 2007) but specialists as well (Althoff *et al.* 2012; Machado *et al.* 2005; Molbo *et al.* 2003). Without such assessments it is difficult to interpret the true host species associations of an insect. For example, species regarded as generalists are often found to incorporate, unexpectedly, additional insect species which are often specialists (Rafter *et al.* 2013). The prevalence of "generalist" habits may thus be overrepresented until these complexes are disentangled. Conversely, some generalist species have been found to be complexes of species that also use multiple hosts, such as the thrips *Franklinella occidentalis* (Rugman-Jones *et al.* 2010) and whitefly *Bemisia tabaci* (Dinsdale *et al.* 2010). In these situations, the ecological quantification of host use cannot proceed until the species status of the organisms concerned is established. Accurately assessing the presence of cryptic species has not always been easy, and only with the widespread use of phylogenetic and population genetic analyses have estimations of species diversity become more straightforward, at least in situations where divergence is deep. These molecular techniques have also provided insights into how insect-plant interactions evolve, and the spatiotemporal scales over which such evolution takes place. Such molecular tools must, however, be used carefully with adequate consideration of the biology of each organism (in particular the host plants across which to test samples of the insects concerned) if species limits are to be accurately established, especially in cases where divergence is more recent.

Understanding the species-specific biology and host associations of herbivorous insects with different levels of host specialisation, in detail, should provide the basis from which the appropriateness of different generalisations and theories that relate to the evolution of herbivorous insect host relationships can be judged. This is the general aim of this thesis, as developed further in the following sections.

1.2 Research approach and thesis structure

The specific aim of this thesis is to contrast the genetic structure of an extreme specialist with that of an extreme generalist insect. Then, using the insights gained from the investigation of these two systems, develop and test hypotheses about the role of geography and host use in the evolution of insect-plant interactions more generally. I first explore the patterns of host association and gene flow, as well as the evolutionary history, of the extreme generalist insect (*Nezara viridula*; Hemiptera, Pentatomidae) followed by that of the extreme specialist insect (*Cycadothrips chadwicki*; Thysanoptera, Aeolothripidae). Aspects of the biology of these insects suggest that they are species complexes, and phylogenetic and microsatellite analyses were designed to determine the relationship between their populations. These analyses aim to characterise the population genetic structure of these insects relative to the geographical distribution of their host plants. My investigation into the patterns of species diversity and host associations within these systems will allow for a thorough interpretation of the processes that have generated their species diversity, and identify what role their host associations have played in this. Hypotheses about the role of host specificity and geography are then made and tested with reference to the results of similar studies on other herbivorous insects, in Chapter Six.

Nezara viridula is a highly polyphagous sucking pest of global significance, and genetic and biological variability is reported across populations of this species. What this variability means for understanding the ecology of this insect is unclear. Chapter Two investigates the biogeographical history of *N. viridula* and the relationship between the two genetic lineages of this insect that now overlap in distribution. The global and Australian distribution of the mitochondrial lineages of *N. viridula* is investigated as well as whether the individuals with mitochondrial DNA belonging to different lineages mate where they co-occur in Australia. Chapter Three investigates the spatiotemporal genetic structure of *N. viridula* across eastern Australia to determine whether their populations are relatively localised or are well dispersed in agricultural areas. A better understand of the long-term patterns of movement of this bug will help to investigate how host plant distribution and host plant availability influence the variable abundance of *N. viridula*.

Cycadothrips are the obligate pollinators of 12 *Macrozamia* cycads that are endemic to Australia, and this system has the potential to contribute significantly to our understanding of the origins, evolution, and persistence of mutualisms. Chapter Four investigates the species status and phylogenetic relationships of *Cycadothrips chadwicki* from the *Macrozamia* species that this thrips pollinates. A comprehensive investigation into the evolutionary relationship of *Cycadothrips* populations will provide insights into the role that host plant associations have played in generating the species diversity in the *Macrozamia*-*Cycadothrips* mutualism. An understanding of the species status of *Cycadothrips* from different *Macrozamia* host species will provide a solid foundation for further investigations into how speciation might occur in this mutualism. Chapter Five comprehensively assesses the spatiotemporal patterns of gene flow among populations of the *C. chadwicki* lineage that pollinates *M.*

lucida and southern populations of *M. macleayi* in southern Queensland (which is referred to as *C. chadwicki* sp. 4 as detailed in Chapter 4). The geographical distribution of *Macrozamia* appears to play a prominent role in structuring the species associations of the *Macrozamia-Cycadothrips* mutualism overall. Gene flow among *C. chadwicki* sp. 4 populations associated with geographically separated populations of their *Macrozamia* hosts should be low, and gene flow should be high where their hosts are more regularly distributed.

Each chapter is written as a manuscript in preparation for publication, and to avoid repetition only a brief outline of each system is included in the following sections.

1.3 *Nezara viridula* and its diverse host plants

Nezara viridula (Hemiptera, Pentatomidae) is a highly polyphagous insect pest that feeds on plants from more than 30 different families (Todd 1989). The insect is incredibly widespread geographically and is found on all continents except Antarctica (Todd 1989). *Nezara viridula* produces multiple generations within a single year (typically between two and five) (Todd 1989; Velasco *et al.* 1995) and the sequential use of host plants is necessary for *N. viridula* to persist (Panizzi 1997; Velasco *et al.* 1995). Individuals of *N. viridula* are thus associated with different host species depending on the location and the time of year (Jones & Sullivan 1982; Panizzi 2000; Velasco *et al.* 1995). During winter in temperate regions, *N. viridula* typically enters diapause in response to cold temperatures and decreased day length (Todd 1989). This is usually characterised by suspended reproductive capacity, a change in body colour from green to brown, or both (Musolin 2007; Todd 1989; Waterhouse & Norris 1987). During summer in some tropical regions, these insects even responded to high temperatures by aestivating (Singh 1972). Three distinct genetic lineages have been documented within *N. viridula* (Kavar *et al.* 2006; Li *et al.* 2010). However, the species status, morphology, geographical distribution, ecology, and mating behaviour of the insects concerned, has not been investigated for each of these lineages.

The relevance of the three genetic lineages to the uncertainty surrounding *N. viridula* remains unclear, but the variability in the biology and ecology of this insect may relate directly to them. The reproductive capacity of *N. viridula* varies considerably on different hosts (Panizzi 2000; Panizzi & Meneguim 1989; Velasco & Walter 1993), and populations of *N. viridula* from different regions differ in some of the hosts they find suitable, specifically *Ricinus communis* (castor oil plant) and *Zea mays* (corn) (Panizzi 2000), as well as at different temperatures (Ali & Ewiess 1977; Vivan & Panizzi 2005). Differences have been reported in mating behaviour, including vibratory communication, between *N. viridula* from distinct regions (Čokl *et al.* 2001; Jeraj & Walter 1998; Kon *et al.* 1988; Žunič *et al.* 2011). Until the genetic relationship of *N. viridula* from different locations is established, each of these issues cannot be investigated in a targeted manner. How speciation might occur in such a widespread insect cannot be understood without understanding the species status of different *N. viridula* populations and their respective biogeographical and evolutionary histories.

Within Australia most of the hosts that *N. viridula* uses are crop or weed plants (Velasco *et al.* 1995), and the local abundance of the insect is highly variable causing it to reach numbers that make it a significant pest only occasionally. Understanding why the seasonal and annual abundance of *N. viridula* is so variable is difficult because the insect is highly polyphagous, and host availability changes frequently according to local environmental conditions and crop availability (Panizzi 1997; Velasco *et al.* 1995). This insect relies on the sequential use of different host plant species throughout a year, and so its abundance is tightly linked to host plant availability (Panizzi & Meneguim 1989; Velasco *et al.* 1995). Further, the distribution of *N. viridula* in eastern Australia covers a large area, over 3000 km from north to south. The geographical scale at which host availability influences the local abundance of *N. viridula* is unclear and must be investigated.

To address these issues I establish the biogeographical history of *N. viridula* globally, as well as the contemporary population genetic structure of this insect in Australia relative to its host plants. The evolutionary history of *N. viridula* is investigated using a phylogenetic analysis of Australian individuals, as well as all publicly available sequence data from global populations. The spatiotemporal population genetic structure of *N. viridula* from eastern Australia is then investigated in depth to determine how its populations are related to one another in space and time. The results are then interpreted in relation to what is known about the host use and pest status of this insect in eastern Australia, particularly with respect to cotton.

1.4 *Cycadothrips* and the *Macrozamia-Cycadothrips* mutualism

Thrips in the genus *Cycadothrips* (Thysanoptera, Aeolothripidae) (Terry *et al.* 2005) are host plant specialists, the obligate pollinators of a group of endemic Australian cycads in the genus *Macrozamia* (Zamiaceae, Cycadales). Out of about 40 *Macrozamia* species, 12 are pollinated by *Cycadothrips* and only three species of *Cycadothrips* are presently described, one from Western Australia, one from the Northern Territory, and one from both Queensland and New South Wales (Forster *et al.* 1994; Jones *et al.* 2001; Mound 1991; Terry *et al.* 2008). The *Cycadothrips* species from eastern Australia, *C. chadwicki*, is the pollinator of 10 *Macrozamia* species with a broad distribution that covers thousands of kilometres, with many large breaks in that distribution. *Cycadothrips* are attracted to volatile compounds produced by *Macrozamia* cycads during the pollination widow and these thrips feed on pollen from male cycads (Terry *et al.* 2001). The volatile compounds produced by the cycads also attract *Cycadothrips* to female cycad cones and pollination is achieved through this mechanism (Terry *et al.* 2001). The specificity of different *Cycadothrips* species to these volatiles is presently unknown. How so many *Macrozamia* species persist while sharing a single pollinator is unclear, especially when we consider that some of these cycads exist in sympatry.

Cycadothrips chadwicki likely represents a complex of species, or at minimum, the movement of *Cycadothrips* among different cycad species is highly restricted during the pollination window. The *C. chadwicki* populations that are associated with *Macrozamia* show some variation, primarily in body size and coloration (Forster *et al.* 1994), and so at least in some circumstances there is limited or no gene flow occurring between populations. The evolutionary relationship between *Macrozamia* species is further complicated by most *Macrozamia* being pollinated by weevil species, and two *Macrozamia* even being pollinated by both *Cycadothrips* and weevils (Jones *et al.* 2001; Terry *et al.* 2001). Only the *Macrozamia* pollinated by *Cycadothrips* are considered with depth here. The diversity of *Macrozamia* pollination biology and the geographical relationships between their populations raises several questions regarding the origins, age and persistence of the species diversity in this system. How these results contribute to our understanding of speciation in the insects and plants involved in obligate pollination mutualisms is also considered.

My research tests the hypothesis that *C. chadwicki* is a species complex using molecular analyses. First a phylogeny of *Cycadothrips* is produced, followed by an analysis of gene flow among closely related

C. chadwicki populations using microsatellite markers. The divergence times of *Cycadothrips* found on different *Macrozamia* species is estimated and compared to the known divergence times of their host cycads. These patterns of host-pollinator species associations and the divergence estimates provide the basis for interpreting the relative roles of coevolution, host switching and co-diversification in this system. Microsatellite markers are used to comprehensively investigate spatiotemporal gene flow within one *C. chadwicki* lineage across its two host *Macrozamia* species to quantify pollinator population connectivity and so estimate how thrips dispersal might influence cycad pollination. This allows for consideration of how speciation might proceed in this mutualism more generally.

1.5 Gene flow and insect host use

Why phytophagous insects are so speciose, and how different patterns of host use evolve, have been central questions in research on insect-plant interactions. For herbivorous insects, the availability of each host plant species changes through space and time (Strong 1979), and this is particularly the case over long evolutionary time scales (Petit *et al.* 1999; Zachos *et al.* 2001; Dynesius & Jansson 2000; Walther *et al.* 2002). One process that may affect insects with different patterns of host association differentially is the likelihood of population fragmentation occurring through changes in the distribution of their host plants. Host plant interactions may, therefore, influence insect population genetic structure through the number of hosts of each particular with which an insect interacts. Geographical isolation and the cessation of gene flow is important for the majority of speciation events (Mayr 1942; Mayr 1970; Futuyma 1980; Fitzpatrick 2009). Aspects of the landscape, the dispersal capability of specific insect species, and host phenology, also affect population genetic structure (Peterson and Denno 1998; Avise 1994). The population genetic structure of insect populations can therefore be used to investigate the generality of these patterns by comparing the amount of gene flow that occurs between insect populations that are associated with different host plant species.

Gene flow is the movement of genetic material between populations of a species and, rarely, between different species (Slatkin 1985; Toews & Brelsford 2012). Gene flow can result from migration, movement of gametes, extinction and recolonization of entire populations, or the movement of extra-nuclear DNA such as mitochondria (Slatkin 1985). Gene flow is not the only factor that affects the genetic structure of organisms, with genetic drift, selection, and mutation also affecting the genetic makeup of populations (Avise 1994). Indirect methods are used to assess gene flow, which usually involves statistical analysis of genetic markers that are investigated over the geographical distribution of organisms (Slatkin 1985; Avise 1994). Indirect methods are used because direct measurements of gene flow, for example direct observations of mating and dispersal, are difficult to make (Bossart and Prowell 1998; Avise 1994).

A rich body of molecular population genetics data is available from a variety of herbivorous insects, but relatively few studies have attempted to use such data to make broader generalizations about insect plant interactions (but see Peterson and Denno (1998)). I therefore investigated the feasibility of using the population genetic literature of phytophagous insects to perform a secondary analysis, one that investigates the relationship between the spatial genetic structure of insect populations and their diet breadth. For such an approach to be successful, accurate species delimitation and host use categorization, as described above, are crucial. Using the investigations into such questions for *N. viridula* and *C. chadwicki* an approach to addressing this question is formulated, and then tested with a preliminary data set. This also includes exploration of various options regarding the choice of genetic measures, appropriateness of insect species, study suitability, and marker selection. The specific prediction tested is that populations of relatively specialist insects should be more fragmented and so have increased genetic differentiation over shorter geographic distances.

Chapter Two – Deep divergence within the herbivorous generalist bug *Nezara viridula* – global biogeography, invasion, and secondary contact in Australia

2.1 Introduction

The association of insect pests with agriculture has for many facilitated their invasion on other continents (Anderson *et al.* 2016; Cristescu 2015; Rafter *et al.* 2013). The process of invasion has important evolutionary consequences for invasive populations, affecting the genetic diversity of the initial colonisers (Blows & McGuigan 2015; Colautti & Lau 2015). The genetic history of invasive insects frequently shows that the process has not occurred as a single population simply establishing in a new area (Garnas *et al.* 2016), and multiple invasions can add to the complexity of the genetic structure of invasive populations (Blackburn *et al.* 2011; Walther *et al.* 2009). In such situations, there is potential for secondary contact and the subsequent admixture of genetically distinct and previously allopatric insect populations (Dlugosch *et al.* 2015; Garnas *et al.* 2016). In the case of herbivorous insects this can have consequences for host range, insecticide resistance, and tolerance for environmental conditions, and this may also affect their survival and spread.

The consequences of such admixture for understanding the ecology of contemporary insect populations can be overlooked if this genetic diversity is not well understood. The biogeographical history, evolutionary history, and contemporary population genetic structure are particularly relevant in those taxa where there is evidence of undetected species or sub-species. For phytophagous insects this will necessarily include an assessment of any potential differences in host use among genetically distinct lineages, the change in available host plant species across and within their native and invasive distributions (purely because of host availability), as well as any potential differences in host use among distinct genetic lineages. For truly generalist insects this is difficult because of the large spatial and temporal variation in their host plant interactions, and potential changes in host availability in their native and invasive ranges. These issues all potentially apply to the generalist bug *Nezara viridula* (Linnaeus) (Hemiptera: Pentatomidae), as detailed below.

Nezara viridula is polyphagous and feeds on plants from over 30 different families, of which many are important agricultural or horticultural species (Todd 1989). The broad host range, economic significance, and association with agriculture help explain its global distribution (Panizzi 1997; Todd 1989). The fruit, pods, and seeds of host plants are fed on preferentially and *N. viridula* moves through a sequence of host plants as they become more or less suitable for feeding and oviposition (Todd 1989). Individual female *N. viridula* are reported as flying as far as 1 km in a single day (Kiritani and Sabasa 1969) and wind may also displace these bugs large distances (Aldrich 1990). Development of *N. viridula* proceeds from the egg stage, through five nymphal instars, eggs take about five days to incubate and 23 to 58 days to progress through from egg to adult (Kariya 1961). A higher temperature provides a higher rate of development (Todd 1989). Temperature and host availability interact to produce a variable number of generations each year, depending on region (Todd 1989; Velasco *et al.* 1995).

Three mitochondrial lineages have been detected worldwide for *N. viridula*. One lineage is restricted to Africa and the other two are distributed more widely (see below) and even sympatrically in Brazil and Japan (Kavar *et al.* 2006; Li *et al.* 2010). Despite the ecology and biology of *N. viridula* having been investigated with intensity for decades, the question of whether multiple species or sub-species exist within this taxon remains unresolved despite the presence of separate genetic lineages. The geographic origin of *N. viridula* also remains uncertain, despite a number of reasoned suggestions (Ferrari *et al.* 2010; Jones 1988; Kavar *et al.* 2006; Li *et al.* 2010). Both of these points, the biogeographical history of *N. viridula* and the presence of multiple genetic lineages, make interpreting the ecology of this insect (Kiritani 2011; Musolin *et al.* 2011; Tougou *et al.* 2009) more difficult. Whether *N. viridula* is one or more species could have significant consequences for understanding its ecology and thus guiding future research on this pest.

The three mitochondrial lineages of *N. viridula* are associated, broadly, with Africa, Asia, and Europe (Kavar *et al.* 2006; Li *et al.* 2010). Most species in the genus *Nezara* are confined to Africa (Ferrari *et al.* 2010) and of the species found outside Africa, only *N. viridula* and *N. antennata* appear to have wide distributions (Ferrari *et al*. 2010). The Asian and European lineages of *N. viridula* are the ones that have evidently spread outside of their native ranges (despite this not yet having been unambiguously defined) and become invasive in a number of areas that include the Americas and Australia (Todd 1989). Molecular evidence demonstrates that insects of the European lineage have in at least one case (Japan) even moved into an area where the Asian lineage was already present (Kavar *et al.* 2006). The African lineage is the most divergent and it is separated from the other two (Asian and European) by the closely related species *N. antennata* and so is likely a distinct species (Li *et al.* 2014). The African lineage of *N. viridula* is found only in Africa and is estimated to have diverged from the non-African lineages six million years ago (Mya) (Kavar *et al.* 2006) to 18 Mya (Li *et al.* 2010).

Divergence estimates for the more closely related Asian and European lineages range from 1.2 Mya (Kavar *et al.* 2006) to four Mya (Li *et al.* 2010). In short, these divergences are substantial compared with what is typical within a single species, and even the divergence between the more closely related Asian and European *N. viridula* lineages.

The depth of evolutionary time that separates the mitochondrial lineages of *N. viridula* is potentially associated with ecological differences, but a case cannot yet be made because the relationship between these ecological differences and the genetic lineages has not been investigated. Differences exist in vibratory sound communication of *N. viridula* from different locations, an important aspect of mating in these insects (Jeraj & Walter 1998; Kon *et al.* 1988; Ryan *et al.* 1996; Virant-Doberlet *et al.* 2000). These differences are associated with sex-biases in mating across some populations (Jeraj & Walter 1998). Potential differences in host range also suggest possible ecological differences, as at least *Ricinus communis* and *Zea mays* are considered suitable for some of the global populations of *N. viridula* and not others (Panizzi 1997; Panizzi & Meneguim 1989). Populations of *N. viridula* from different locations may also have distinctive pheromone profiles (Aldrich *et al.* 1987; Miklas *et al.* 2000; Ryan *et al.* 1995) and they display an incredible range in copulation times, from 1h to165h (McLain 1980) when compared with times that appear more typical of Pentatomids (2.7h to 4.7h, (Silva *et al.* 2012); 5.6h to 11.2h (Rodrigues *et al.* 2009)), which may indicate differences in mating behaviour across populations of *N. viridula*. In other insects with similar degrees of mitochondrial divergence, some represent multiple species with different patterns of host use (Rafter *et al.* 2013) while others with even high amounts represent a single species (Toon *et al.* 2016), so *N. viridula* warrants closer investigation.

Several authors have justified suspicion of cryptic species within the taxon *Nezara viridula* (Jeraj & Walter 1998; Kavar *et al.* 2006; Ryan *et al.* 1996; Virant-Doberlet *et al.* 2000). A number of approaches have been taken to address this question, including assessing mitochondrial phylogenetics (Kavar *et al.* 2006; Li *et al.* 2010) and local population genetic structure (Sosa-Gómez *et al.* 2005), morphology (Ferrari *et al.* 2010; Qi *et al.* 1999), and different aspects of their mating behaviour (Jeraj & Walter 1998; Ryan *et al.* 1995; Virant-Doberlet *et al.* 2000). These approaches have so far provided no conclusive evidence that *N. viridula* represents a complex of species. However, none have assessed, on the basis of a suitably structured sampling program, both the historic and contemporary genetics of *N. viridula* populations where the lineages are found in sympatry and across host plants. This would resolve the ambiguity regarding the species status of *N. viridula* by assessing whether mating occurs between the mitochondrial lineages. Specifically, the extent of gene flow across the nuclear genome must be assessed in regions where the mitochondrial lineages co-occur before the species status of these mitochondrial lineages can be determined (e.g. (Toon *et al.* 2016)).

This study was designed to establish the genetic relationship between the two most widely distributed mitochondrial lineages of *N. viridula*, the Asian and European. Phylogenetic analyses using mitochondrial (COI) and nuclear genes (EF1α and Tubα1) were used to investigate biogeographical history of these lineages with the context of the genus *Nezara* as a whole. Further, both the Asian and European mitochondrial lineages are present in Australia, and this allowed investigation of the population genetic structure of *N. viridula* at sites where the mitochondrial lineages co-occur. Microsatellites were used to investigate the extent of gene flow between individuals belonging to the different mitochondrial lineages to determine whether mating occurs across those lineages. Gene flow was also assessed more generally across host plants and across geography to determine whether either mtDNA lineage was restricted in these respects. The phylogenetic analysis also included sequences from global populations of *N. viridula* that were available from online databases, while the microsatellite analyses were performed only on Australian *N. viridula* populations.

2.2 Methods and materials

2.2.1 Sampling, DNA sequencing, and molecular approach

Individuals of *N. viridula* were sampled across Australia from 2014 to 2016, from 33 different host plant species across a about 3,000 kilometres (north to south) in eastern Australia, and in two locations over 400km apart in the north-west of the continent. Most sampling locations were within agricultural regions, and sampling sites in these regions were separated by at most 150km. Some pinned and field collected specimens were provided by the North Australian Quarantine Strategy (NAQS), and some older samples were obtained from freezer storage or pinned specimens at The University of Queensland. Specific site, date, and host plant information are presented in Table S2.1. DNA extraction was performed on each *N. viridula* adult by either removing a leg and extracting with 20% Chelex (Bio-rad) solution or dissecting out flight muscle tissue for a salt extraction (Appendix 2.1). Adult samples were used to avoid biasing the results through inclusion of nymphs, because they have limited mobility and are gregarious for some time after hatching from the egg mass (Todd 1989). However, where the sample size from a region or host plant was small, a maximum of one 4th or 5th instar nymph was included from any particular site and host combination to bolster sample size.

Primers for 12 microsatellite loci were developed to assess the contemporary population genetic structure of *N. viridula* (see Appendix 2.2 for microsatellite development). Data was obtained for 571 Australian *N. viridula* individuals that were genotyped at these loci. Some individuals had poor amplification and 14 for which data was missing at five or more loci were removed from the final data set (not included in the count above). Individuals collected from each site were grouped into aggregate populations (i.e. regardless of host plant) for microsatellite analyses. These groups represent individuals collected at the same location and at the same time (see Table 2.1). Individual population assignment analyses were later carried out in STRUCTURE (see below) to determine whether these groupings were appropriate.

Mitochondrial DNA (mtDNA), specifically the gene region cytochrome C oxidase subunit I (COI), was used to relate the Australian insects to those from other continents. The primers used to amplify mitochondrial sequences were selected to give fragments that overlapped with the cytochrome C oxidase subunit I (COI) barcoding gene region (see Appendix 2.3). This gene region was successfully sequenced in both directions for 176 Australian individuals, and for six individuals from outside Australia (the Philippines and Egypt) to establish which mtDNA haplotypes were present in Australian populations. An additional 304 Australia *N. viridula* were sequenced at COI in a single direction but only to determine which of the two mtDNA lineages the insects belonged to. As such, these single direction sequences were not used in phylogenetic analyses because of the higher chance of sequencing errors. Mitochondrial sequences used in previous phylogenetic analyses (Kavar *et al.* 2006; Li *et al.* 2010) were obtained from GenBank, as were COI sequences from various published and unpublished barcoding projects (Table S2.2).

Two nuclear DNA (nuDNA) gene regions were sequenced for phylogenetic analysis of *N. viridula*, as previous studies only used mtDNA, and the pattern in mtDNA versus nuDNA can differ (Toews & Brelsford 2012). The Tubulin alpha 1 (Tubα1) gene region was sequenced for 178 individuals and the elongation factor 1 alpha (EF1α) gene region for 172 individuals. Primers for both genes were developed *de novo* (see Appendix 2.3.). DNAsp 5.1 (Librado & Rozas 2009) was used to estimate the frequency of the nuclear DNA haplotypes from diploid sequences.

Table 2.1. Groupings of *Nezara viridula* samples by site, with population and geographical region designations. See Table 2.1 for the site codes. Geographical region labels that include state abbreviations are as follows: NSW – New South Wales, NT – Northern Territory, QLD – Queensland, and WA – Western Australia. Populations marked with an asterisk (*) were excluded from all analyses that did not group populations into regions. Latitude and longitude here represent the mid-point between the sample sites that make up each population. The populations within geographical regions were at most 30km from one another.

Pop.	Pop. Site Codes	n	Region	Lat.	Long.
KUN	GH01, KU01, KU02, KU03, KU04, KU05, KU06, KU07, KU08, KU09, KU10	118	Northern WA	-15.672	128.725
DAR	DW01, DW02, DW03, DW04	23	Northern NT	-12.510	131.107
LHR	LR01	2*	Far North QLD	-12.800	143.316
TOW	TV01	24	Northern QLD	-19.285	146.822
GRU	GU01	8*	Northern QLD	-19.571	147.137
BOW	BO01, BO02, BO03, BO04	60	Northern QLD	-20.052	148.161
EMRa	EM01, EM02	19	Central QLD	-23.549	148.202
EMRb	EM03, EM04	40	Central QLD	-23.549	148.202
BIL	BI01, BI02	32	Central QLD	-24.404	150.520
DAL	DA01, DA02	32	Southern QLD	-27.283	151.275
GAT	DA05	24	Southern QLD	-27.541	152.337
NARa	NA01, NA02	25	Central NSW	-30.255	149.553
NARb	NA04	24	Central NSW	-30.255	149.553
BBA	NA03	24	Central NSW	-30.542	150.010
BRZ	BR01, BR02, BR03	22	Central NSW	-31.186	150.433
GRIa	DP02, GR01, GR02	31	Southern NSW	-34.505	146.190
GRIb	GR04	25	Southern NSW	-34.441	146.037
HAY	HA01, HA02	23	Southern NSW	-34.472	144.753

2.2.2 Phylogenetic analyses

The global distribution of each mitochondrial lineage of *N. viridula* (Asian and European) was assessed using phylogenetic analyses of all available COI sequence data. Three analyses were performed because the COI gene sequences had two different fragment lengths: (i) a 149 base pair (bp) overlapping region common to all sequences (overlapping fragment - OF), (ii) a 557 bp fragment of the more usual COI barcoding region (long fragment - LF), (iii) a 348bp fragment that was used in previous phylogenetic studies of *N. viridula* (short fragment - SF) (Kavar *et al.* 2006; Li *et al.* 2010). Both the LF and SF data sets are subsets of the OF dataset. The nuDNA sequences (418bp EF1α and 458bp Tubα1) were not incorporated into a single phylogenetic analysis because of the nuclear-mitochondrial discordance detected (see below) mean that the mtDNA and nuDNA of some contemporary populations do not relate directly to the pre-contact lineages of these insects. Instead, haplotype networks were created independently for the mitochondrial and each of the two nuclear genes using a TCS network (Clement *et al.* 2000) as implemented in PopArt (Leigh & Bryant 2015).

Jmodeltest 2.1.4 (Darriba *et al.* 2012; Guindon & Gascuel 2003) was used to compare evolutionary models using Akaike Information Criteria (AIC) for each COI dataset for use in phylogenetic analyses. The evolutionary model with the best AIC estimates for the COI datasets was Jukes-Cantor (JC) (Jukes & Cantor 1969) in each case and so this was used in each of the COI phylogenetic analyses. Phylogenetic analyses were performed using the MrBayes 3.2.6 (Huelsenbeck & Ronquist 2001) plugin as implemented in Geneious 9.0.5. Only sequences that were unique within each dataset were used. The COI sequences of *Eurydema gebleri* (Pentatomidae) (KP207595.1) and *Oechalia schellenbergii* (EF641159.1) (Pentatomidae) were included in each COI phylogenetic analysis, with the former used as the outgroup. Each phylogenetic analysis used uniform branch lengths with a gamma range of 0.0112 to 0.0175 to reflect the most likely rate of sequence substitution for COI (Papadopoulou *et al.* 2010). A chain length of 1,100,000 was used with a burn-in of 100,000. The overlapping sequence dataset (OF) had a low number of informative sites, and so the molecular clock was used only to constrain the tree and so no estimates of divergence times are included. The probable lower and upper limits of divergence times are covered by Kavar *et al.* (2006) and Li *et al.* (2010) respectively.

2.2.3 Microsatellite analyses

The twelve microsatellite primer pairs were developed specifically to function across *N. viridula* individuals from both the Asian and European lineages. The individuals used to develop these primers were sampled as distantly from one another as possible in Australia, with the Asian lineage represented by insects collected from Kununurra in Western Australia, and the European lineage by insects from Griffith in New South Wales. Microsatellite peaks were confirmed and binned manually using the Geneious 9.0.5 microsatellite plugin 1.4.2. Summary microsatellite statistics (Table 2.2) were calculated using the 16 aggregate populations that had 19 or more genotyped individuals (see Table 2.1) to assess the suitability and quality of the microsatellite markers. Null allele estimates and global F_{ST}, both with and without ENA correction, were calculated using FreeNA (Chapuis & Estoup 2007). Deviations from Hardy-Weinberg Equilibrium (HWE) were measured using Hardy-Weinberg exact tests in Genepop 4.2 (Raymond & Rousset 1995; Rousset 2008), as were estimates of linkage disequilibrium.

Microsatellite data were analysed using an individual based Bayesian clustering algorithm, STRUCTURE 2.3.4 (Falush *et al.* 2003; Falush *et al.* 2007; Pritchard *et al.* 2000) that tests the probability of the assignment of each individual to a chosen number of populations (K). A number of K values are assessed, the most appropriate K values are estimated, and then all biologically meaningful K values are discussed and used to direct further analyses (Porras-Hurtado *et al.* 2013; Puechmaille 2016). STRUCTURE analyses were performed using the admixture model with correlated alleles, with values of K (assumed number of populations) from one to 10, and with 10 replicates for each K value. A burn-in of 100,000 iterations, followed by 1,000,000 iterations, was used. Discriminatory Analysis of Principle Coordinates (DAPC) was also performed to assess the clustering of individuals, as well as whether insects from broad regions differed. DAPC used the *adegenet* package (Jombart *et al.* 2010) for R (R 2016), and using 100 cross-validation replicates to choose the number of principle and discriminatory coordinates retained and using all genotyped individuals, with the output coloured according to sampling region. Pairwise population F_{ST} values were obtained in Genepop 4.2 (Raymond & Rousset 1995; Rousset 2008), with and without high null allele loci (>10%), to test whether these loci influenced the analyses unduly (see Tables S2.5 and S2.6 respectively). Differences in allelic richness across the populations identified by STRUCTURE were calculated manually (Table S2.5) to compare their genetic diversity.

Table 2.2. Statistics of the microsatellite loci included in the population genetic analysis of *Nezara viridula.* Only samples collected from the 16 populations with 19 or more individuals were used to generate these statistics (see Table 2.1) with the exception of the total number of alleles (∑Na).

Locus	**∑*Na*[a]**	**$\hat{N}$a[b]**	***Ho*[c]**	***He*[d]**	**HWE[e]**	**Est. Null Alleles**	**gF_{ST}**	**gF_{ST} ENA**
NEZA01 (51)	5	3.2	0.594	0.572	0 (0)	0.02	0.13	0.13
NEZA02 (82)	6	3.2	0.586	0.540	0 (0)	0.01	0.08	0.07
NEZA03 (55)	2	1.1	0.064	0.056	0 (0)	0.00	0.15	0.15
NEZA04 (77)	8	4.4	0.695	0.638	3 (0)	0.02	0.07	0.07
NEZA05 (89)	4	2.3	0.347	0.428	5 (3)	0.09	0.14	0.12
NEZA06 (50)	5	4.1	0.517	0.499	1 (1)	0.01	0.07	0.07
NEZA07 (96)	5	3.1	0.485	0.514	0 (0)	0.03	0.14	0.13
NEZA08 (127)	3	2.0	0.145	0.151	0 (0)	0.02	0.29	0.28
NEZA09 (95)	8	3.4	0.583	0.526	0 (0)	0.01	0.18	0.18
NEZA10 (86)	5	4.1	0.524	0.626	8 (3)	0.09	0.06	0.05
NEZA11 (52)	4	3.9	0.590	0.594	0 (0)	0.01	0.05	0.05
NEZA12 (78)	3	3.0	0.590	0.577	1 (0)	0.02	0.10	0.09

[a] Total number of alleles

[b] Average number of alleles

[c] Average observed heterozygosity

[d] Average expected heterozygosity

[e] Deviations from Hardy-Weinberg Equilibrium and their within loci Bonferroni corrected values shown in bracket

2.3 Results

2.3.1 Global phylogenetics

All of the global *N. viridula* sequences obtained could be assigned to either the Asian or European mitochondrial lineage of *N. viridula* (bar one which is dealt with below). These lineages were resolved as separate clades in all three phylogenetic analyses (Figure 2.1, Figure S2.1 and Figure S2.2), even with the low number of informative sites in the overlapping fragment (OF) data set. Based on COI sequences generated from this study, and obtained from others, the distribution of mtDNA haplotypes is as follows. The Asian lineage of *N. viridula* is found in sequences obtained from individuals sampled from Australia, Brazil, China, Japan, Korea, and India. The European lineage is found in Canada, Egypt, Greece, Iran, Italy, the Galapagos, Guadeloupe, Madeira, the Philippines, and the USA (California, Florida, Georgia and Texas) and within the same populations as the Asian one in Australia, Brazil, and Japan. The only sequence not readily assigned to either linage in the OF dataset was one from Brazil, designated NV-COI-4 in previous research (Kavar *et al.* 2006), but this is resolved as an early branch of the Asian lineage in the analysis shown in Figure S2.1 and so belongs to the Asian mitochondrial lineage.

Two haplotypes, each from a single individual, were highly divergent relative to other *N. viridula* sequences. One was from Botswana and the other from India (labelled *Nezara* sp. in Figure 2.1). These two sequences are separated from other *N. viridula* sequences by sequences of the species *N. antennata* (Figure 2.1). The sample from Botswana represents the previously described African lineage (Kavar *et al.* 2006) and the sequence from India represents a further lineage that has not been described. Neither has been investigated further.

2.3.2 Phylogenetics of Australian *Nezara viridula*

A total of 17 COI haplotypes was found across 480 Australian *N. viridula* individuals but only six of these were from individuals that had been sequenced in both directions. All Australian *N. viridula* COI sequences were clearly associated with the Asian or European mitochondrial lineages, and these two lineages have a mostly separate distribution from one another in Australia (Figure 2.2). The Asian lineage is found in the north of Australia, including the northern parts of eastern Australia, and the European one in the southern and central parts of eastern Australia.

The two mitochondrial lineages are sympatric in only a single Australian location, Bowen in Queensland (Figure 2.2). At this site most haplotypes belonged to the Asian mitochondrial lineage (81%). Nine haplotypes were found in the Asian haplotype group with five found in individuals sequenced in both directions. The European haplotype group had eight haplotypes but only one in individuals sequenced in both directions. Most sequences from each lineage represented only a single haplotype – 98% in the European lineage and 84% in the Asian lineage.

The most common nuclear gene haplotypes for both Ef1α and Tubα1 in Australian *N. viridula* were associated with both mitochondrial lineages (Figure S2.3). Tubα1 had a total of five haplotypes. The three most common Tubα1 haplotypes (96%) were present in both mitochondrial lineages, and the remaining two were found only in individuals with mtDNA of the Asian lineage. Ef1α was more diverse, with 16 haplotypes, but again the two haplotypes with the highest frequency (83%) were associated with both of the mitochondrial lineages. Of the remaining haplotypes, seven were associated only with the European mtDNA lineage, four with the Asian lineage, and the other three were associated with both mtDNA lineages.

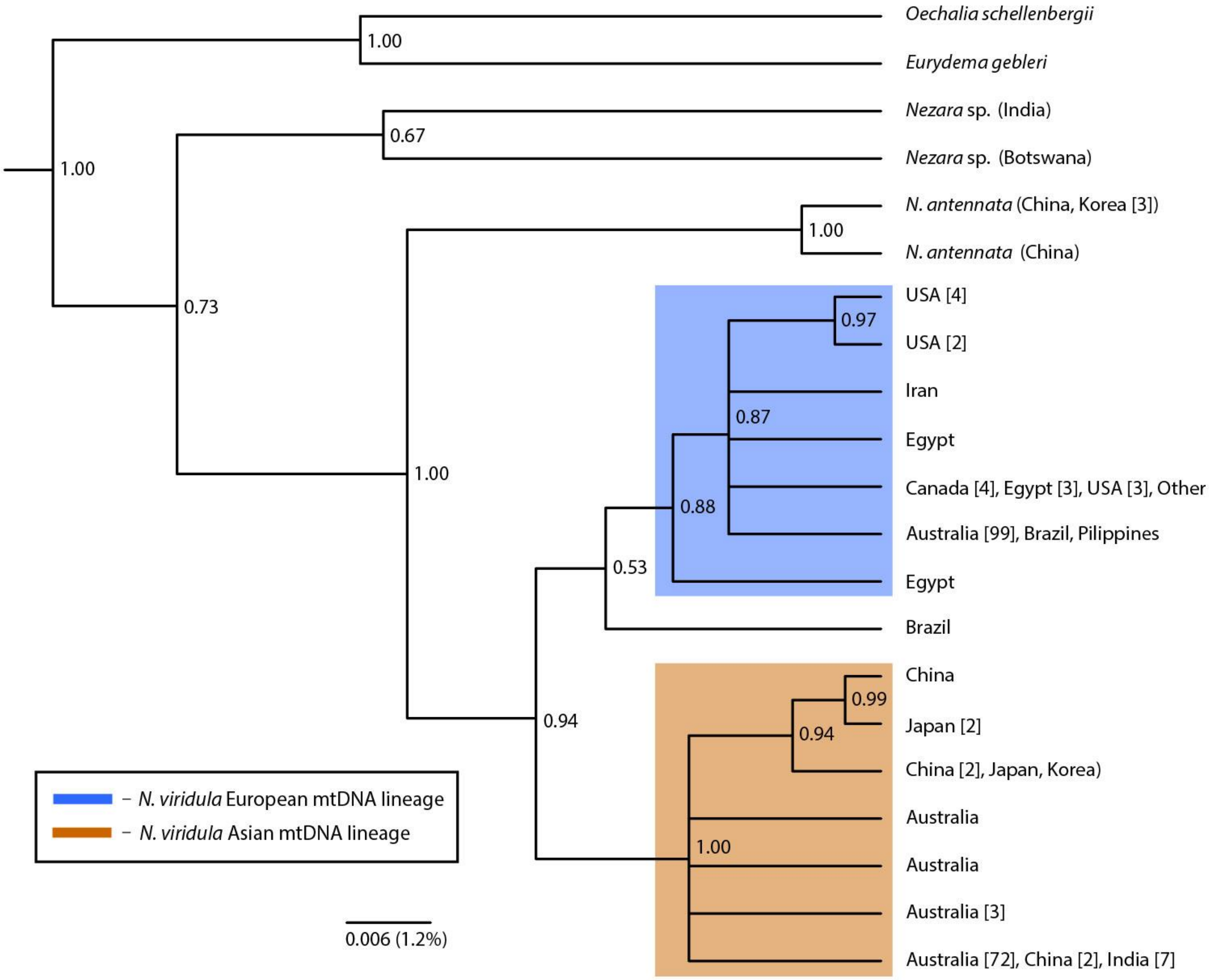


Figure 2.1. Tree representing the phylogeny of *Nezara viridula*, based on a MrBayes 3.2.6 analysis and using the 149 bp COI gene dataset (OF in the text) from all available individuals of both *N. viridula* and *N. antennata* (Table S2.2). Only unique sequences were used and if the number of samples that shared a sequence is greater than one then the number is shown in square brackets. *Eurydema gebleri* and *Oechalia schellenbergii* (both also Pentatomidae) were included, with *O. schellenbergii* designated as the outgroup. The label 'Other' denotes a haplotype of *N. viridula* common to individuals sampled from Greece, Guadeloupe, Galapagos, California, Japan, Italy and Brazil, from the Kavar (2006) data set. Posterior probabilities are shown as node labels. The results of the same analyses, but using the 348 bp (SF) and 557 bp (LF) datasets (see text), can be found in Supporting Information Figures S2.1 and S2.2 respectively.

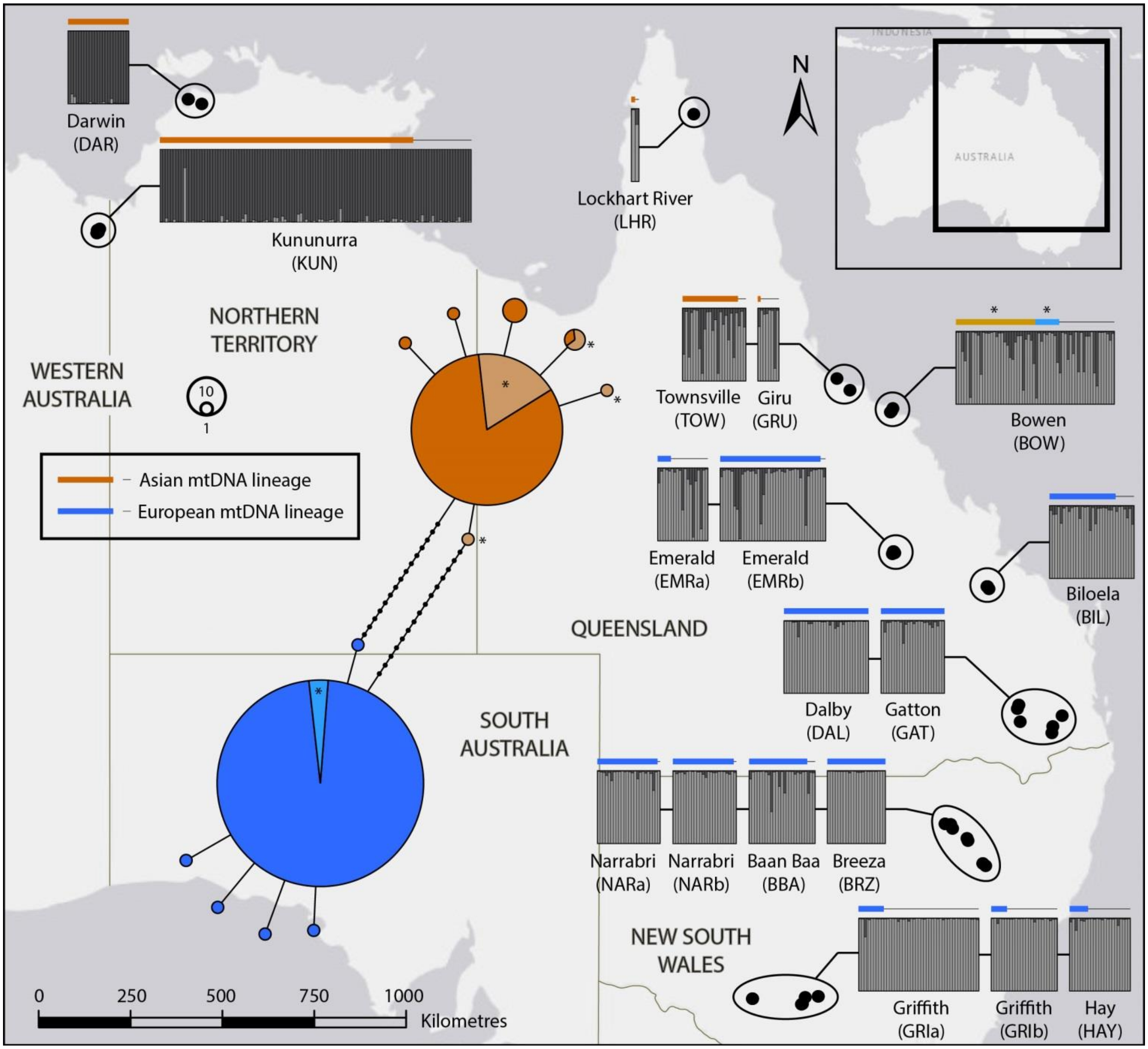


Figure 2.2. A map of *Nezara viridula* sampling sites with sampling locations represented by black dots. The circles define nearby sampling sites (black dots) into regions (Table 2.1). Superimposed on the map is a haplotype network showing the relationship between the COI haplotypes of 480 Australian *N. viridula* individuals (all sequences), and also a K=2 STRUCTURE analysis of the microsatellite data, from 571 individuals (Figure S2.4), separated into boxes of individuals associated with sampling regions. Horizontal bars above the boxes indicate the mitochondrial haplotype of the corresponding individual in the STRUCTURE analysis boxes. For both of the analyses represented, the Asian mitochondrial lineage is shown in orange and the European in blue. Individuals that were genotyped but did not have their COI gene sequenced are represented above the boxes by a thin black line. Bowen is the only location in which bugs with both mitochondrial haplotypes were found together. Where haplotypes of both lineages were found in the same location, they are shown with an asterisk and a lighter shade of either orange or blue to distinguish them from others in the haplotype network and the bars above the STRUCTURE plots. See Table 2.1 for sample label state abbreviations.

2.3.3 Microsatellite analysis of Australian *Nezara viridula*

The 12 microsatellite loci developed in this study amplified well across both mitochondrial lineages. During development a number of loci were rejected because of their high estimate for null alleles (>20%) in insects from only one lineage or the other (Appendix 2.2). Of the final 12 loci some showed departures from Hardy-Weinberg Equilibrium (HWE) even after sequential Bonferroni correction within loci, but not in all populations, and so this was not of concern. Of these loci, two (NEZA05 and NEZA10) had a high null allele estimate, even though none had >10% average across populations, and these two loci also had the highest number of deviations from HWE (Table 2.2). No loci showed evidence for linkage disequilibrium. Pairwise F_{ST}s were calculated with and without the loci NEAZ05 and NEZA10 to assess the impact of these loci on the outcome (see Tables S2.5 and S2.6) and only the southern populations of *N. viridula* (Griffith and Hay, Figure 2.2) were affected.

For the STRUCTURE analysis of the microsatellite data, the most likely inferred value of K was 2 (see Figure S2.4 for K values two through four, and Figure S2.5 for the likelihood of K and Delta K). This K=2 analysis groups all the eastern populations of *N. viridula* together, and separates them from those in north-western Australia (Kununurra and Darwin). Most importantly, this analysis shows the discordance between the nuDNA and mtDNA in eastern Australia. Here, gene flow occurs across individuals with mtDNA of each lineage, despite the distribution of the mtDNA being mostly separate (Figure 2.3). Although eastern *N. viridula* represent a single interbreeding population (STRUCTURE analysis in Figure 2.2), individuals from northern and central Queensland were more frequently assigned to either of the two populations than were individuals from southern Queensland and New South Wales in the K=2 STRUCTURE analysis (Figure 2.2). However, in the K=3 analysis few eastern individuals were assigned to north-western populations (Figure S2.4). On the eastern side of Australia, only two individuals were sampled far north of the mtDNA boundary plotted in Figure 2.3 (i.e. at Lockhart River), and these individuals were also assigned to the eastern populations in the K=2 and K=3 analyses. The results of DAPC show the same two groups that were evident in the STRUCTURE analysis (Figure 2.4), north-western Australia and eastern Australia. Within the northern and eastern populations there were no further large genetic differences that might indicate a reduction in gene flow across *N. viridula* collected from different host plants (Figure 2.2).

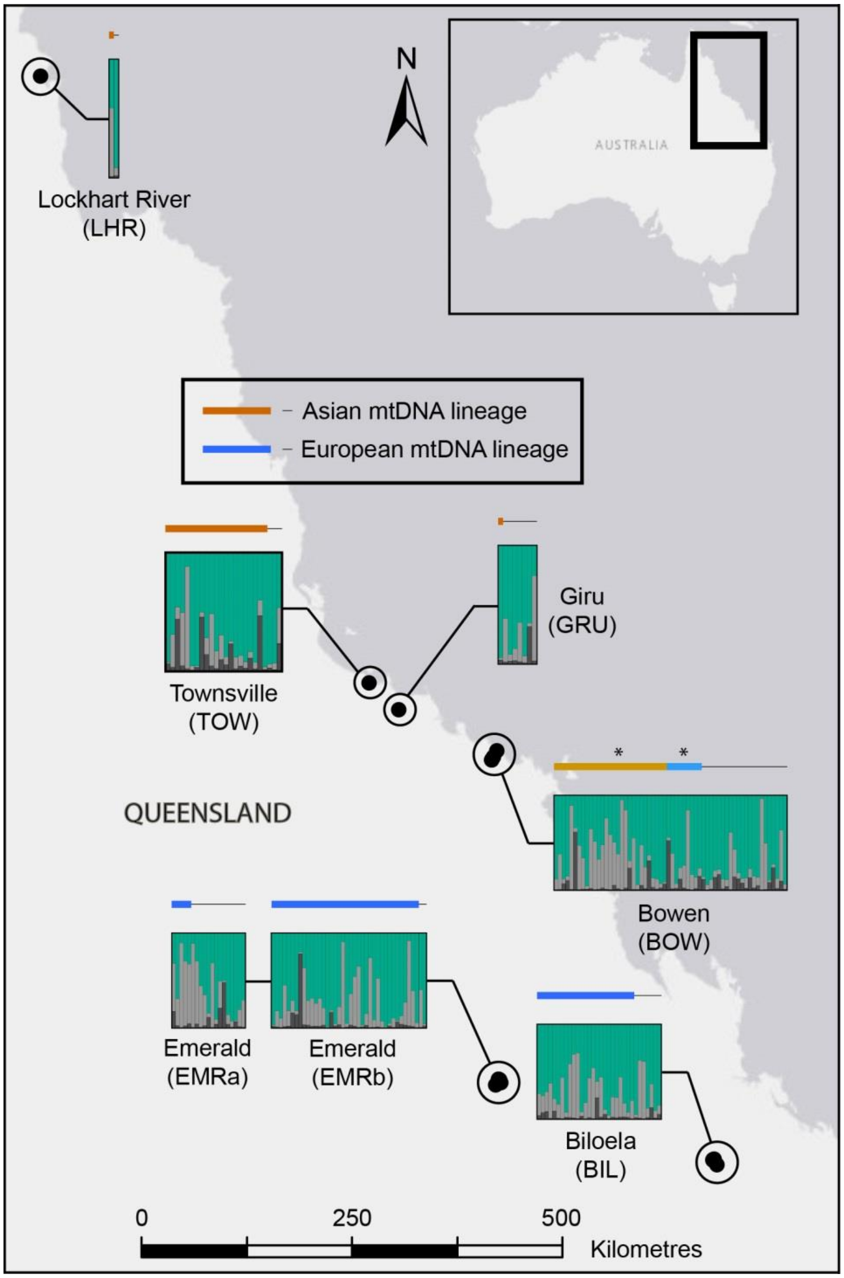


Figure 2.3. Map of north-east Australia showing the geographical area of overlap in the distribution of the two *Nezara viridula* mitochondrial lineages. Microsatellite data are displayed in the boxes and represent the results of the K=3 STRUCTURE analysis (Figure S2.4). Dark grey bars represent assignment to the same population as most individuals from north-west Australia. Horizontal bars over the boxes indicate the mitochondrial haplotype of the corresponding individual in the STRUCTURE analysis. The Asian mitochondrial lineage is shown in orange and the European one in blue. Individuals that were genotyped but did not have their COI gene sequenced are represented in black in the horizontal line above the boxes. Bowen is the only location in which bugs from both mitochondrial haplotypes were found together, and so these haplotypes are represented with an asterisk and a lighter shade of either orange or blue to distinguish them from others in the haplotype network and the bars above the STRUCTURE plots.

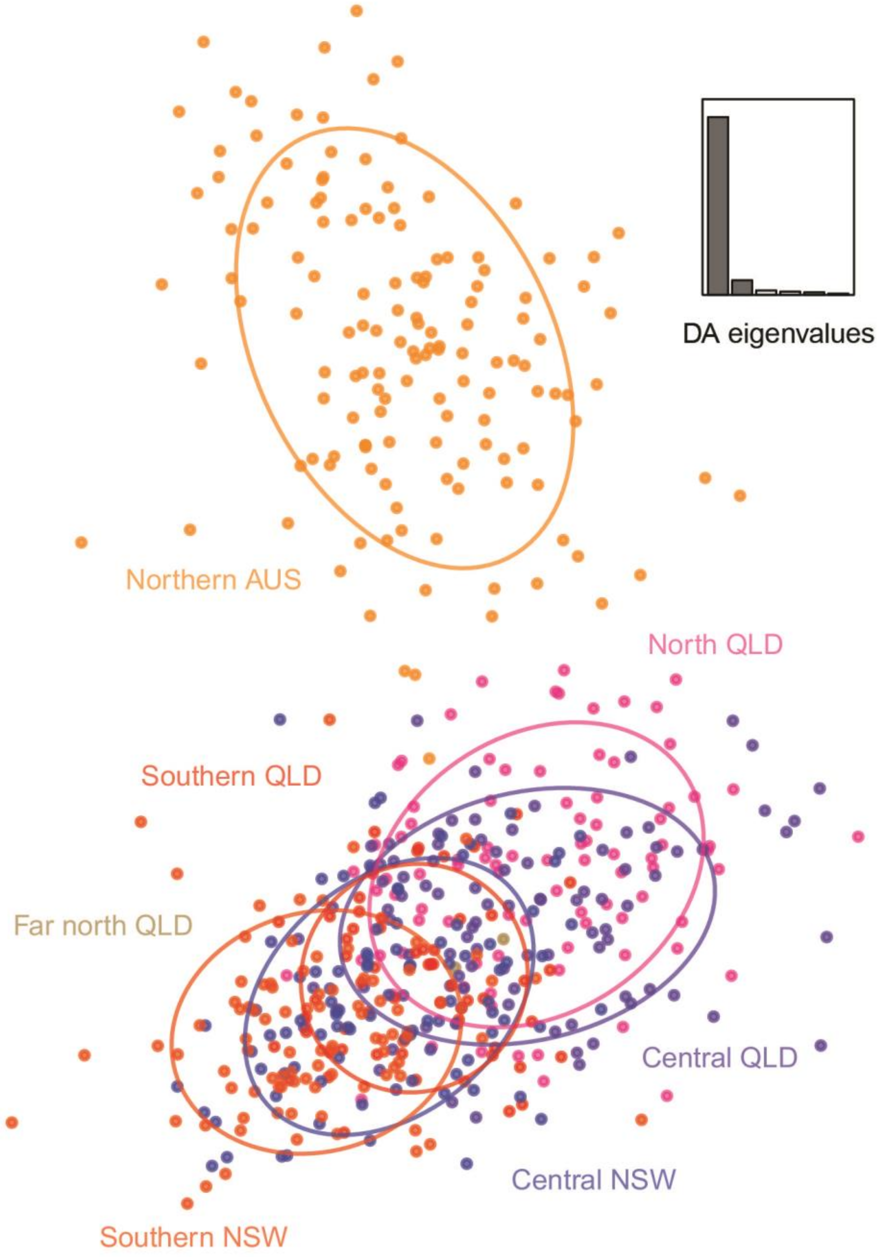


Figure 2.4. A Discriminatory Analysis of Principle Components (DAPC) using microsatellite genotype data from *Nezara viridula* sampled from Australia. Individuals in the DAPC are coloured according their region of origin rather than mitochondrial haplotype. See Table 2.1 for sample label state abbreviations.

2.4 Discussion

2.4.1 The species status of *Nezara viridula* and secondary contact in Australia

The two lineages of *N. viridula* have only recently been brought into secondary contact in Australia through human-assisted movement. Secondary contact, at least in eastern Australia, has resulted in mating between the two lineages as evidenced by the discordance in the mtDNA and nuDNA result. The mostly separate distribution of mtDNA haplotype groups A and E in eastern Australia indicates that secondary contact has taken place *in situ*, and the presence of both mtDNA groups is not the result of both having arrived in the same invasive population. The microsatellite results (nuDNA) show no barrier to gene flow between individuals with Asian lineage mtDNA and those with European mtDNA where they are found together. This result means that the mitochondrial haplotypes of individual *N. viridula* do not always reflect their nuclear genetic identity and evolutionary history and so we must be able to refer to the nuclear gene pools and the mitochondrial lineages separately. The two clades of *N. viridula* will therefore be referred to as *N. viridula* A (Asian) and *N. viridula* E (European), with the mitochondrial haplotype groups similarly referred to as mtDNA group A and mtDNA group E.

The frequency of mating events between *N. viridula* A and *N. viridula* E that has generated the pattern of nuDNA and mtDNA discordance is still unclear, and so the species status of the lineages remains uncertain. That two geographically and genetically distinct *N. viridula* populations occur in Australia (*N. viridula* A in the north and *N. viridula* E in the east) is clear from the K=2 STRUCTURE analysis and DAPC (Figures 2.2, 2.3 and S2.6). The microsatellite data show that gene flow between the populations regions is low – it has not resulted in widespread introgression in the nuclear genome. The geographical distance separating the north-western and eastern populations means that attributing low gene flow to mating differences is not appropriate because the frequency of movement between the genetically distinct eastern Australian and north-western Australian populations of *N. viridula* (Figure 2.2) is unknown. Further, although the K=2 STRUCTURE analysis shows some individuals of *N. viridula* E in eastern Australia assigned (with varying posterior probabilities) to the *N. viridula* A cluster, the K=3 analysis demonstrates that this assignment (at K=2) is in some part the result of further genetic structure among the eastern *N. viridula* E populations (Figure S2.4).

The two possible explanations for the observed pattern depend on the frequency at which mating occurs between *N. viridula* A and *N. viridula* E when and where they co-occur. (i) *N. viridula* represents two species and their secondary contact in Australia has led to rare hybridisation events.

This has resulted in a single nuclear gene pool, but with the introgression of mtDNA lineage A into northern populations of *N. viridula* E. Under this scenario the lack of *N. viridula* A individuals in northern Queensland requires explanation. There may be differences in the ecological requirements of *N. viridula* A and *N. viridula* E that cause them to be differentially distributed. Alternatively, the difference in their distribution could simply reflect the extent to which each gene pool has dispersed since their introduction to Australia. (ii) *Nezara viridula* represents two divergent sub-species or lineages of a single species that readily interbreed and so have completely introgressed through secondary contact in eastern Australia. If this explanation is correct, the lack of a strong signal of introgression in the nuclear genome of *N. viridula* in eastern Australia requires explanation. Perhaps this difference simply reflects the relative abundance of the two when secondary contact was first made, with few *N. viridula* A contributing to these founding populations. Alternatively, the source of mtDNA group A in northern Queensland may not be the populations sampled from north-western Australia, perhaps instead having arrived across the Torres Strait islands or a separate invasion into Queensland.

It is clear that *N. viridula* A and *N. viridula* E represent distinct evolutionary lineages but the extent to which the ecology of the different lineages of *N. viridula* might differ because of their separate evolutionary histories is not clear. No large differences in population genetic structure were found outside of the two main groups, and this indicates that there is no reduction in gene flow across *N. viridula* sampled from different host plants within the two groups. The exact nature of the relationship between *N. viridula* A and *N. viridula* E will only be resolved by mating experiments alongside the genetic analysis of populations on other continents where the two lineages might co-occur. Whether there are ecological differences associated with the genetic lineages will also require experimental investigation of *N. viridula* A and *N. viridula* E. The most conservative option is to treat them as meaningfully distinct so that their relationship can be established with more certainty. Spatially independent populations of *N. viridula* A and *N. viridula* E should be investigated as distinct entities, and the nuclear genetic composition of populations should be investigated alongside that of their mtDNA haplotypes. In particular, this should be investigated for *N. viridula* populations in Brazil and Japan, where secondary contact is also likely to have occurred between *N. viridula* A and E.

Bugs of *N. viridula* A were probably present in northern Australia prior to the invasion of the *N. viridula* E bugs. The considerably higher allelic diversity found across microsatellite loci in these insects (Table S2.5) supports this view, as does the greater number of individuals with non-dominant mtDNA haplotypes, and the presence of four colour morphs in these populations (G, O, F, R – (Kiritani 1970)) when compared with the single colour morph (G – (Kiritani 1970)) found in eastern Australia. The absence of colour morphs other than the G type is consistent with invasive populations of *N. viridula* on other continents (Yukawa & Kiritani 1965), and the lower genetic diversity is typical of invasive populations more generally (Cristescu 2015). Haplotypes belonging to both mtDNA groups A and E are found in *N. viridula* from eastern Australia but they are distributed separately except for a small area of overlap (Figure 2.2). It is unclear from these data why there is a boundary in the distribution of the mtDNA haplotype groups, but movement of this insect may be limited over large distances due to the ephemerality of available hosts (this is tested and discussed in Chapter 3). These differences may also reflect a difference in the number of founding individuals for each population (Blackburn *et al.* 2011), or there may be ongoing movement between northern Australian populations and those in south-east Asia.

2.4.2 The biogeographical history of the genus *Nezara*

The African lineage of *N. viridula* is represented by only a single individual from Botswana (Kavar *et al.* 2006) and this individual is the most genetically distant of all the COI haplotypes available from the genus *Nezara* (Figure 2.1). The African lineage of *N. viridula* is separated phylogenetically from the Asian and European lineages by the species *N. antennata* (Figure 2.1). Mating does occur between *N. antennata* and *N. viridula*, but these pairings do not produce offspring (Kiritani *et al.* 1963; Kon *et al.* 1988). The specimen from Botswana may represent a misidentified individual of another *Nezara* species rather than a new species within what is currently described as *N. viridula*, but *N. viridula* is widely reported from Africa (Afreh-Nuamah 1983; Ivbijaro & Bolaji 1990; Khaemba & Mutinga 1982; Poutouli 1995; Sithole *et al.* 1986). The conclusion that the African lineage represents a distinct species was also reached by Li *et al.* (2014) based on the same evidence, but more than the single sample from Botswana will be required from across Africa to test the relationship of African *N. viridula* to those found elsewhere.

Another distinct *Nezara* sequence is available from India (GenBank accession: KJ408787) from a COI barcoding project. As with the African lineage, this individual is more distantly related to *N. viridula* A and *N. viridula* E than to *N. antennata*, and so, based on this evidence, is probably a separate species. It is most likely that this individual has been misidentified and is in fact the species *N. yunnana* based on the location from which it was sampled, and the proximity of *N. yunnana* to *N. viridula* and *N. antennata* in the cladistics analysis of Ferrari *et al.* (2010). Also possible, but less probable, is that this specimen represents a previously undescribed species of *Nezara.* Although it is unclear what this specimen truly represents, it can aid in inferring the biogeographical history of the genus *Nezara,* as outlined below.

The centre of diversity for the genus *Nezara* is Africa but, excluding invasive populations, three of the four genetic lineages outside of Africa are found throughout Asia (Ferrari *et al.* 2010). These are *N. antennata, N. yunnana,* and *N. viridula* A. The exception is *N. viridula* E which is found throughout the Middle East, the Mediterranean, and Western Europe. All of the lineages found outside of Africa cluster together in the phylogenetic analysis of Figure 2.1, and diverged from one another following their divergence from the lineage represented by the specimen from Botswana. The genus *Nezara* has, therefore, been present and diversifying outside of Africa, probably in Asia, for millions of years (Kavar *et al.* 2006; Li *et al.* 2012). The most recent divergence event (Figure 2.1) among the *Nezara* lineages found outside of Africa is between *N. viridula* A and *N. viridula* E. Therefore, *N. viridula* E has resulted from the movement of this insect out of Asia, rather than out of Africa.

2.5 Conclusions

The global pattern of movement of *N. viridula* is relatively complicated due to the presence of two genetic lineages that co-occur in some instances, and which mate with one another at unknown frequencies. Previous phylogenetic studies only used mtDNA sequences (Kavar *et al.* 2006; Li *et al.* 2010), and here mtDNA haplotypes are shown to be potentially misleading regarding how they relate to the nuclear genome (and so also *N. viridula* A and *N. viridula* E). This means that the genetic relationship between contemporary populations as we now understand them may not be accurate, because much of this information relies on the mtDNA haplotypes. How geographically separated *N. viridula* populations might differ ecologically is uncertain as *N. viridula* populations may resemble *N. viridula* A, *N. viridula* E, or some combination of the two. This study does however provide a path forward. Mating tests should be carried out using individuals that represent *N. viridula* A and *N. viridula* E to determine whether mating is limited between them. Further, Brazilian and Japanese populations of this insect also possess different mitochondrial haplotypes, and so these populations offer a unique opportunity to investigate secondary contact under different circumstances. Once the nuclear genetic relationship of global populations is established, their ecology can be investigated with confidence.

2.6 Supporting information

Table S2.1. *Nezara viridula* samples that were sequenced and genotyped in this study. They are ordered by their sampling locality and defined by collection date, host plant(s), genes sequenced, and numbers genotyped at microsatellite loci (m). The site code links the information in this table to raw data files. For COI and Ef1α, numbers in brackets represent those individuals for which only a single direction was sequenced. Samples at all sites were collected by D. Brookes except those marked with an asterisk. Entries with multiple host plants represent sites where insects were collected from multiple host plants growing immediately next to one another, and so plants could not be sampled independently of one another. A dash (-) indicates that no samples from that host in that location were used in an analysis. Totals are COI: 176 (304), Tubα1: 178, Ef1α: 43 (129), and m: 571.

Site code	**Locality**	**Long.**	**Lat.**	**Date (dd-mm-yy)**	**Host plant/s**	**COI**	**Tubα1**	**Ef1α**	**m**
GH01	Kununurra	-15.720	128.691	08/*ix*/2014	*Vigna unguiculata* subsp. *sesquipedalis* (snake bean)	(16)	-	-	16
KU01	Kununurra	-15.720	128.691	08/*ix*/2014	*Momordica charantia* (bitter melon)	6 (16)	4	1 (4)	24
KU02	Kununurra	-15.652	128.707	09/*ix*/2014	*Salvia hispanica* (chia)	6 (17)	6	(6)	24
KU03	Kununurra	-15.652	128.707	09/*ix*/2014	*Carthamus tinctorius* (safflower)	6 (2)	6	2 (4)	8
KU04	Kununurra	-15.692	128.722	10/*ix*/2014	*Solanum lycopersicum* (tomato)	1	1	(1)	1
KU05	Kununurra	-15.692	128.722	10/*ix*/2014	*Phaseolus vulgaris* (green bean)	1	1	1	1
KU06	Kununurra	-15.692	128.721	10/*ix*/2014	*Zea mays* (sweet corn)	6 (9)	6	3 (3)	15
KU07	Kununurra	-15.692	128.721	10/*ix*/2014	*Ipomoea batatas* (sweet potato)	2	2	1 (1)	2
KU08	Kununurra	-15.652	128.708	11/*ix*/2014	*Gossypium hirsutum* (cotton)	6	6	5	25
KU09	Kununurra	-15.652	128.708	11/*ix*/2014	*Oryza sativa* (rice)	1	1	(1)	1
KU10	Kununurra	-15.617	128.753	10/*ix*/2014	*Zea mays* (sweet corn)	1	-	-	1
DW01	Darwin	-12.443	130.930	15/*ix*/2014	*Momordica charantia* (bitter melon)	6 (8)	6	4 (2)	16
DW02	Darwin	-12.443	130.930	15/*ix*/2014	*Solanum lycopersicum* (tomato)	1	1	(1)	1
DW03	Darwin	-12.443	130.930	15/*ix*/2014	*Abelmoschus esculentus* (okra)	2	2	(2)	2
DW04*	Darwin	-12.555	131.257	12/*v*/2015	*Passiflora edulis* (passionfruit)	4	4	1 (3)	4
LR01*	Lockhart	-12.800	143.316	20/*vii*/2015	*Cucurbita* sp.	1	-	-	2
TV01	Townsville	-19.285	146.822	29/*ix*/2014	*Solanum lycopersicum* (tomato)	6 (15)	6	3 (2)	24
GU01	Giru	-19.571	147.137	31/*ix*/2014	*Solanum nigrum*	1	-	-	8
BO01	Bowen	-20.078	148.151	31/*ix*/2014	*Solanum nigrum*	6 (6)	6	3 (3)	15
BO02	Bowen	-20.106	148.135	31/*ix*/2014	*Ricinus communis* (castor oil plant)	6	6	3 (3)	7
BO03	Bowen	-20.106	148.135	31/*ix*/2014	*Cleome gynandra* (tickweed)	5	5	1 (4)	5
BO04	Bowen	-20.010	148.189	01/*x*/2014	*Gossypium hirsutum* (cotton)	6 (13)	6	(4)	33
BI01	Biloela	-24.381	150.496	06/*ii*/2015	*Gossypium hirsutum* (cotton)	6 (8)	6	1 (5)	16

Continued

Table S2.1 continued.

BI02	Biloela	-24.379	150.494	06/*ii*/2015	*Cajanus cajun* (pigeon pea)	6 (5)	6	2 (3)	16
BI03	Biloela	-24.372	150.511	06/*ii*/2015	*Medicago sativa* (lucerne)	2	2	(2)	-
BI04	Biloela	-24.432	150.539	23/*xi*/2015	*Medicago sativa* (lucerne)	(22)	-	-	-
EM01	Emerald	-23.553	148.233	15/*xi*/2014	*Vigna radiata* (mung bean), *Arachis hypogaea* (peanuts)	5	6	2 (4)	18
EM02	Emerald	-23.578	148.178	15/*xi*/2014	*Medicago sativa* (lucerne)	3	3	(3)	1
EM03	Emerald	-23.518	148.203	05/*ii*/2015	*Cajanus cajun* (pigeon pea)	5 (13)	6	(6)	20
EM04	Emerald	-23.525	148.216	04/*ii*/2015	*Vigna radiata* (mung bean)	6 (14)	6	3 (2)	20
DA01	Dalby	-27.283	151.275	14/*iv*/2015	*Cajanus cajun* (pigeon pea)	6 (10)	6	1 (5)	16
DA02	Dalby	-27.283	151.275	14/*iv*/2015	*Gossypium hirsutum* (cotton)	6 (10)	6	(6)	16
DA03	Dalby	-27.687	151.307	02/*ix*/2015	*Medicago polymorpha*, *Raphanus raphanistrum*, *Vicia sativa* subsp. *sativa*	(19)	-	-	-
DA04	Dalby	-27.367	151.244	01/*ix*/2015	*Rapistrum rugosum* (turnipweed), *Xanthium* sp. (Noogoora burr)	(11)	-	-	-
DA05	Gatton	-27.541	152.337	23/*viii*/2015	*Amaranthus* sp., *Brassica* sp., *Malva parviflora*, *Medicago polymorpha*, *Medicago sativa*	(23)	-	-	24
DA06	Gatton	-27.248	151.282	01/*ix*/2015	*Rapistrum rugosum* (turnipweed)	(2)	-	-	-
DA07	Gatton	-27.799	152.106	11/*viii*/2015	*Rapistrum rugosum* (turnipweed)	(4)	-	-	-
NA01	Narrabri	-30.186	149.473	04/*xii*/2014	*Vigna radiata* (mung bean)	5 (17)	5	1 (5)	24
NA02	Narrabri	-30.305	149.657	03/*xii*/2014	*Helianthus annuus* (sunflower)	1	1	(1)	1
NA03	Narrabri	-30.542	150.010	04/*xii*/2014	*Medicago sativa* (lucerne)	5 (16)	6	1 (5)	24
NA04	Narrabri	-30.186	149.473	19/*ii*/2015	*Gossypium hirsutum* (cotton)	6 (17)	6	(6)	24
BR01	Breeza	-31.158	150.396	23/*vi*/2014	*Rapistrum rugosum* (turnipweed), *Xanthium* sp. (Noogoora burr)	6 (4)	6	1 (5)	10
BR02	Breeza	-31.222	150.468	05/*xii*/2014	*Xanthium* sp. (Noogoora burr)	6 (4)	6	(6)	10
BR03	Breeza	-31.158	150.396	23/*vi*/2014	*Urtica* sp. (stinging nettle)	(2)	-	-	2
GR01	Griffith	-34.419	146.363	24/*ii*/2015	*Cajanus cajun* (pigeon pea)	5	6	(6)	23
GR02	Griffith	-34.419	146.363	24/*ii*/2015	*Glycine max* (soybean)	5	6	1 (4)	8
GR04	Griffith	-34.441	146.037	25/*ii*/2015	*Medicago sativa* (lucerne)	6	6	1 (4)	25
DP02	Griffith	-34.596	145.977	25/*ii*/2015	*Cajanus cajun* (pigeon pea)	-	-	-	15
HA01	Hay	-34.472	144.753	25/*ii*/2015	*Gossypium hirsutum* (cotton)	2	2	(2)	2
HA02	Hay	-34.472	144.753	25/*ii*/2015	*Cajanus cajun* (pigeon pea)	5	6	1 (5)	21

Table S2.2. Source of all COI Sequence and their inclusion in the three COI analyses. Multiple analyses were performed because of the different fragment lengths of the COI gene sequences: (i) a 149 base pair (bp) overlapping region common to all sequences (overlapping fragment – OF, Figure 2.1), (ii) a 557 bp fragment of the more usual COI barcoding region (long fragment – LF, Figure S2.2), (iii) a 348bp fragment that was used in previous phylogenetic studies of *N. viridula* (short fragment – SF, Figure S2.1) (Kavar *et al.* 2006; Li *et al.* 2010). Both the LF and SF datasets are subsets of the OF dataset. The sequence labelled 'Other' in the location column was found in *N. viridula* from Brazil, California, Greece, Guadeloupe, Iran, Italy, Japan, Madeira, and Slovenia. Unpublished citations refer to sequences for which there are only GenBank accessions.

Species	**No. of Seq.**	**Host Plant/s**	**Location**	**COI Datasets**	**GenBank Accession No.**	**Citation**
N. viridula	176	Various (see Table S2.1)	Australia (see Table 2.1)	OF, LF	-	This study
N. viridula	5	Unknown	Egypt	OF, LF	-	This study
N. viridula	1	*Oryza sativa* (rice)	Philippines	OF, LF	-	This study
N. viridula	4	Unknown	Canada	OF, LF	KU601563:KU601566	Dhami *et al.* (2016)
N. viridula	1	Unknown	Canada	OF, LF	KF303511	Gariepy *et al.* (2014)
N. viridula	3	*Carya illinoinensis* (pecan)	Georgia, USA	OF, LF	KJ642018:KJ642020	Brown *et al.* unpublished
N. viridula	1	Unknown	Florida, USA	OF, LF	KR044112	Gwiazdowski *et al.* (2015)
N. viridula	1	Unknown	Texas, USA	OF, LF	KR037758	
N. viridula	4	*Gossypium hirsutum* (cotton)	Georgia, USA	OF, LF	JX548492:JX548495	Tillman *et al.* (2015)
N. viridula	3	Unknown	India	OF, LF	KR028339:KR028341	Reetha *et al.* unpublished
N. viridula	2	*Lablab purpureus*	India	OF, LF	KJ559399, KJ559406	Karthika and Krishnaveni unpublished
N. viridula	1	Unknown	India	OF, LF	HQ236460	Tembe *et al.* unpublished
Nezara sp.	1	Unknown	India	OF, LF	KJ408787	Kaur and Sharma unpublished
N. viridula	1	Unknown	India	OF, LF	GQ306225	Tembe *et al.* unpublished
N. viridula	1	Unknown	India		KJ866507	Rakshit *et al.* unpublished
N. viridula	1	Unknown	South Korea	OF, LF	GQ292245	Jung *et al.* (2011)
N. antennata	2	Unknown	South Korea	OF, LF	GQ292247, GQ292248	Jung *et al.* (2011)
N. antennata	1	Unknown	South Korea	OF, LF	KC135971	Jung unpublished
N. antennata	2	Unknown	China	OF, LF	FJ418861, FJ418862	Li *et al.* (2010)
N. viridula	4	Unknown	China	OF, SF	FJ418856:FJ418859	Li *et al.* (2010)

Continued

Table S2.2 continued.

N. viridula	1	Unknown	Iran		FJ418860	Li *et al.* (2010)
N. viridula	1	Unknown	Other	OF, SF	AY839161	Kavar *et al.* (2006); Li *et al.* (2010)
N. viridula	1	Unknown	Madeira	OF, SF	AY839162	Kavar *et al.* (2006)
N. viridula	1	Unknown	Brazil	OF, SF	AY839163	Kavar *et al.* (2006)
N. viridula	1	Unknown	Brazil	OF, SF	AY839164	Kavar *et al.* (2006)
N. viridula	1	Unknown	China, Japan	OF, SF	AY839165	Kavar *et al.* (2006); Li *et al.* (2010)
N. viridula	1	Unknown	Japan	OF, SF	AY839166	Kavar *et al.* (2006)
Nezara sp.	1	Unknown	Botswana	OF, SF	AY839167	Kavar *et al.* (2006)
N. viridula	1	Unknown	Japan	OF, SF	-	Muraji *et al.* (2001)

Table S2.3. Below the diagonal are pairwise F_{ST}s for all populations of *Nezara viridula* with sample size 19 or greater (Table 2.1) and using all 12 microsatellite loci. The results of pairwise exact G-tests for genotypic differentiation are shown above the diagonal, with levels of significance: * = 0.01 - 0.05, ** = 0.01 – 0.001, *** = >0.001. If more than one nearby location is included in a single population, the average pairwise geographical distance is used. Values that changed from significant to not significant, and *vice versa,* with the exclusion of NEZA05 and NEZA10 (Table S2.4) are highlighted in grey.

	KUN	**DAR**	**TOW**	**BOW**	**EMRa**	**EMRb**	**BIL**	**DAL**	**GAT**	**NARa**	**NARb**	**BBA**	**BRZ**	**GRIa**	**GRIb**	**HAY**
KUN	-	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
DAR	0.021	-	***	***	***	***	***	***	***	***	***	***	***	***	***	***
TOW	0.108	0.138	-	*	ns	**	*	***	***	***	***	***	***	***	***	***
BOW	0.140	0.169	0.011	-	ns	ns	ns	***	***	**	***	***	***	***	***	***
EMRa	0.144	0.179	0.012	-0.005	-	ns	ns	ns	*	ns	*	ns	***	***	***	***
EMRb	0.143	0.187	0.019	0.004	0.006	-	ns	***	***	**	***	***	***	***	***	***
BIL	0.150	0.184	0.012	-0.003	-0.014	0.002	-	*	*	*	*	***	***	***	***	***
DAL	0.204	0.234	0.049	0.021	0.009	0.032	0.011	-	ns	ns	ns	*	**	***	***	*
GAT	0.184	0.226	0.056	0.024	0.016	0.029	0.012	0.000	-	ns	*	***	***	***	***	**
NARa	0.190	0.230	0.039	0.015	0.012	0.014	0.008	-0.005	-0.005	-	ns	ns	*	**	**	**
NARb	0.195	0.238	0.034	0.027	0.012	0.027	0.007	0.008	0.013	0.003	-	*	*	***	***	**
BBA	0.188	0.232	0.053	0.022	0.006	0.031	0.017	0.007	0.015	0.004	0.011	-	**	***	***	**
BRZ	0.262	0.318	0.102	0.073	0.049	0.075	0.054	0.016	0.030	0.014	0.015	0.023	-	ns	ns	**
GRIa	0.265	0.332	0.114	0.071	0.050	0.073	0.062	0.021	0.036	0.024	0.035	0.024	-0.004	-	ns	ns
GRIb	0.267	0.323	0.116	0.080	0.070	0.072	0.071	0.033	0.039	0.026	0.047	0.044	0.021	0.015	-	*
HAY	0.225	0.281	0.062	0.045	0.037	0.039	0.035	0.011	0.029	0.016	0.016	0.017	0.024	0.009	0.022	-

Table S2.4. Below the diagonal are pairwise F_{ST}s for all populations of *Nezara viridula* with sample size 19 or greater (Table 2.1) and using 10 microsatellite loci, with loci NEZA05 and NEZA10 having been excluded for having moderately high null allele estimates (~10%, Table 2.2). The results of pairwise exact G-tests for genotypic differentiation are shown above the diagonal with levels of significance: * = 0.01 - 0.05, ** = 0.01 – 0.001, *** = >0.001. Values that changed from significant to not significant, and *vice versa*, with the exclusion of NEZA05 and NEZA10 are highlighted in grey.

	KUN	**DAR**	**TOW**	**BOW**	**EMRa**	**EMRb**	**BIL**	**DAL**	**GAT**	**NARa**	**NARb**	**BBA**	**BRZ**	**GRIa**	**GRIb**	**HAY**
KUN	-	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
DAR	0.02	-	***	***	***	***	***	***	***	***	***	***	***	***	***	***
TOW	0.11	0.13	-	*	ns	**	**	***	***	***	***	***	***	***	***	***
BOW	0.15	0.16	0.02	-	ns	ns	ns	***	***	***	***	***	***	***	***	***
EMRa	0.15	0.17	0.01	0.00	-	ns	ns	ns	**	ns	*	ns	***	***	***	**
EMRb	0.14	0.17	0.02	0.00	0.00	-	ns	***	***	***	***	***	***	***	***	***
BIL	0.16	0.18	0.02	0.00	-0.01	0.00	-	*	*	*	*	***	***	***	***	***
DAL	0.21	0.23	0.06	0.03	0.01	0.04	0.01	-	ns	ns	ns	*	**	**	***	ns
GAT	0.21	0.24	0.07	0.03	0.02	0.03	0.02	0.00	-	ns	*	***	***	***	**	ns
NARa	0.21	0.23	0.05	0.02	0.01	0.02	0.01	0.00	0.00	-	ns	ns	***	**	**	*
NARb	0.21	0.25	0.05	0.03	0.01	0.03	0.01	0.01	0.02	0.01	-	**	*	***	***	ns
BBA	0.20	0.23	0.07	0.03	0.01	0.03	0.02	0.01	0.02	0.00	0.01	-	**	***	***	*
BRZ	0.28	0.33	0.13	0.09	0.06	0.09	0.07	0.02	0.04	0.02	0.02	0.03	-	ns	ns	**
GRIa	0.27	0.32	0.13	0.08	0.05	0.08	0.07	0.02	0.03	0.02	0.03	0.02	-0.01	-	ns	ns
GRIb	0.27	0.31	0.12	0.07	0.05	0.07	0.06	0.02	0.02	0.01	0.04	0.03	0.01	0.00	-	*
HAY	0.21	0.24	0.06	0.03	0.02	0.04	0.02	0.00	0.01	0.01	0.00	0.00	0.02	0.01	0.01	-

Table S2.5. Allelic richness of *Nezara viridula* populations from northern Australia (141 individuals from northern NT and northern WA) and those populations from eastern Australia (383 individuals from all other regions but excluding populations LHR and GRU (Figure 2.2)). The motif length of the locus is shown in brackets following the locus name. The proportion of the total number of alleles for each group is show in brackets following the number of alleles for the northern and eastern populations. The total number of alleles found in all populations is also show. The specific base pair length of all alleles is shown in the final column, with a superscript 1 indicating alleles only found in northern populations, and a superscript 2 indicating alleles only found in eastern populations.

Locus	**Northern**	**Eastern**	**Total**	**Allele size (no. base pairs)**
NEZA01 (tri-)	5 (1.00)	3 (0.60)	5	179^1, 182, 185, 188, 191^1
NEZA02 (tri-)	5 (0.83)	4 (0.67)	6	155^1, 158^1, 164, 167, 170^2, 173
NEZA03 (tri-)	2 (1.00)	1 (0.50)	2	164, 167^1
NEZA04 (tri-)	8 (1.00)	5 (0.63)	8	196, 199, 202, 205^1, 208,211, 214^1, 220^1
NEZA05 (tetra-)	4 (1.00)	3 (0.75)	4	120, 124, 128, 136^1
NEZA06 (tri-)	4 (0.80)	4 (0.80)	5	312, 315, 318, 321, 324^1, 330^2
NEZA07 (tetra-)	3 (0.60)	5 (1.00)	5	218, 226, 230^2, 234, 238^2
NEZA08 (tri-)	3 (1.00)	2 (0.67)	3	204, 207, 210^1
NEZA09 (tetra-)	8 (1.00)	4 (0.50)	8	106^1, 110, 118, 122^1, 126, 130^1, 134^1, 138
NEZA10 (tri-)	5 (1.00)	5 (1.00)	5	137, 140, 143, 146, 149, 152
NEZA11 (tri-)	4 (1.00)	4 (1.00)	4	206, 209, 212, 215
NEZA12 (tri-)	3 (1.00)	3 (1.00)	3	262, 265, 268
Average across loci	4.5 (0.93)	3.8 (0.79)	4.8	

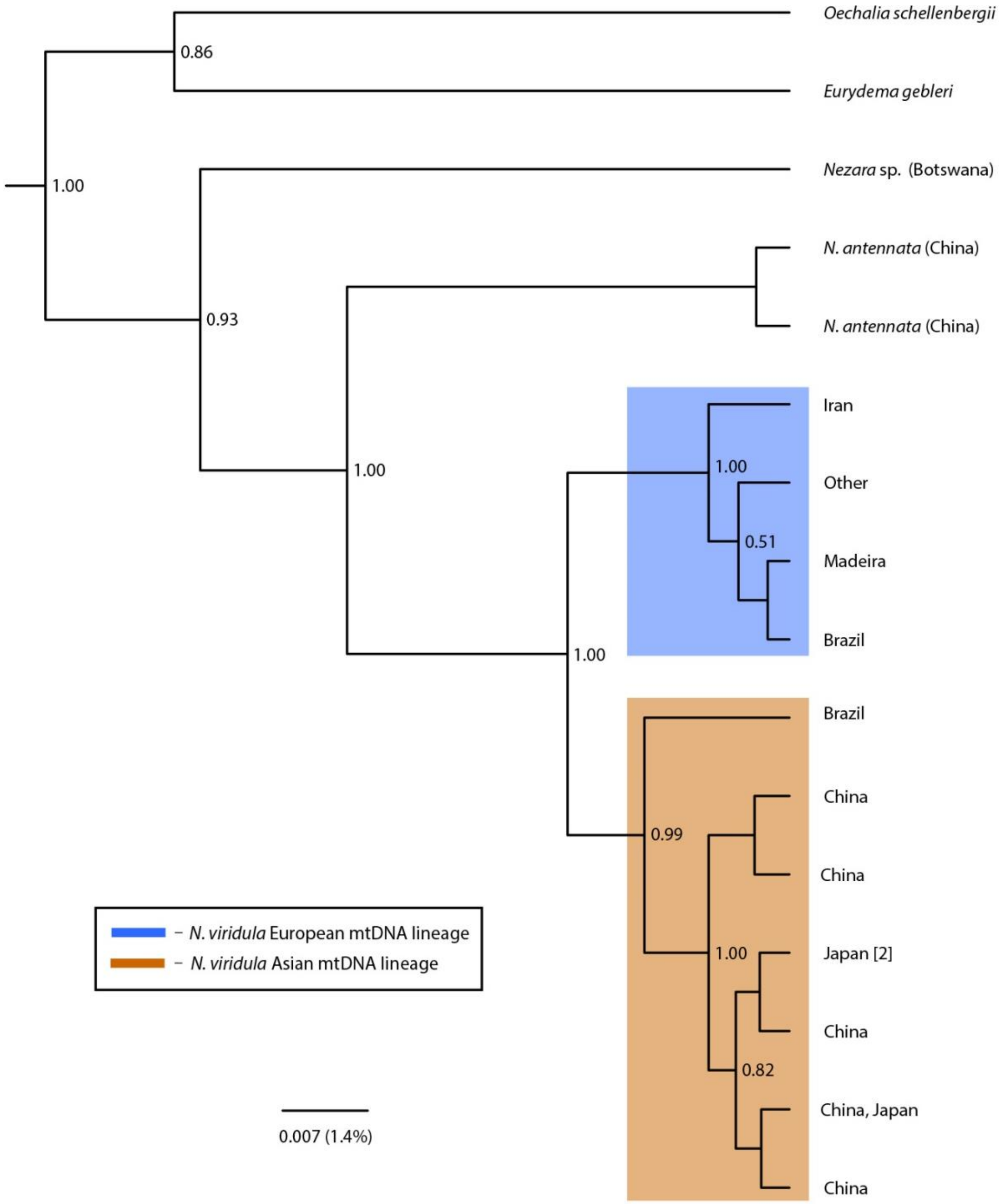


Figure S2.1. Tree representing the phylogeny of *Nezara viridula*, based on a MrBayes 3.2.6 analysis and using the 348 bp COI gene dataset (SF in the text) from all available individuals of both *N. viridula* and *N. antennata* (Table 2.1.3). Only unique sequences were used and if the number of samples that shared a sequence is greater than one then the number is shown in square brackets. The primary source of these sequences are the Kavar *et al.* (2006) and Li *et al.* (2010) papers, specifically outlined in Table S2.2. *Eurydema gebleri* and *Oechalia schellenbergii* (both also Pentatomidae) were included, with *O. schellenbergii* designated as the outgroup. The label ‘Other’ denotes a haplotype of *N. viridula*, common to individuals sampled from Greece, Guadeloupe, Galapagos, California, Japan, Italy and Brazil, from the Kavar (2006) dataset. Posterior probabilities are shown as node labels.

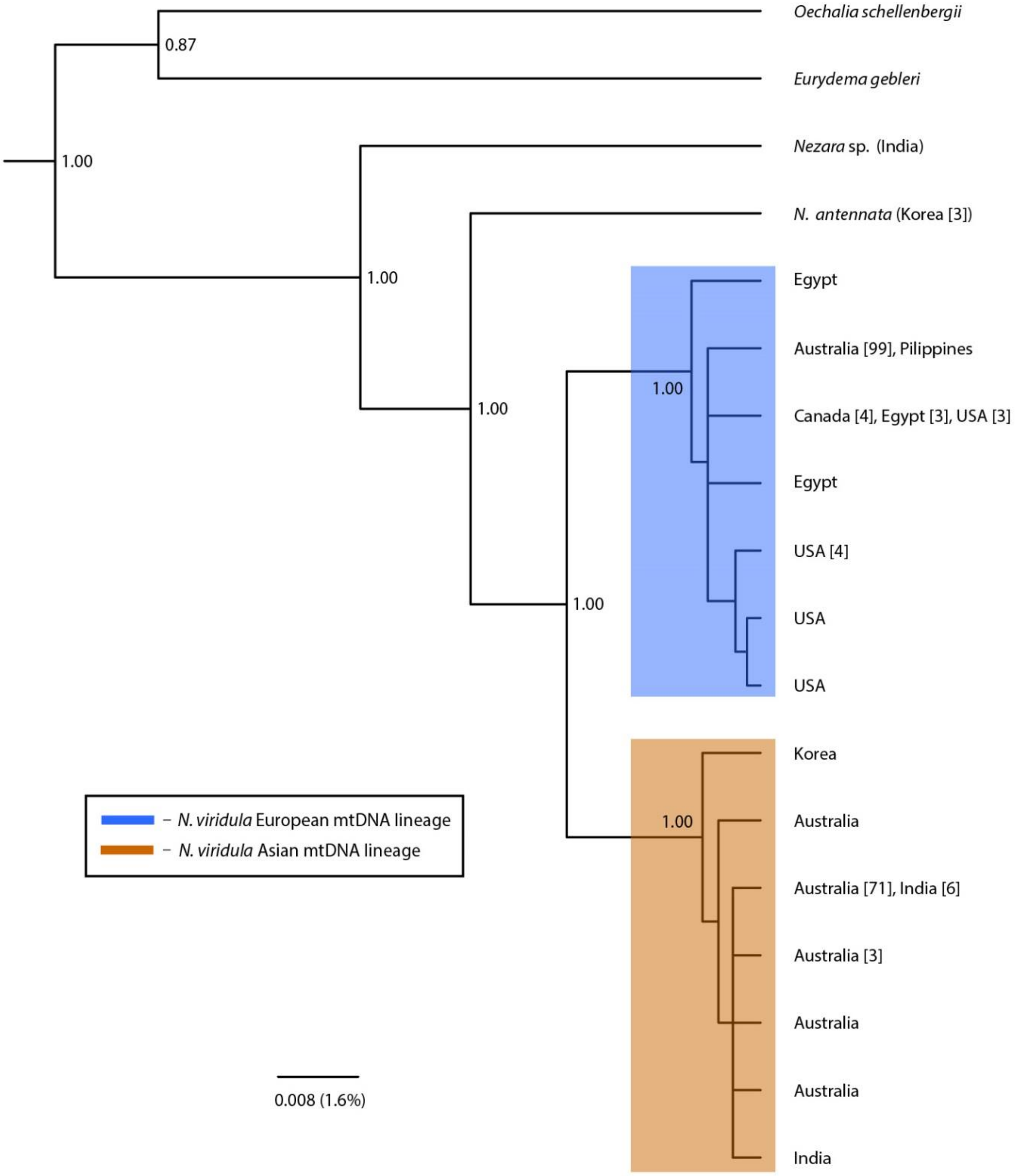


Figure S2.2. Tree representing the phylogeny of *Nezara viridula*, based on a MrBayes 3.2.6 analysis and using the 348 bp COI gene dataset (SF in the text) from all available individuals of both *N. viridula* and *N. antennata* (Table S2.2). Only unique sequences were used and if the number of samples that shared a sequence is greater than one then the number is shown in square brackets. The primary source of these sequences is this study and the DNA barcoding projects from various countries, specifically outlined in Table S2.2. *Eurydema gebleri* and *Oechalia schellenbergii* (both also Pentatomidae) were included, with *O. schellenbergii* designated as the outgroup. Posterior probabilities are shown as node labels.

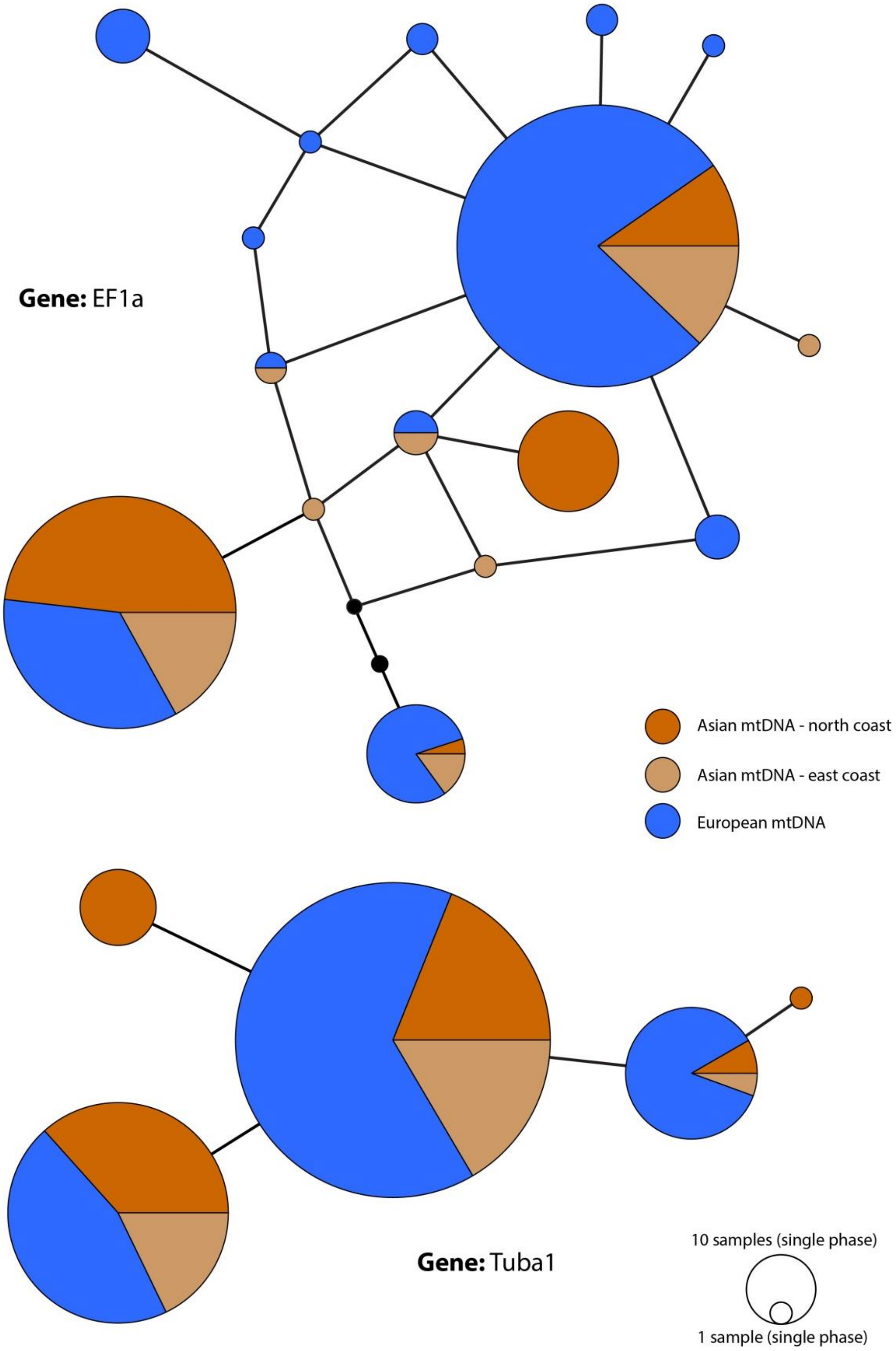


Figure S2.3. Haplotype networks for the EF1α and Tubα1 genes sequenced for individuals of *Nezara viridula* collected in Australia, with haplotype colour matched to the mitochondrial lineage to which each individual belongs (see key). Haplotypes that correspond to the Asian mtDNA lineage are further subdivided into those from individuals collected from northern Australia and those from eastern Australia (Figure 2.2). A single sample represents one phase of the nuDNA of a single individual as estimated by DnaSP 5.1 (Librado & Rozas 2009).

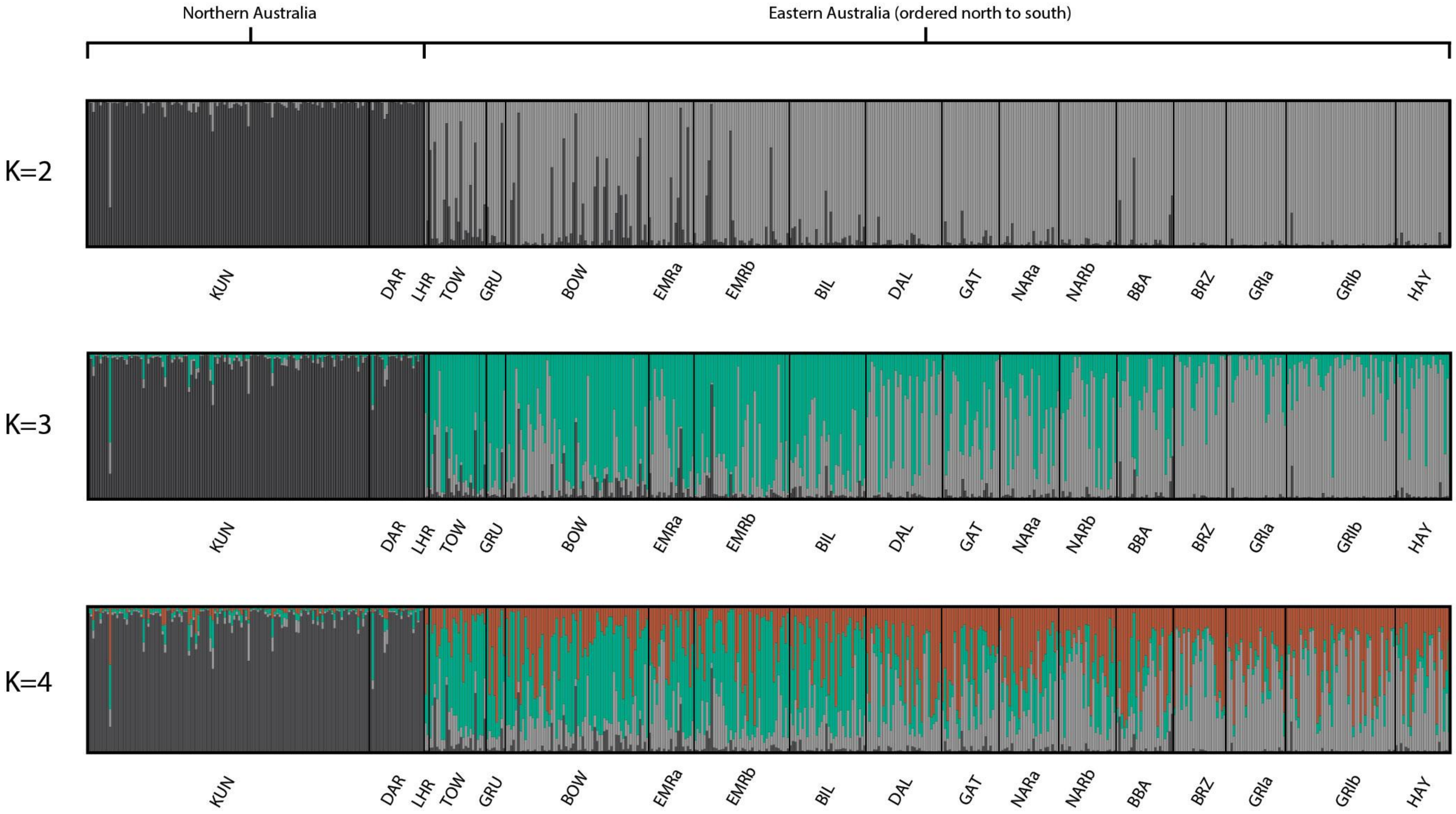


Figure S2.4. The results of the STRUCTURE analyses for genotyped Australian *Nezara viridula* individuals (Table S2.1) for K=2 through K=4 and labelled according to their sampling location. Sample sites are organised into two groups, the northern populations (Kununurra and Darwin) and eastern populations (all others to the right) (Figure 2.2). The eastern sites are then organized in geographical sequence from north to south. K=2 was the most likely value of K as per the Evanno (2005) method.

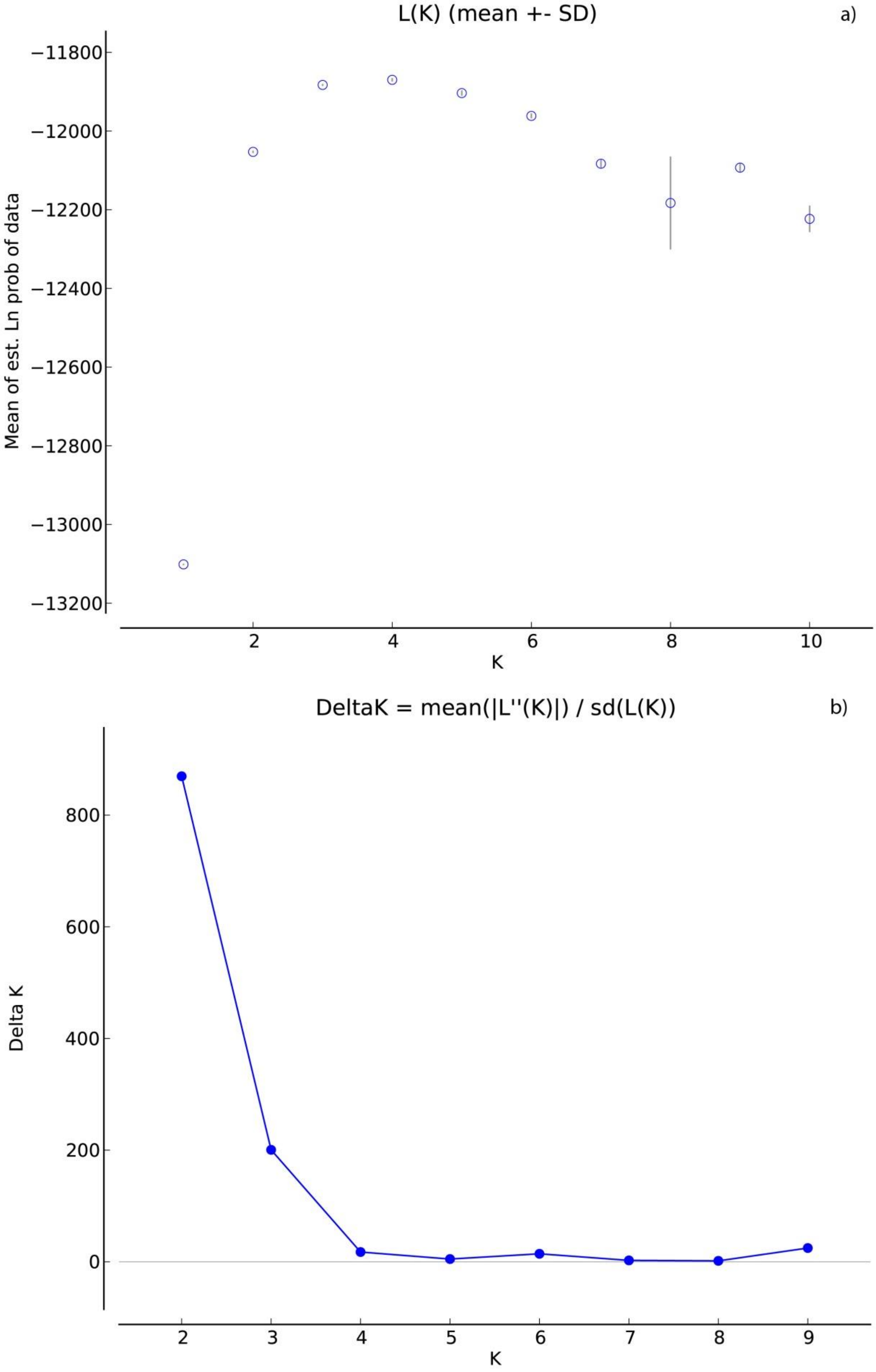


Figure S2.5. Output from STRUCTURE Harvester showing the likelihood of K (a) and Delta K (b). This indicates that K=2 is the most likely value of K.

Chapter Three – Spatiotemporal population genetics of *Nezara viridula* in eastern Australia relative to cotton

3.1 Introduction

Nezara viridula is an occasional pest of many crops in eastern Australia where it has highly variable seasonal and local abundance (Clarke 1992; Wilson *et al.* 2013). In recent years *N. viridula* has become a more significant pest of cotton (*Gossypium hirsutum*) in eastern Australia because of a reduction in the use of the broad-spectrum pesticides, that previously suppressed their numbers, following the widespread planting of transgenic *Bt* cotton (Trapero *et al.* 2016; Wilson *et al.* 2013). Predicting when and why *N. viridula* reaches numbers that require control is difficult because this insect is an extreme generalist that feeds on host plants from over 30 families (Todd 1989) and the spatiotemporal availability of host plants is not consistent (Rew *et al.* 2005; Tillman *et al.* 2009; Walker *et al.* 2005). This variability leads to uncertainty in interpreting the ecology of this insect (Wilson *et al.* 2013), particularly its pattern of dispersal over short and long times scales, and therefore the spatial scale at which host use influences its abundance in a given region. Cotton appears to be a good host for *N. viridula* (Olson *et al.* 2016; Olson *et al.* 2011), but only when fruits are at the right age for feeding. This means that host use and movement of *N. viridula* outside of the cotton system must influence the observed variability in the abundance and pressure of this insect within cotton.

The available evidence suggests that movement of *N. viridula* into cotton from other host plants can occur over small spatial scales (Olson *et al.* 2012; Tillman 2014; Tillman *et al.* 2014; Tillman *et al.* 2009), something that is found in similar pentatomid pests that include *Halyomorpha halys* (Nielsen *et al.* 2013; Venugopal *et al.* 2014; Wallner *et al.* 2014), *Chinavia hilaris* (Tillman *et al.* 2015), as well as *Euschistus conspersus* and *Thyanta pallidovirens* (Pease & Zalom 2010). Long distance and inter-annual movement of pentatomids, including *N. viridula*, has been given less attention. The variability in host plant abundance that *N. viridula* experiences may mean that long distance movement is an important part of their life history. One pentatomid, *H. halys,* has been shown to be capable of long distance movement, flying as much as 117 km in one day (on a flight mill), and with a mean distance of 2.4 km for males and 2.1 km for females (Lee 2015). The flight propensity and flight distance of *H. halys* also changes at different times of the year and at different temperatures (Lee 2015), and monitoring of its spread in the USA (Nielsen *et al.* 2013; Wallner *et al.* 2014) suggests that, while long-distance dispersal in pentatomids does occur, it may happen gradually.

In eastern Australia, *N. viridula* has a distribution that stretches over about 3000km from north to south, and here it feeds mostly on crops and agricultural weeds. The availability and suitability of each host plant that *N. viridula* is associated with varies spatially and temporally (Herbert & Toews 2012; Panizzi 1997), with each host species providing, in its different life stages, some combination of nutrition, water, shelter, and reproductive suitability (Panizzi 1997, 2000; Velasco & Walter 1992). Reproduction occurs on some host plants, but *N. viridula* may merely persist on others until a suitable reproductive host becomes available (Panizzi 2000; Velasco & Walter 1992), and this limits the number of major generations annually (Velasco *et al.* 1995). The sequential use of different host plants by *N. viridula* (Todd 1989) is thus necessary for *N. viridula* populations to persist, and population numbers may increase only on some of these. Investigating the timing and extent of *N. viridula* movement, relative to the distribution and availability of its hosts, is therefore crucial for understanding the ecology of this insect.

Use of a particular host species is spatially limited, and so the extent to which regional populations of *N. viridula* in eastern Australia are localised will affect the changing abundance of their populations. If the movement of *N. viridula* individuals is limited, *N. viridula* may rely on host plants being available in only a narrow area, and so their populations may experience frequent localised contractions in abundance. Alternatively, individuals of *N. viridula* may move considerable distances to find suitable host plants, similar to another Australian cotton pest bug, the mirid *Creontiades dilutus* (Hereward *et al.* 2013). In this case local abundance would be driven by their host interactions and immigration from distant areas. Further, *Nezara viridula* are reported to have developed resistance to insecticides in the past (Georghiou & Lagunes-Tejeda 1991; Sosa-Gómez *et al.* 2005). The extent of insect movement and the frequency of genetic bottlenecks will have implications for how resistance is managed (Wilson *et al.* 2013) should this become a problem again in the future.

Here I use microsatellites to genotype *N. viridula* individuals collected in eastern Australia to allow for inferences to be made about the extent of gene flow, local persistence, and connectivity of regional populations across their host plants. Assessment of gene flow will allow for the investigation of whether populations of *N. viridula* from different agricultural regions are genetically differentiated from one another or highly similar to one another. Sampling was performed across two different agricultural seasons to determine whether any observed patterns change over time. If *N. viridula* predominantly disperses over only short distances, and is restricted to agricultural areas by some combination of host availability and dispersal ability, then populations from eastern Australia should be strongly structured genetically across their range with very low temporal genetic differentiation within regions. In this case, host scarcity in winter, through cropping phases, or because of drought, may also cause *N. viridula* populations to experience regular genetic bottlenecks. If *N. viridula* regularly disperses across regions by active migration there should be little or no genetic differentiation across regions or seasons.

3.2 Methods and materials

3.2.1 Temporal sampling and microsatellite genotyping

Individuals of *N. viridula* E (Chapter 2) were collected into 100% ethanol from as many host plant species as possible, between spring 2014 and autumn 2016. Tables S2.1 and S2.2 show the samples that were used, their collection dates, and how populations were grouped for analyses. All samples taken from the same region at the same time were treated as a single population regardless of their host plants when collected. This is justified because the geographical distance separating sites within regions is negligible compared with the distance separating regions. Whether there might be any genetic different between *N. viridula* individuals collected from separate hosts was considered for all individual assignment (STRUCTURE and DAPC) analyses.

A total of 659 *N. viridula* individuals from 21 populations in eastern Australia were genotyped at 12 microsatellite loci, developed and sequenced as outlined in Appendices 2.2 and 2.3. The 21 populations used here are further grouped into ten localities in five regions – Bowen and Townsville in northern Queensland, Biloela and Emerald in central Queensland, Dalby and Gatton in southern Queensland, Narrabri and Baan Baa in northern New South Wales, as well as Griffith and Hay in southern New South Wales. Further population subdivisions represent populations sampled at different times from the same or similar locations. Summary microsatellite statistics can be found in Table 3.1. In Table 3.2 samples are separated within year into early, middle, or late year, to assess whether there are any genetic differences between *N. viridula* sampled at different times within a year.

3.2.2 Microsatellite analysis

Microsatellite peaks were confirmed and binned manually using the Geneious 9.0.5 microsatellite plugin 1.4.2. Null allele estimates and global F_{ST}, with and without ENA correction, were calculated using FreeNA (Chapuis & Estoup 2007). One locus, NEZA03, was fixed for a single allele in all populations and so was excluded from further analysis. Deviations from Hardy-Weinberg Equilibrium (HWE) were measured using Hardy-Weinberg exact tests in Genepop 4.2 (Raymond & Rousset 1995;

Rousset 2008), as were estimates of linkage disequilibrium. Sequential Bonferroni correction was performed manually.

Table 3.1. Statistics of the microsatellite loci included in the population genetic analysis of *Nezara viridula.* Only samples collected from the 17 populations with 19 or more individuals were used to generate these statistics (see Table 3.1), with the exception of the total number of alleles (∑Na).

Locus	∑*Na*[a]	*Ho*[b]	*He*[c]	HWE[d]	Est. Null Alleles	gF_{ST}	gF_{ST} ENA
NEZA01 (51)	3	0.597	0.579	0 (0)	0.01	0.01	0.01
NEZA02 (82)	4	0.526	0.518	1 (0)	0.02	0.02	0.02
NEZA03 (55)	1	-	-	-	-	-	-
NEZA04 (77)	6	0.717	0.668	2 (0)	0.01	0.02	0.02
NEZA05 (89)	3	0.296	0.429	4 (0)	0.10	0.02	0.02
NEZA06 (50)	6	0.626	0.604	2 (1)	0.02	0.01	0.01
NEZA07 (96)	5	0.437	0.530	4 (1)	0.08	0.05	0.04
NEZA08 (127)	2	0.085	0.091	1 (0)	0.01	0.01	0.02
NEZA09 (95)	4	0.608	0.561	0 (0)	0.00	0.01	0.01
NEZA10 (86)	5	0.536	0.670	5 (2)	0.08	0.01	0.01
NEZA11 (52)	4	0.662	0.647	2 (0)	0.01	0.00	0.00
NEZA12 (78)	3	0.607	0.605	2 (0)	0.03	0.03	0.03

[a] Total number of alleles.

[b] Average observed heterozygosity.

[c] Average expected heterozygosity.

[d] Deviations from Hardy-Weinberg Equilibrium and their within loci Bonferroni corrected values shown in bracket

Locus by locus Analysis of Molecular Variance (AMOVA (Excoffier *et al.* 1992)) using number of different alleles (F_{ST}-like) to examine the contributions of space and time to genetic structure using Arlequin 3.5 (Excoffier & Lischer 2010) with 1000 permutations. Two AMOVAs were performed, one with populations and no higher grouping, and a second in which populations were grouped into one of two seasons (as per Table 3.2). Population structure was further assessed using Discriminatory Analysis of Principle Coordinates (DAPC) which were performed using the *adegenet* package (Jombart *et al.* 2010) for R (R 2016), and using 100 cross-validation replicates to choose the number of principle and discriminatory coordinates retained. DAPC were performed on all populations grouped regionally rather than temporally, as well as within regions across different sampling times. Tests for the genetic effects of recent population bottlenecks (heterozygote excess) were conducted with Bottleneck 1.2.02

(Piry *et al.* 1999) using Wilcoxon tests and the two-phase model (TPM) with 0.36 variance and a single-step mutation proportion of 95% due to the relatively low allelic diversity present in these populations. Wilcoxon tests were used because they are more appropriate with a lower number of loci (Piry *et al.* 1999). Bottleneck uses a heterozygote-excess test that is sensitive to the timing of any genetic bottleneck and so should detect only more recent genetic bottlenecks (Peery *et al.* 2012). Further, Bottleneck is more effective when the number of alleles was low prior to any genetic bottleneck occurring (Peery *et al.* 2012). The arrival of *N. viridula* in Australia has been recent and probably coincided with a genetic bottleneck. This means that the heterozygote-excess method is appropriate for testing for the more recent genetic bottlenecks that are of interest.

Table 3.2. Groupings of sample sites for the genetic analysis of *Nezara viridula*. Population names represent multiple samples taken from nearby sites at similar times. The sites and host plants that correspond to each population name can be found in Table S3.1. Population names with an asterisk are spatial samples only. Regional samples include total the following number of sites and individuals – (i) northern QLD, two populations and 83 individuals, (ii) central Queensland, five populations and 146 individuals, (iii) southern QLD, four populations and 129 individuals, (iv) central NSW, six populations and 157 individuals, and (v) southern NSW, four populations and 144 individuals. The column ‘n’ shows the number of *N. viridula* genotyped from each population. Sample labels with state abbreviations are as follows: NSW – New South Wales, and QLD – Queensland.

Pop. name	**n**	**Region**	**Season**	**Lat.**	**Long.**
EMR-Y1a	19	Central QLD	2014-2015	-23.549	148.202
EMR-Y1b	40	Central QLD	2014-2015	-23.549	148.202
EMR-Y2	32	Central QLD	2015-2016	-23.518	148.203
BIL-Y1	32	Central QLD	2014-2015	-24.404	150.520
BIL-Y2	23	Central QLD	2015-2016	-24.432	150.539
DAL-Y1	32	Southern QLD	2014-2015	-27.283	151.275
DAL-Y2	32	Southern QLD	2015-2016	-27.520	151.261
GAT-Y2a	31	Southern QLD	2015-2016	-27.541	152.337
GAT-Y2b	34	Southern QLD	2015-2016	-27.960	152.080
NAR-Y1a	25	Central NSW	2014-2015	-30.255	149.553
NAR-Y1b	24	Central NSW	2014-2015	-30.255	149.553
NAR-Y2	32	Central NSW	2015-2016	-30.206	149.602
BBA-Y1	24	Central NSW	2014-2015	-30.542	150.010
BBA-Y2	30	Central NSW	2015-2016	-30.605	150.025
GRI-Y1	71	Southern NSW	2014-2015	-34.419	146.363
GRI-Y2	13	Southern NSW	2015-2016	-34.610	145.947
HAY-Y1	23	Southern NSW	2014-2015	-34.472	144.753
HAY-Y2	37	Southern NSW	2015-2016	-34.472	144.753
BRZ-Y1	22	Central NSW	2014-2015	-31.186	150.433
BOW-Y1	60	Northern QLD	2014-2015	-20.052	148.161
TOW-Y1	24	Northern QLD	2014-2015	-19.285	146.822

Pairwise population F_{ST}s, as well as exact G-tests for population differentiation, were computed in Genepop 4.2 (Raymond & Rousset 1995; Rousset 2008). Two loci, NEZA05 and NEZA10, were previously determined to have an effect on pairwise F_{ST}s (Chapter 2). G-test results were examined for each locus to assess their relative impact on the result. Pairwise F_{ST}s were used in Mantel tests to estimate isolation by distance (IBD) using IBD Web Service 3.23 (Jensen *et al.* 2005) with 10,000 randomizations. Three IBD analyses were performed including: (i) all populations with individuals sampled grouped together regardless of sampling time, (ii) populations from the late 2014 and early 2015, and (iii) samples from late 2015, early 2016, but also northern Queensland samples (TOW and BOW Table S3.1). Each of these IBD analyses was performed with raw geographic distance and with the log of geographic distance.

3.3 Results

3.3.1 Summary microsatellite statistics

Twelve microsatellite loci were used in this study, these having been developed on individuals of *N. viridula* from both *N. viridula* A and *N. viridula* E (Appendix 2.2). Nine of these loci showed evidence for deviation from Hardy-Weinberg equilibrium (HWE) in at least one population, although this was reduced to a total of four populations across three loci after sequential Bonferroni correction (Holm 1979). No loci showed evidence for linkage disequilibrium. One locus, NEZA03, was monomorphic for all genotyped individuals, and three loci had relatively high null allele estimates (Table 3.1) but none of these was greater than 10%.

3.3.2 Spatial and temporal genetic structure

Genetic differentiation is low between populations of *N. viridula* sampled north to south over about 2500km in eastern Australia (Figure 3.1), as indicated by the results of STRUCTURE analyses and DAPC. Over this distance there is a significant correlation between geographic and genetic distance as indicated by the IBD analysis over all samples (Figure 3.2). However, all analyses indicate that gene flow occurs between all populations. This spatial genetic was also evident when the analyses were carried out with samples from each season (Figures S2.1 and S2.2). Within each of the sampled regions there was no further genetic structure detected in STRUCTURE, but DAPC did detect some small differences between populations (outlier circles in Figure 3.1).

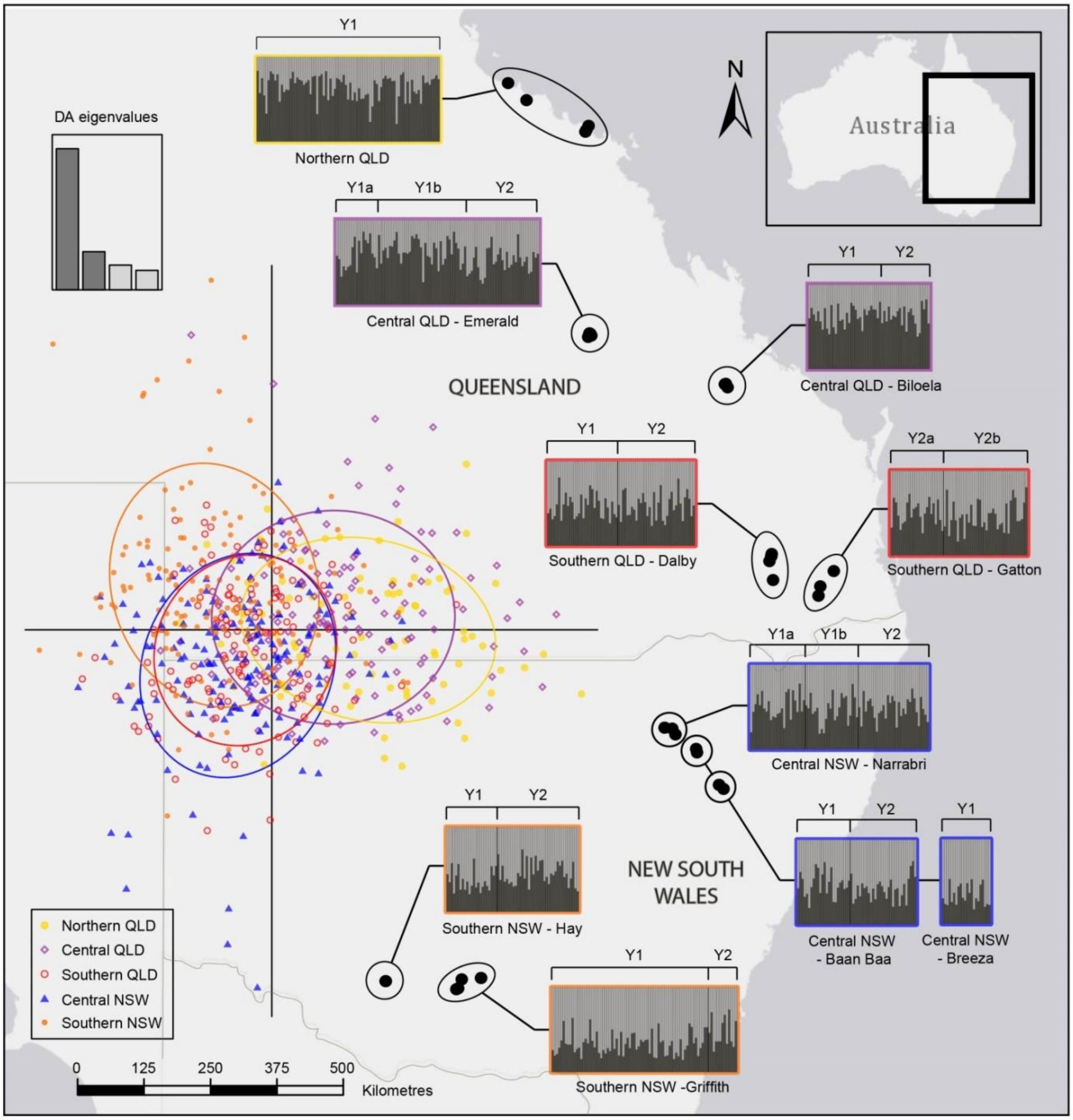


Figure 3.1. A map of genotyped *Nezara viridula* individuals grouped according to region and the season they were sampled. Populations represented by the STRUCTURE analysis boxes are labelled Y1 if collected during the 2014-2015 season, and Y2 if collected during the 2015-2016 season. A further label of ‘a’ or ‘b’ designates samples taken at different times within a season. Site names represent broad localities within regions. The scatter plot shows a Discriminatory Analysis of Principle Components (DAPC) with populations grouped according to region regardless of sampling time. The boxes represent a K=2 STRUCTURE analysis of the same genotyped individuals used in the DAPC.

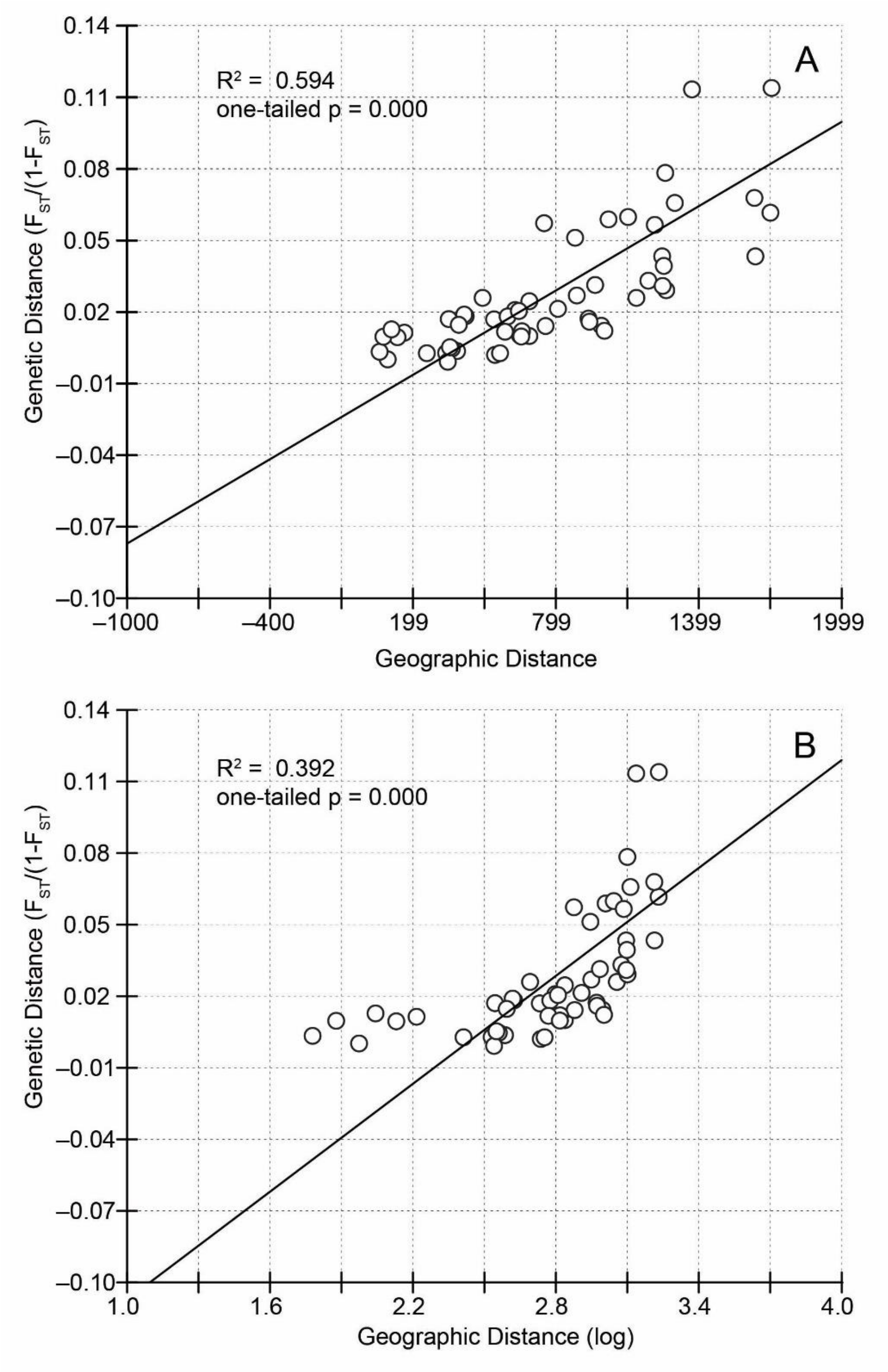


Figure 3.2. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* individuals (Table 3.2), with individuals grouped according to sample location irrespective of time sampled. Geographic distance is used in Figure 3.1A and log transformed geographic distance is used in Figure 3.1B.

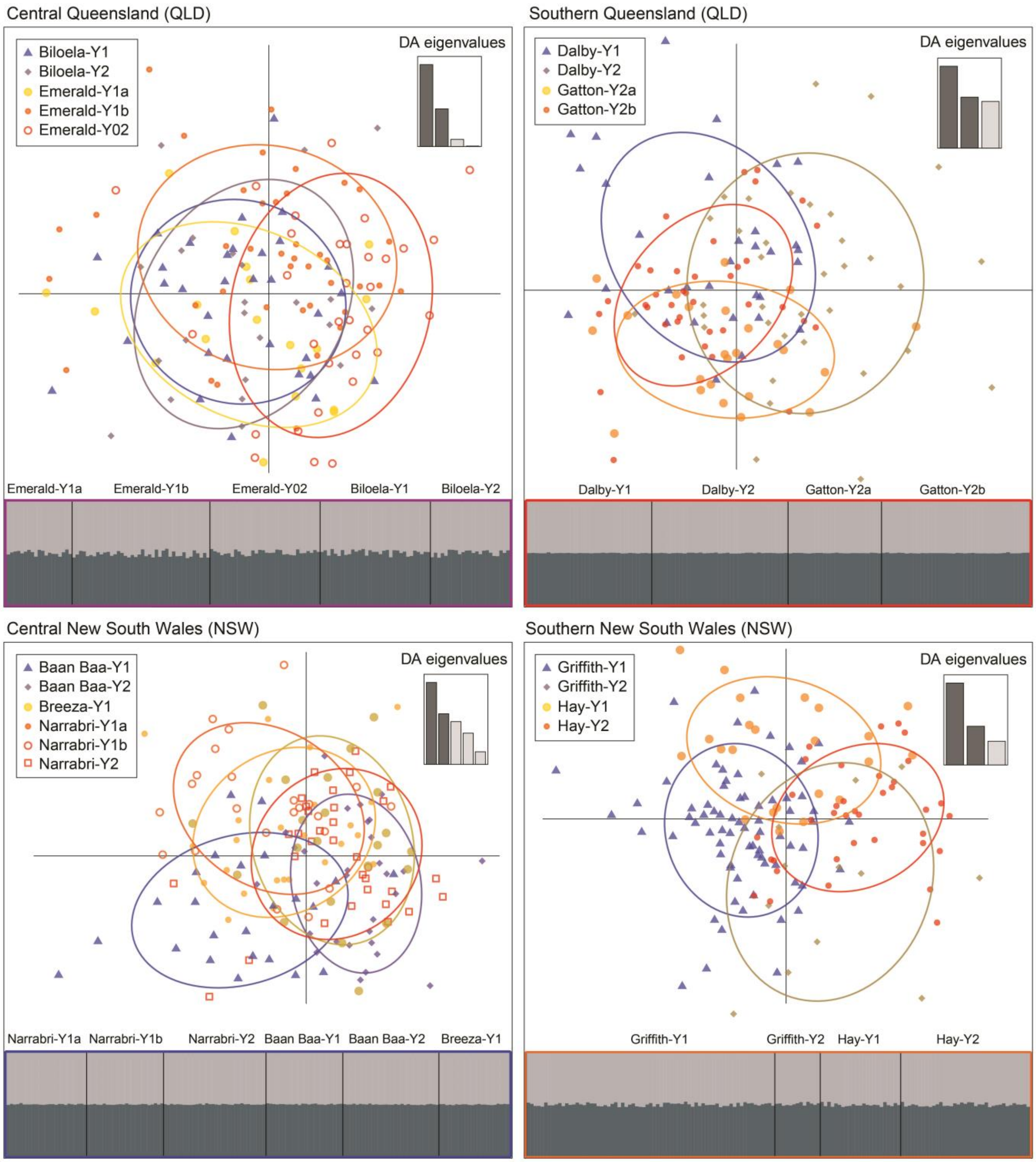


Figure 3.3. Within-region analyses of genotyped *Nezara viridula* individuals using Discriminatory Analysis of Principle Coordinates (DAPC) and K=2 STRUCTURE analyses. *Nezara viridula* individuals are grouped within regions according to the season in which they were sampled. Populations sampled during the 2014-2015 season are labelled Y1, and those from the 2015-2016 season labelled Y2. Labels with an ‘a’ or ‘b’ indicate population samples taken at different times within a season. See Figure 3.1 for these same analyses but conducted with all populations, rather than within regions, and for a map of the sample locations. The boxes of STRUCTURE analyses here are coloured the same as in Figure 3.1.

3.3.3 Temporal genetic structure and genetic bottlenecks

In Bottleneck analyses most *N. viridula* populations showed significant evidence for genetic bottlenecks – 15 of 21 when using the TPM and 13 of 21 when using the SSM (Table 3.3). The time of sampling, whether, early, middle, or late season, did not impact whether a genetic bottleneck occurred (Table 3.3). None of the genetic diversity in *N. viridula* populations was partitioned among seasons in AMOVA (Table 3.4). For individuals sampled at different times there was no difference in their population assignment for each within-region STRUCTURE analysis (Figure 3.3). Samples taken from similar locations at different times were genetically distinct from one another in four out of 11 instances according to pairwise exact G-tests (Table 3.2), but with population GR-S2 excluded because of the low number of individuals sampled.

Table 3.3. The results of an analysis of *Nezara viridula* microsatellite genotypes using the software Bottleneck 1.2.02 to perform Wilcoxon tests for heterozygosity excess (one-sided p) and for both heterozygosity excess and heterozygosity deficiency (two-sided p). Time sampled is divided into three categories, early (E – August, September, and October), middle (M – November, December, January), and late (L – February, March, April). In this regard, populations TOW and BOW are excluded because of regional climatic differences and BRZ is excluded because the samples that constitute that population do not represent a single sample time. Bottleneck parameters were a variance of 0.36 and using both the two-phase mutation model (TPM) and single step mutation model (SMM). The TPM used a single-step mutation proportion of 95%.

Site Code	Time Sampled	TPM one-sided p	TPM two-sided p	SSM one-sided p	SSM two-sided p
TOW	-	0.010	0.021	0.027	0.054
BOW	-	0.002	0.003	0.002	0.003
EMR01a	M	0.074	0.147	0.074	0.147
EMR01b	L	0.000	0.001	0.000	0.001
EMR02	L	0.074	0.147	0.074	0.147
BIL01	L	0.006	0.012	0.006	0.012
BIL02	M	0.051	0.102	0.074	0.147
DAL01	L	0.006	0.012	0.006	0.012
DAL02	E	0.002	0.003	0.002	0.003
GAT02a	E	0.042	0.083	0.051	0.102
GAT02b	L	0.006	0.012	0.006	0.012
NAR01a	M	0.042	0.083	0.051	0.102
NAR01b	L	0.000	0.001	0.000	0.001
NAR02	L	0.120	0.240	0.139	0.278
BBA01	M	0.002	0.005	0.003	0.007
BBA02	L	0.042	0.083	0.042	0.083
BRZ	-	0.289	0.577	0.350	0.700
GRI01	L	0.087	0.175	0.120	0.240
GRI02	L	0.034	0.067	0.034	0.067
HAY01	L	0.034	0.067	0.042	0.083
HAY02	L	0.010	0.021	0.010	0.021

Table 3.4. AMOVA were conducted to assess that partitioning of *Nezara viridula* genetic differences, primarily among populations and among two seasons. Global locus by locus AMOVA results are presented for an F_{ST}-like analysis conducted in Arlequin (Excoffier & Lischer 2010). AMOVA a) includes all populations of *N. viridula* (Table 3.2), AMOVA b) includes populations from regions that were sampled in both the 2014-2015 and the 2015-2016 seasons (so excluding TOW and BOW) and with populations separated into groups according to season.

Source of variation	Sum of Squares	Variance explained	F-statistics	P val.
a) all populations				
Among populations	148.9	2.4%	0.023	0.000
Among individuals within populations	1921.5	5.6%	0.057	0.000
Within individuals	1776.5	92.1%	0.079	0.000
b) temporally sampled populations				
Among seasons	6.2	0.0%	0.000	0.712
Among populations	110.2	1.9%	0.019	0.000
Among individuals within populations	1673.3	5.6%	0.057	0.000
Within individuals	1550.0	92.6%	0.074	0.000

3.4 Discussion

3.4.1 Connectivity and persistence of *N. viridula* populations

Populations of *N. viridula* in eastern Australia are relatively localized despite there being a high amount of gene flow between them over the long term. This is evident in the strong pattern of IBD (Figure 3.2) that is present in at least two growing seasons (Figures S2.1 and S2.2). This suggests that the regional abundance of *N. viridula* will be predominantly driven by regional host availability rather than the movement of populations over large distances. However, what constitutes a single region is unclear from the available data as insects from some regions are very similar to one another in pairwise exact G-tests (Table 3.2) and DAPC (Figure 3.1). These analyses show all within-region populations are somewhat similar to one another (Figure 3.1). Where populations within regions do differ from each other, it is not clear whether this is a consequence of the apparently frequent genetic bottlenecks, the result of *N. viridula* movement into an area, the result of a low sample size, or some combination of these.

Fluctuations in host availability within a year are probably responsible for *N. viridula* undergoing regular genetic bottlenecks (Table 3.3). Populations may increase in size rapidly when suitable host plants are available, as *N. viridula* females lay egg masses of about 80 eggs (Panizzi 2000; Todd 1989). Sequential host use is however crucial for the longer-term persistence of *N. viridula* populations (Todd 1989) but not all hosts are equally suitable for oviposition or nymphal survival (Panizzi 2000; Velasco *et al.* 1995) and some plants called hosts are entirely unsuitable in this regard (Velasco *et al.* 1995). In eastern Australia none of the host plants that *N. viridula* uses are available year-round, and so dramatic changes in host availability may occur at multiple times – at the onset of winter (when these insects undergo diapause), in early summer as the environment dries and spring weeds become less abundant, and throughout the year as various crops are harvested. It has been suggested that *N. viridula* completes two major generations in south-east Queensland (Velasco *et al.* 1995), and if few of the host plants are available to complete these generations or allow for the persistence of *N. viridula*, the abundance of *N. viridula* will not remain high through the entire year.

Crucial to the interpretation of these data is that these samples were collected during a period when pest pressure from *N. viridula* was generally low in the areas sampled in eastern Australia (Table S3.1). Indeed, in almost all areas these insects were in low numbers and difficult to find compared with previous years during which they were studied (Wilson 2010), and no insects were found outside of agricultural areas despite the native pentatomids *Plautia affinis* and *Piezodorus oceanicus* being found there on host plants they share with *N. viridula*, particularly *Medicago sativa* (pers. obs.). The results presented in this paper are thus representative of *N. viridula* populations when they are at relatively low abundance and so this situation is indicative of the conditions that might cause their numbers to remain low rather than increase. The conditions during sampling were generally dry, and outside of crops, *N. viridula* populations appeared restricted to narrow bands of vegetation along riversides or drainage areas.

The results presented here for *N. viridula* show a vastly different pattern of genetic structure to that of another cotton pest bug, the native green mirid *Creontiades dilutus*. This has important implications for the management of *N. viridula*. The abundance of *C. dilutus* in cotton appears to be driven to some extent by the use of host plants well outside of cotton growing regions (Hereward *et al.* 2013). Coupled with long distance movement, the small amount of spatial genetic structure observed among the populations of *C. dilutus* does not persist through time (Hereward *et al.* 2013). This means that environmental conditions well away from the regions identified in Figure 3.1 are likely to have a stronger impact on the local abundance of *C. dilutus* in these regions when compared with *N. viridula*. Understanding what factors lead to high abundance must therefore be investigated independently for the major growing regions where *N. viridula* is a pest in order to better predict years of high and low abundance. Further, if *N. viridula* becomes a more frequent pest it may be at risk of developing resistance to insecticides because its populations remain relatively localized and experience frequent genetic bottlenecks. At present, any alleles that confer resistance to insecticides may be diluted during years when *N. viridula* is not in high abundance.

3.4.2 Implications for investigating the host use of *N. viridula*

The sequential and seasonal host use of *N. viridula* needs to be understood on a regional basis, in a similar way to the studies conducted by Velasco *et al.* (1995) and Wilson (2010). The data presented here, however, can only suggest what the extent of these regions might be, as there is not enough genetic resolution or temporal sampling to delimit them and understand their persistence in these regions. Population connectivity and host availability may differ significantly in years where *N. viridula* is a more significant pest and so it will be crucial to investigate how the population genetic structure of *N. viridula* changes in eastern Australia as environmental and agricultural conditions change. None of the microsatellite loci used here have alleles that are unique to particular regions, and so ideally a different molecular technique would be used in future, one that might offer a greater amount of variation. Restriction site associated DNA sequencing (RADSeq) techniques have successful been tested for *N. viridula* (Brookes unpublished), and this technique will be used in future to assess gene flow across the same samples used here (Table S3.1) and future samples taken under different conditions. The frequent genetic bottlenecks observed in the microsatellite data presented here may cause local populations to be genetically distinct enough that their movement can be effectively monitored in the long term with a large number of loci. Such methods could also effectively characterise population genetic structure using fewer individuals during periods when *N. viridula* is in low abundance.

The genetic background of *N. viridula* populations may also influence host use (Chapter 2), as global populations of *N. viridula* are derived from two evolutionary lineages, the Asian and the European (Kavar *et al.* 2006; Li *et al.* 2010; Chapter 2), whose host use may differ in some respects. Both lineages are present in Australia and mating between these lineages has occurred (Chapter 2). The extent to which these mating events have impacted the ecology of Australian *N. viridula* in eastern Australia is uncertain (Chapter 2). Several features in the life of these insects contribute to the difficulty in understanding the host use of this bug, including the existence of divergent genetic lineages with potential differences in host use, variable contribution of these genetic lineages to any one regional population, and spatiotemporal variability in host availability. Discerning whether there are any biological differences between bugs will require resolution of their genetic relationships along with a comprehensive understanding of host use and changing abundance in different regions. Progress has been made in this regard for Australian *N. viridula*, but without similar research in other regions a broader understanding of these relationships will be lacking.

3.5 Supporting information

Table S3.1. Collection information related to the *Nezara viridula* genotyped, including locality, date, and host plants. The site code links the information in this table to raw data files. Samples at all sites were collected by D. Brookes. Entries with multiple host plants represent sites where insects were collected from multiple host plants growing immediately next to one another, and so plants could not be sampled independently of one another. Insects from site codes with an asterisk, totaling 314 individuals, were genotyped previously and the results included in analyses presented in Chapter 2. Insects from all other site codes, totaling 240 individuals, were genotyped for the aims of Chapter 3. The column 'n' shows the number of individuals genotyped at the microsatellite loci as described in Appendices 2.2 and 2.3.

Site code	**Pop. name**	**Locality**	**Long.**	**Lat.**	**Date (dd-mm-yy)**	**Host plant/s**	**n**
EM01*	EMR-Y1a	Emerald	-23.553	148.233	15/*xi*/2014	*Vigna radiata* (mung bean), *Arachis hypogaea* (peanuts)	18
EM02*	EMR-Y1a	Emerald	-23.578	148.178	15/*xi*/2014	*Medicago sativa* (lucerne)	1
EM03*	EMR-Y1b	Emerald	-23.518	148.203	05/*ii*/2015	*Cajanus cajun* (pigeon pea)	20
EM04*	EMR-Y1b	Emerald	-23.525	148.216	04/*ii*/2015	*Arachis hypogaea* (peanuts)	20
EM05	EMR-Y2	Emerald	-23.518	148.203	30/*iii*/2016	*Cajanus cajun* (pigeon pea)	32
BI01*	BIL-Y1	Biloela	-24.381	150.496	06/*ii*/2015	*Gossypium hirsutum* (cotton)	16
BI02*	BIL-Y1	Biloela	-24.379	150.494	06/*ii*/2015	*Cajanus cajun* (pigeon pea)	16
BI04	BIL-Y2	Biloela	-24.432	150.539	23/*xi*/2015	*Medicago sativa* (lucerne)	23
DA01*	DAL-Y1	Dalby	-27.283	151.275	14/*iv*/2015	*Cajanus cajun* (pigeon pea)	16
DA02*	DAL-Y1	Dalby	-27.283	151.275	14/*iv*/2015	*Gossypium hirsutum* (cotton)	16
DA03	DAL-Y2	Dalby	-27.687	151.307	02/*ix*/2015	*Medicago polymorpha*, *Raphanus raphanistrum*, *Vicia sativa* subsp. *sativa*	20
DA04	DAL-Y2	Dalby	-27.367	151.244	01/*ix*/2015	*Rapistrum rugosum* (turnipweed), *Xanthium* sp. (Noogoora burr)	12
DA05*	GAT-Y2a	Gatton	-27.541	152.337	23/*viii*/2015	*Amaranthus* sp., *Brassica* sp., *Malva parviflora*, *Medicago polymorpha*, *Medicago sativa*,	24
DA06	GAT-Y2a	Gatton	-27.248	151.282	01/*ix*/2015	*Rapistrum rugosum* (turnipweed)	3
DA07	GAT-Y2a	Gatton	-27.799	152.106	11/*viii*/2015	*Rapistrum rugosum* (turnipweed)	4
GA02	GAT-Y2b	Clifton	-27.960	152.080	01/*iii*/2016	*Datura stramonium* (Jimson weed)	34
NA01*	NAR-Y1a	Narrabri	-30.186	149.473	04/*xii*/2014	*Vigna radiata* (mung bean)	24

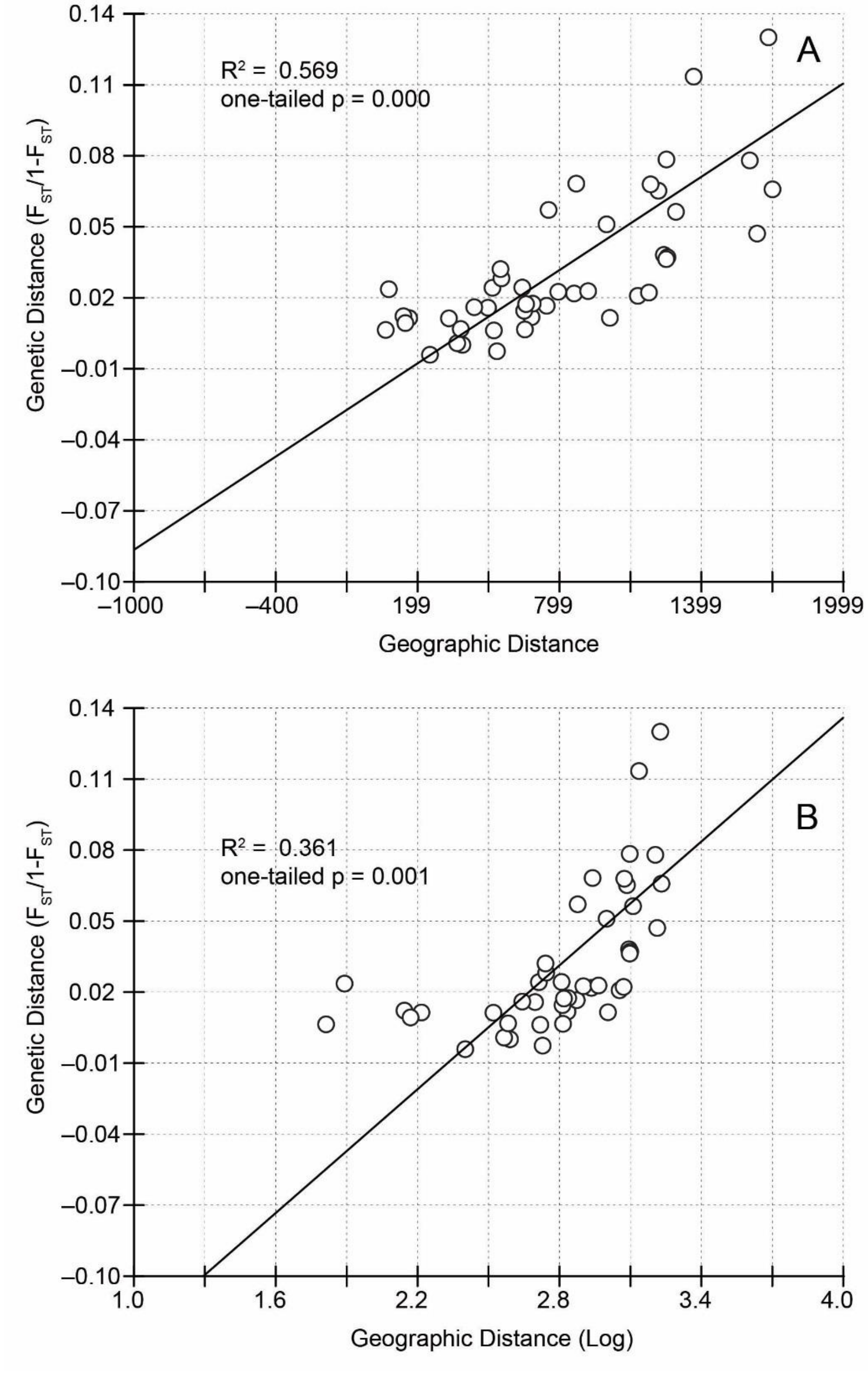
Continued

Table S3.1 continued

NA02*	NAR-Y1a	Narrabri	-30.305	149.657	03/*xii*/2014	*Helianthus annuus* (sunflower)	1
NA03*	BBA-Y1	Narrabri	-30.542	150.010	04/*xii*/2014	*Medicago sativa* (lucerne)	24
NA04	NAR-Y1b	Narrabri	-30.186	149.473	19/*ii*/2015	*Gossypium hirsutum* (cotton)	24
NA05	NAR-Y2	Narrabri	-30.204	149.607	01/*iii*/2016	*Cajanus cajun* (pigeon pea)	32
NA06	BBA-Y2	Narrabri	-30.605	150.025	03/*iii*/2016	*Cajanus cajun* (pigeon pea)	30
GR01*	GRI-Y1	Griffith	-34.419	146.363	24/*ii*/2015	*Cajanus cajun* (pigeon pea)	23
GR02*	GRI-Y1	Griffith	-34.419	146.363	24/*ii*/2015	*Glycine max* (soybean)	8
GR04*	GRI-Y1	Griffith	-34.441	146.037	25/*ii*/2015	*Medicago sativa* (lucerne)	25
DP02*	GRI-Y1	Griffith	-34.596	145.977	25/*ii*/2015	*Cajanus cajun* (pigeon pea)	15
DP03	GRI-Y2	Griffith	-34.610	145.947	06/*iii*/2016	*Cajanus cajun* (pigeon pea)	13
HA01*	HAY-Y1	Hay	-34.472	144.753	25/*ii*/2015	*Gossypium hirsutum* (cotton)	2
HA02*	HAY-Y1	Hay	-34.472	144.753	25/*ii*/2015	*Cajanus cajun* (pigeon pea)	21
HA03	HAY-Y2	Hay	-34.472	144.753	06/*iii*/2016	*Cajanus cajun* (pigeon pea)	37
TV01	TOW-Y1	Townsville	-19.285	146.822	29/*ix*/2014	*Solanum lycopersicum* (tomato)	24
BO01	BOW-Y1	Bowen	-20.078	148.151	31/*ix*/2014	*Solanum nigrum*	15
BO02	BOW-Y1	Bowen	-20.106	148.135	31/*ix*/2014	*Ricinus communis* (castor oil plant)	7
BO03	BOW-Y1	Bowen	-20.106	148.135	31/*ix*/2014	*Cleome gynandra* (tickweed)	5
BO04	BOW-Y1	Bowen	-20.010	148.189	01/*x*/2014	*Gossypium hirsutum* (cotton)	33
BR01	BRZ-Y1	Breeza	-31.158	150.396	23/*vi*/2014	*Rapistrum rugosum* (turnipweed), *Xanthium* sp. (Noogoora burr)	10
BR02	BRZ-Y1	Breeza	-31.222	150.468	05/*xii*/2014	*Xanthium* sp. (Noogoora burr)	10
BR03	BRZ-Y1	Breeza	-31.158	150.396	23/*vi*/2014	*Urtica* sp. (stinging nettle)	2

Table 3.2. Pairwise exact G-test results for genic differentiation across temporally separated populations of *Nezara viridula* collected in the 2014-2015 and 2015-26 seasons (Table 3.2), with lines separating the regions. Non-significant pairwise associations are highlighted green. Population GR-Y2 is excluded for having too few sampled individuals.

	TOW-Y1	BOW-Y1	EMR-Y1a	EMR-Y1b	EMR-Y2	BIL-Y1	BIL-Y2	DAL-Y1	DAL-Y2	GAT-Y2a	GAT-Y2b	NAR-Y1a	NAR-Y1b	NAR-Y2	BBA-Y1	BBA-Y2	BRZ-Y1	GRI-Y1	HAY-Y1	HAY-Y2
TOW-Y1	-																			
BOW-Y1	0.03	-																		
EMR-Y1a	0.13	0.63	-																	
EMR-Y1b	0.00	0.00	0.19	-																
EMR-Y2	0.00	0.00	0.05	0.00	-															
BIL-Y1	0.02	0.61	0.98	0.05	0.01	-														
BIL-Y2	0.24	0.10	0.72	0.31	0.00	0.78	-													
DAL-Y1	0.00	0.00	0.08	0.00	0.00	0.04	0.01	-												
DAL-Y2	0.00	0.00	0.09	0.00	0.04	0.01	0.00	0.37	-											
GAT-Y2a	0.00	0.00	0.01	0.00	0.01	0.03	0.00	0.44	0.30	-										
GAT-Y2b	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.29	0.05	0.39	-									
NAR-Y1a	0.00	0.00	0.06	0.00	0.07	0.03	0.04	0.60	0.57	0.44	0.23	-								
NAR-Y1b	0.00	0.00	0.02	0.00	0.00	0.04	0.02	0.11	0.03	0.03	0.02	0.25	-							
NAR-Y2	0.00	0.00	0.06	0.00	0.08	0.02	0.02	0.02	0.10	0.22	0.86	0.07	0.03	-						
BBA-Y1	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.04	0.16	0.00	0.00	0.06	0.02	0.01	-					
BBA-Y2	0.00	0.00	0.02	0.00	0.05	0.00	0.00	0.05	0.02	0.03	0.56	0.02	0.00	0.60	0.00	-				
BRZ-Y1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.02	0.02	0.01	0.01	0.02	-			
GRI-Y1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	-		
HAY-Y1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	-	
HAY-Y2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.01	0.00	0.00	0.00	0.04	0.00	0.00	0.00	-



FigureS3.1. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* collected in late 2014 and early 2015 (see Table 3.1) , and from northern Queensland in 2014 (see Table 3.1). Geographic distance is used in Figure S3.1A and log transformed geographic distance is used in Figure S3.1B.

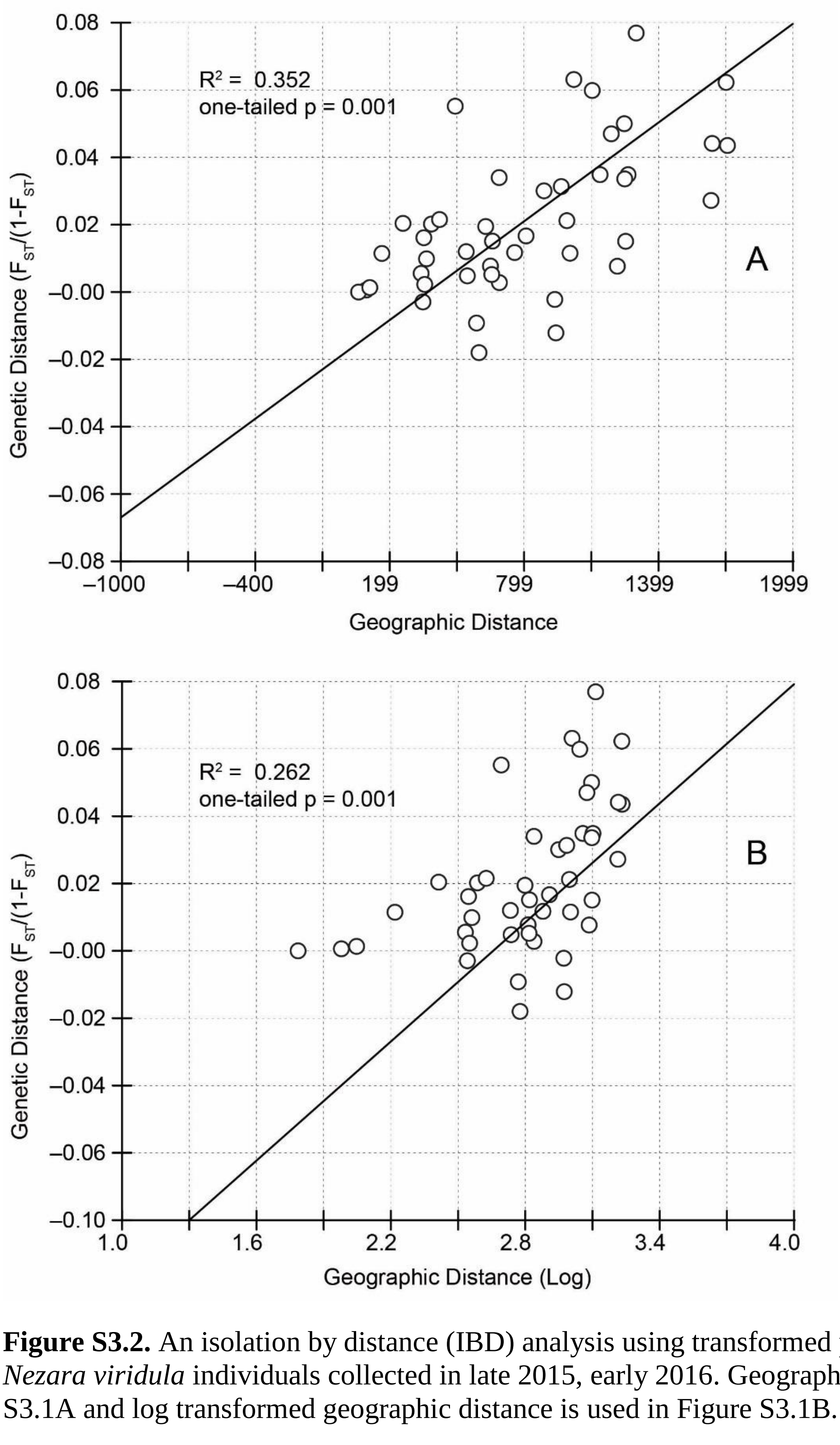


Figure S3.2. An isolation by distance (IBD) analysis using transformed pairwise population F_{ST}s of *Nezara viridula* individuals collected in late 2015, early 2016. Geographic distance is used in Figure S3.1A and log transformed geographic distance is used in Figure S3.1B.

Chapter Four – Evolutionary dynamics of a cycad obligate pollination mutualism – Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators

4.1 Introduction

Obligate pollination mutualisms are among the most intricate of insect-plant interactions simply because each partner species relies entirely on the other for its own persistence. Such tight associations present a unique opportunity to garner insight into the origins and evolutionary dynamics of mutualistic interactions. Early research into obligate pollination mutualisms suggested a ratio of one host species to one pollinator species is most common (Pellmyr 2003; Wiebes 1979), and evolutionary interpretation of these interactions was generally based on the assumption of a dominant role for coevolution (Althoff *et al.* 2014; Anstett *et al.* 1997; Ehrlich and Raven 1964; Hembry *et al.* 2014; Herre, 1989).

Recent research has shown that a number of processes have played a significant role in the diversification of obligate pollination mutualism, including host switching (Cruaud *et al.* 2012; Hembry *et al.* 2013; Jousselin *et al.* 2008; Kawakita and Kato 2009; Machado *et al.* 2005; Smith *et al.* 2008) and co-diversification in allopatry (Althoff *et al.* 2012; Smith *et al.* 2008). Also, these host-pollinator relationships are often less specific than early research suggested (Althoff *et al.* 2012; Cook and Segar 2010; Haine *et al.* 2006; Kawakita and Kato 2006; Machado *et al.* 2005; Molbo *et al.* 2002) and in some cases non-pollinating sister species of pollinating species have been shown to have similar patterns of host association as pollinating species (Althoff *et al.* 2012; Jousselin *et al.* 2008).

Understanding the evolutionary dynamics of obligate pollination mutualisms requires detailed knowledge of the nature of the interaction, the patterns of insect-host species association, and the phylogenetic history of the interacting lineages. This information provides the starting point for evaluating the evolutionary and ecological processes that may have led to the observed associations among lineages in each such mutualistic system. Here we explore the phylogenetic history and evolutionary dynamics of an obligate pollination mutualism between a group of endemic Australian cycads, in the genus *Macrozamia* (Zamiaceae, Cycadales), and their obligate thrips pollinators, *Cycadothrips* (Thysanoptera, Aeolothripidae) (Terry *et al.* 2005), of which three species have been described (Mound and Terry 2001). Both thrips and cycads have long evolutionary histories relative to other insects and seed plants respectively (Grimaldi *et al.* 2004; Norstog and Nicholls 1997; Pott *et al.* 2010). However, the extant species in the genus *Macrozamia* evidently diversified only about five to six Mya (Crisp and Cook 2011; Nagalingum *et al.* 2011; Salas-Leiva *et al.* 2013).

A good understanding of the mechanistic details of the *Macrozamia-Cycadothrips* mutualism has been developed (Terry 2001; Terry *et al.* 2007; Terry *et al.* 2014). The thrips feed on pollen, mate, lay eggs and develop solely within the pollen cones of the dioecious *Macrozamia*, and those within ovulate cones cause no ovulate tissue damage (Mound and Terry 2001; Terry 2001; Terry *et al.* 2005). Studies on *M. lucida* show that cones self-heat for several hours daily, up to 12°C above ambient, simultaneously releasing volatiles in much greater amounts than usual, up to a one million-fold increase for β-myrcene (Roemer *et al.* 2008; Terry *et al.* 2004; Terry *et al.* 2007). The high temperatures and high concentrations of β-myrcene 'push' pollen-laden thrips from the pollen cones (Terry *et al.* 2004; Terry *et al.* 2007; Terry *et al.* 2014) and pollination is achieved when some of these thrips enter ovulate cones (Terry *et al.* 2007). Ovulate cones attract thrips by deception, releasing similar volatiles to the attractive volatiles associated with pollen cones (Terry *et al.* 2004; Terry *et al.* 2007). The thrips are thought to spend the remainder of the year as pupae in soil beneath cycads (Forster *et al.* 1994) although this has yet to be verified.

The central role that coning characteristics play in pollination suggests that these thrips species should be tightly associated with specific cycads or, perhaps, with groups of cycads that share coning characteristics, as occurs in other cycad systems (Suinyuy *et al.* 2012, 2013). One *Macrozamia* pollinator, *C. chadwicki*, pollinates 10 *Macrozamia* species (Forster *et al.* 1994; Jones *et al.* 2001; Mound 1991; Terry *et al.* 2008) that are distributed mostly independently of one another in different environments across a broad area of eastern Australia (Figure 4.1). Some variation is evident between *C. chadwicki* populations, specifically in colour and male size (Forster *et al.* 1994). The *Macrozamia* species with which *C. chadwicki* is associated are diverse in their pollination biology, specifically in the daily onset of cone thermogenesis, volatile profiles, and specific pollinator associations (three of them are also pollinated by specialist weevils) (Jones *et al.* 2001; Roemer *et al.* 2008;

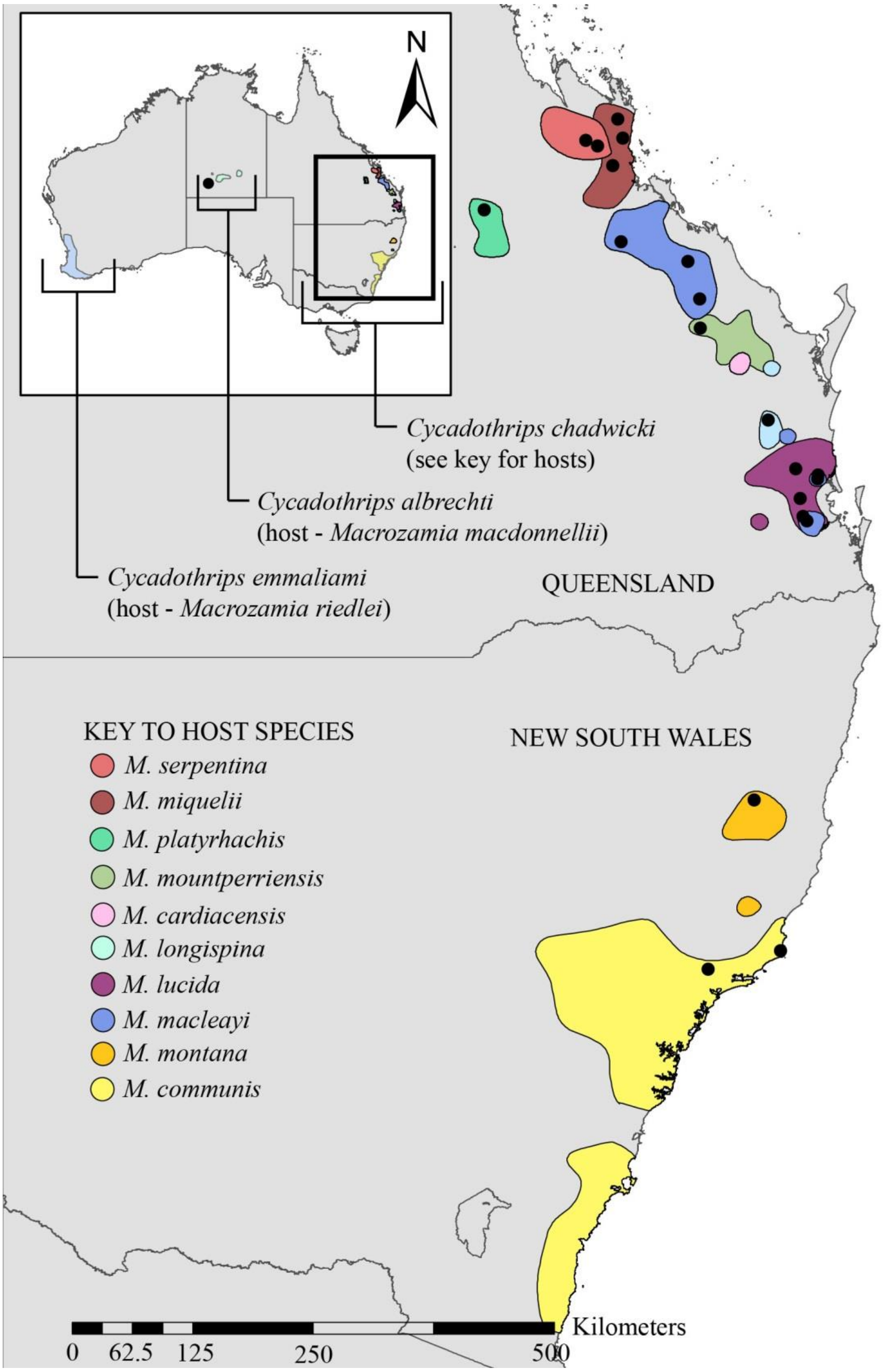


Figure 4.1. The distribution of all 13 species of *Macrozamia* cycad that are pollinated by *Cycadothrips* is represented in these maps of Australia (see key and inset), developed from records obtained through the Atlas of Living Australia (www.ala.org.au) and Queensland Herbarium (Brisbane). Populations of these cycads are patchy, so do not extend over the entire geographical range depicted for each species. The distributions of *M. macleayi* and *M. lucida* overlap in some areas so their population boundaries are not as discrete as depicted. Sites at which *C. chadwicki* were collected for genetic analysis are indicated by black dots. Note that *C. emmaliami* and *C. cardiacensis* were not sampled.

Terry 2001; Terry *et al.* 2008; Terry *et al.* 2004; Terry *et al.* 2005; Wallenius *et al.* 2012), although detailed information is lacking for many species. The diversity exhibited by the hosts suggests more host specialisation within *C. chadwicki* populations than the current taxonomy of these organisms implies.

In this study we provide a phylogenetic and population genetic overview of *Cycadothrips* relative to the *Macrozamia* species they pollinate, and we explore whether *C. chadwicki* represents a number of host-associated cryptic species. We also estimate the divergence times of *Cycadothrips* from different *Macrozamia* species and compare them to the estimated divergence times for *Macrozamia*. We thus evaluate the processes that may have been important in the diversification of this system.

4.2 Methods and materials

4.2.1 Sampling

Cycadothrips chadwicki were sampled from 21 populations and from nine of the 10 *Macrozamia* species they are known to pollinate (Figure 4.1) (Table S4.1). The tenth species, *M. cardiacensis*, is extremely restricted geographically and none of the plants located had produced cones when visited. *Cycadothrips albrechti* was sampled from *M. macdonnellii* but no samples of *C. emmaliami* could be located from *M. riedlei*. About 100 thrips were removed from each cone using a fine brush and placed in 100% ethanol. For DNA extraction, each individual thrips was placed in a separate centrifuge tube, residual ethanol was evaporated overnight, then 500µl of 20% Chelex (100 Resin - BioRad) in TE (10mM Tris HCl and 1mM EDTA (pH8.0)) with RNase (1µl of 25mg/ml RNase per 100ml of 20% Chelex) was added to each centrifuge tube and heated to 97°C for 30 mins on a heating block. This DNA extraction method preserves the whole insect body, and so is available for morphological analysis. Voucher material from each clade that did not share any COI haplotypes with any other clade was deposited in the Queensland Department of Agriculture, Fisheries and Forestry Insect Collection (QDPC), with accession numbers 0-173629 to 0-173667. Morphological analysis of the samples was beyond the scope of this study.

4.2.2 Phylogenetics and Timing of Divergence

Fragments of one mitochondrial gene (Cytochrome C Oxidase 1, COI) and one nuclear gene (28S ribosomal RNA) were PCR amplified across populations of *C. chadwicki* (Figure 4.1) (Table S4.1) and a single population of *C. albrechti*. PCR conditions and primers are outlined in Chapter Appendix S4.1. Phylogenetic analyses were conducted using these two genes. Sequences were edited with CodonCode Aligner 3.7.1 (CodonCode Corporation). Substitution saturation was tested with DAMBE 5.2.78 (Xia 2013) with 2500 replicates. TCS 1.2.1 was used with default settings to create haplotype networks using pairwise parsimony (Clement *et al.* 2000). A second TCS analysis was conducted with the connection limit set to 60 to determine the number of nodes between the three most divergent regions.

Estimations of divergence time were determined using BEAST v1.8.0 (Drummond *et al.* 2012) using individuals with a unique COI haplotype and a corresponding 28S haplotype. The only possible calibration points for *C. chadwicki* are estimates of *Macrozamia* divergence times (Crisp and Cook 2011). The asynchrony between host and pollinator in other similar systems (Smith *et al.* 2008) suggests that calibrating the age of the pollinator to that of the host be discounted. We know of no general substitution rate estimate for the COI gene in thrips. In any case, a generic rate for thrips would probably be inaccurate for *Cycadothrips* because of their unusual life history, specifically the nine to 10 months they are likely to diapause annually. The BEAST analysis was therefore run with a uniform distribution on the uncorrelated lognormal relaxed clock mean prior (ucld.mean), one that was representative of the 95% interval of the COI substitution rate. The mean was set to 0.0145 per site per My (2.8% divergence) and the 95% interval for the upper and lower limit of the substitution rates was 0.0112 per site per My (2.24% divergence) and 0.0178 per site per My (3.56% divergence) respectively. These substitution rates were chosen based on those of other insects, compiled by Papadopoulou *et al.* (2010). Three separate analyses, using the upper, lower and mean substitution rates individually, were also performed. Codons for the COI gene were placed into two partitions with the first and second codon positions in one partition and the third codon position in a second. The substitution rate of the 28S gene was estimated within BEAST.

JModelTest 2.1.4 (Darriba *et al.* 2012; Guindon and Gascuel 2003) was used to compare evolutionary models for COI using maximum likelihood (ML). A HKY model (Hasegawa *et al.* 1985) was selected for the 28S gene to compensate, through its simplicity, for the small number of substitutions (only five substitutions and one insertion within *C. chadwicki*). Analyses were run using a lognormal relaxed clock (uncorrelated) and a birth-death process tree prior. The birth-death model was chosen because the high extinction rate for cycads (Nagalingum *et al.* 2011) suggests that species extinction has probably played a significant role in the evolutionary history of *Cycadothrips*, and this model accounts for that (Nee *et al.* 1994). BEAST was run with a chain length of 50 million, logging parameters every 1000 iterations. Tracer 1.6 (Rambaut *et al.* 2014) was used to assess effective sample size (ESS) of the BEAST analyses. TreeAnnotator 1.8 (part of the BEAST software package) was used, with a burn in of 1000 trees to select a maximum clade credibility tree from among the posterior BEAST trees and to calculate median node heights. Node bars are shown where the posterior probability rounds to 90% in the analysis that used the substitution rate range.

4.2.3 Microsatellite Genotyping and Analysis of Gene Flow

Microsatellites were developed from *C. chadwicki* collected from *M. lucida* and then amplified as per Brookes and Hereward (2013). For this analysis we used the loci CHAD01 through to CHAD10 (Table 4.1). The extensive genetic differentiation we uncovered across *C. chadwicki* samples from the different *Macrozamia* populations meant that the microsatellite analysis could be applied to only eight thrips populations, those from the central clade (Figure 4.2, namely thrips from *M. lucida*, *M. longispina*, *M. mountperriensis* and *M. macleayi* (but not thrips from the northern populations of *M. macleayi*, labelled A in Figure 4.2)). For the thrips collected from the other *Macrozamia* species there was poor or no amplification when the microsatellite primers were initially tested. The only thrips not tested in this way were those from *M. platyrhachis* because too few individuals were available for microsatellite analysis. Genotyping was restricted to females, as the microsatellite optimization process confirmed that *C. chadwicki* is haplo-diploid (Brookes and Hereward, 2013). A total of 287 individuals were genotyped from eight thrips populations collected from four *Macrozamia* species.

Microsatellite peaks were confirmed and binned manually using GeneMarker 2.2.0 and the Geneious 6.1.4 microsatellite plugin. Fisher's exact tests for linkage disequilibrium and for deviation from Hardy-Weinberg equilibrium (HWE) were performed in Genepop 4.0.10 (Raymond and Rousset 1995; Rousset 2008). The presence of null alleles was estimated and Global F_{ST}s were calculated in FreeNA (Chapuis and Estoup 2007). Two loci, CHAD08 and CHAD10, showed evidence for null alleles (each in a different population of *C. chadwicki*) (Table 4.1) and were excluded from subsequent analyses. An individual based Bayesian clustering algorithm, STRUCTURE 2.3.3 (Falush *et al.* 2003; Falush *et al.* 2007; Pritchard *et al.*, 2000), was used to investigate population structure. This algorithm uses a combination of HWE and linkage to assign a posterior probability that an individual belongs to a user specified number of clusters (K). STRUCTURE analyses were carried out under the admixture model with alleles correlated and an initial burn-in of 100,000 iterations followed by 1,000,000 iterations with 20 replicates for each value from K 1 to K 10. The most likely value of K was determined with STRUCTURE HARVESTER (Earl and VonHoldt 2012), which is based on the methods of Evanno (2005). A principal coordinates analysis (PCoA) was computed using pairwise individual genetic distance in Genalex 6.501 (Peakall and Smouse 2006; Peakall and Smouse 2012) so that genetic differentiation and clustering could be assessed without the assumption that populations were in HWE. Pairwise

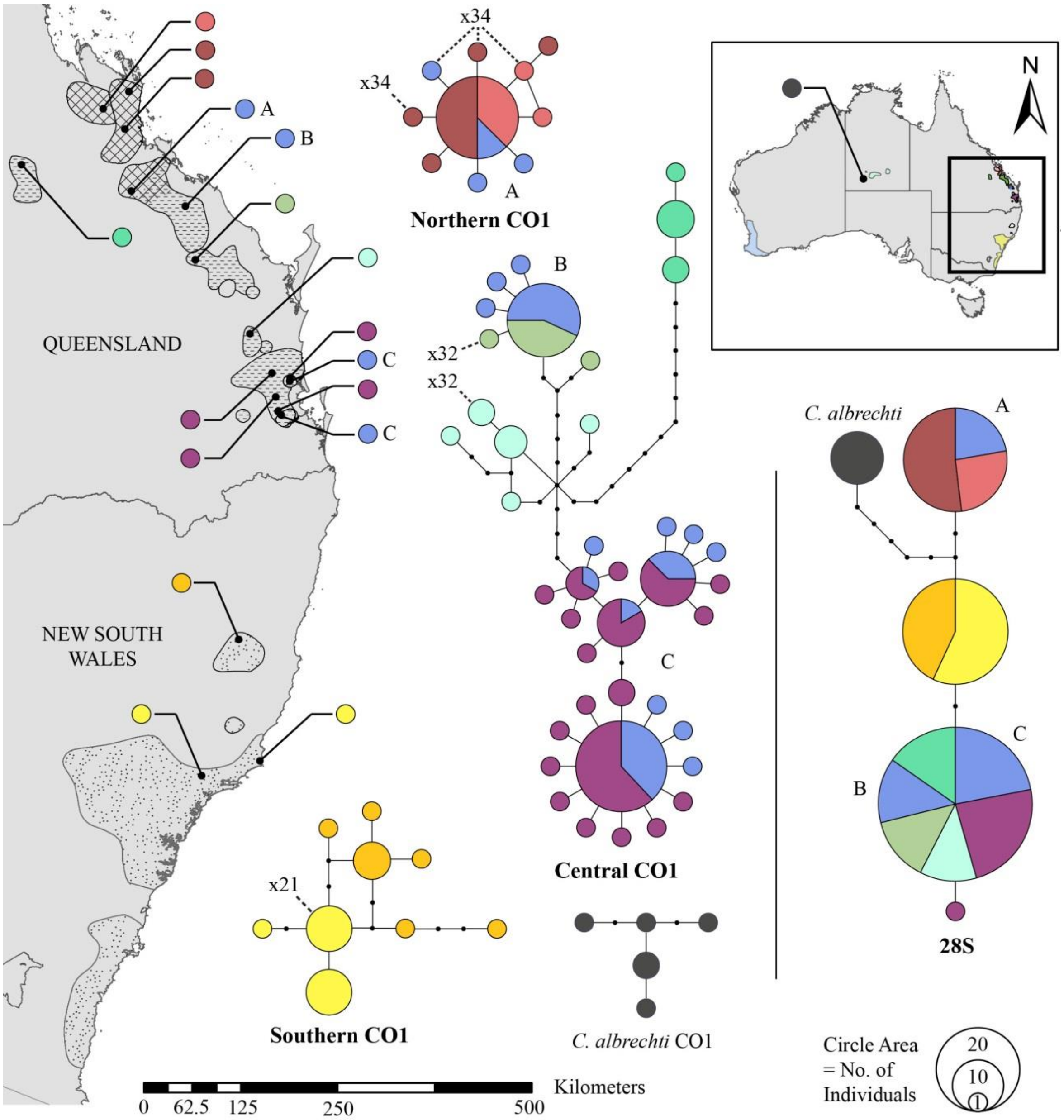


Figure 4.2. Map of eastern Australia to show the association of the 59 COI haplotypes (networks left of vertical line) and four 28S haplotypes (networks right of vertical line) of *Cycadothrips chadwicki* relative to their host *Macrozamia* species. For *C. albrechti*, both the five COI and one 28S haplotypes are shown in dark grey. Each *Macrozamia* species is indicated by a different colour (same as that used in the key to host species and their distribution in Figure 4.1) with respect to the sample locality on the map (see below) and in relation to the thrips collected from them (in the networks). The COI haplotypes of *C. chadwicki* fall into one of three main clades, each of which has its network plotted alongside the map. For both genes, the small black dots represent haplotypes for which samples were not obtained, and the size of each haplotype circle corresponds to the number of individuals with that haplotype (see key in bottom right hand corner of figure). All three *C. chadwicki* networks are joined by a central node, dotted lines with a number represent the number of nodes to this central node. In the map, shading on the distribution of each *Macrozamia* species represents the plants with which the three main thrips clades (north (cross hatch), central (dashed lines) and southern (dotted), are associated. The letters A, B and C associated with different *M. macleayi* populations on the map are cross-referenced to both haplotype networks. For the extent of divergence between these clades also see Figure 4.3.

population F_{ST} values were computed in Genepop 4.0.10 (Raymond and Rousset 1995; Rousset 2008).

Table 4.1. Statistics of the microsatellite loci included in the population genetic analysis of *Cycadothrips chadwicki.*

Locus	$\sum Na^{a}$	$\hat{N}a^{b}$	Ho^{c}	He^{d}	**HWE**^{e}	**No. Null Alleles**	**gF_{ST} Null**	**gF_{ST} No Null**
CHAD01	11	7.5	0.662	0.704	3 (0)	0.040	0.424	0.417
CHAD02	17	9.375	0.701	0.750	1 (1)	0.035	0.224	0.220
CHAD03	13	6.75	0.668	0.657	0 (0)	0.015	0.072	0.071
CHAD04	26	14.875	0.767	0.845	1 (0)	0.030	0.125	0.120
CHAD05	11	4	0.417	0.426	1 (0)	0.015	0.038	0.039
CHAD06	26	14.5	0.926	0.891	1 (0)	0.005	0.075	0.073
CHAD07	24	15.25	0.857	0.893	2 (0)	0.010	0.028	0.027
CHAD08	20	11.75	0.783	0.810	3 (0)	0.017	0.096	0.093
CHAD09	16	9	0.712	0.763	1 (0)	0.041	0.088	0.087
CHAD10	18	9.625	0.717	0.812	3 (1)	0.040	0.057	0.052

[a] Total number of alleles.
[b] Average number of alleles.
[c] Observed heterozygosity.
[d] Expected heterozygosity.
[e] Deviations from Hardy-Weinberg Equilibrium with Bonferroni corrected values in brackets.

4.3 Results

4.3.1 Phylogenetics and Timing of Divergence

Fifty-nine COI haplotypes were recorded across 150 *C. chadwicki* individuals from 17 populations (northern, central and southern in Figure 4.2). By contrast, the slowly evolving 28S sequences had very low diversity, being represented by only four haplotypes across 15 populations and 115 *C. chadwicki* individuals (Figure 4.2). Five unique COI haplotypes and one unique 28S haplotype were recorded for *C. albrechti* from seven and eight individuals respectively. The results from COI and 28S are congruent in showing three highly divergent clades within *C. chadwicki*, namely central, southern and northern (Figure 4.2).

Each of the three divergent *C. chadwicki* clades is, to a large extent, associated with a particular subset of hosts. The 'northern' clade is associated with *M. serpentina*, *M. miquelii* and the northernmost population of *M. macleayi* (A) (Figure 4.2). The 'central' clade is associated with *M. mountperriensis, M. longispina, M. lucida* and the remaining *M. macleayi* populations (B and C)

(Figure 4.2). The 'southern' clade is associated with *M. montana* and *M. communis* (Figure 4.2). Nested within the central clade are three further clades (based on COI sequences) and each is again mostly associated with a particular subset of hosts. One clade contains thrips from *M. macleayi* (B) and *M. mountperriensis,* one contains *M. longispina* thrips and the third contains thrips from *M. macleayi* (C) and *M. lucida*. *Macrozamia macleayi* differs from all the other *Macrozamia* species sampled in that the thrips associated with these plants derive from three different clades, and each of these thrips clades is confined to the geographical distribution of each of the three *M. macleayi* populations.

Divergence times estimated by BEAST are presented as a chronogram from the analysis that used the range of substitution rates that are probable for the COI gene in *Cycadothrips* (Figure 4.3). Substitution models were a GTR +G +I model for the COI gene as determined by JModelTest 2.1.4 (Darriba *et al.* 2012; Guindon and Gascuel 2003) and the HKY model selected for the 28S gene. The *C. albrechti* and *C. chadwicki* lineages are estimated to have diverged about 10.8 Mya (6.0-17.1, 95% HPD). Divergence within the *C. chadwicki* lineage occurred within two different time periods. The older diversification resulted in three clades within *C. chadwicki,* with the northern clade diverging from the central and southern clades about 7.3 Mya (4.4-11.1, 95% HPD) and the central and southern clades diverging from one another about 5.7 Mya (3.2-8.8, 95% HPD). Both of these nodes had low support relative to other nodes, with a posterior probability of 0.91 and 0.90 respectively (Figure 4.3). The remaining nodes are shallow in all clades and there is little support for further division within the northern and southern clades. Of the analyses performed using individual substitution rates, the median substitution rate analysis is most similar to the analysis that used the substitution rate range (Figure S4.1). The analysis that used a range of substitution rates did not reach the same upper and lower divergence estimates as the individual substitution rate analysis. As a result, all three of the single substitution rate analyses are included in the supplementary information to provide a more comprehensive range of possible date estimates (Figures S4.1, S4.2 and S4.3). The divergence estimates from the BEAST analysis that used the range of substitution rates is presented throughout.

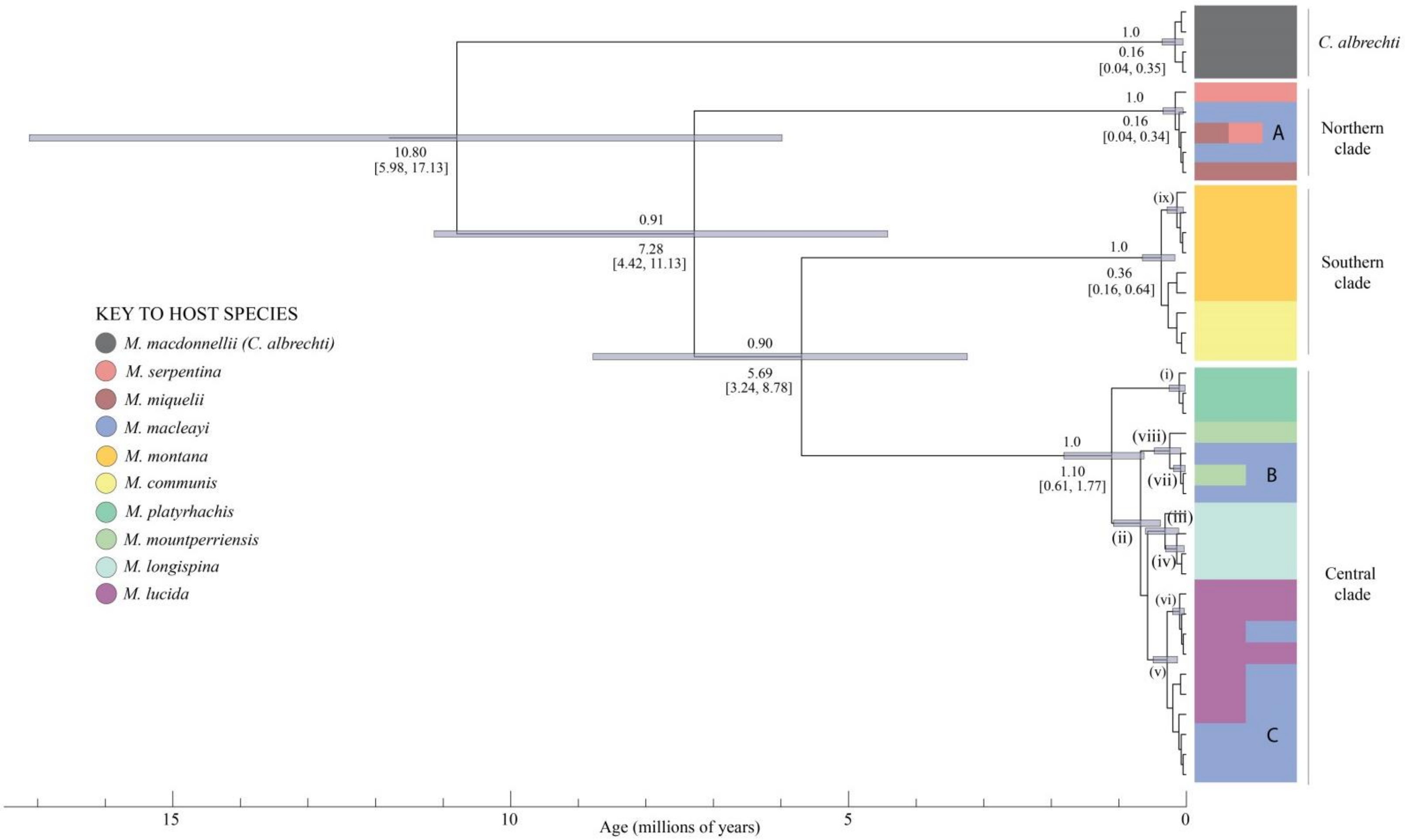


Figure 4.3. Chronogram of *C. chadwicki* samples by host plant. The tree is a maximum clade credibility tree based on COI and 28S data (BEAST 1.8.0) and adapted from FigTree 1.4 outputs. The root is calculated from the uncorrelated molecular clock. Coloured blocks represent the *Macrozamia* species from which each individual was sampled as per the key to host species (and these colours are the same as those used in Figs. 1 and 2). Haplotypes shared across *Macrozamia* species indicate presence and absence so are not coloured proportionately. The letters on the *M. macleayi* blocks correspond to the population from which these thrips were sampled (Figure 4.2). The scale bar represents the number of years (in millions) represented by each branch. The posterior probabilities are displayed as node labels and the 95% Highest Posterior Densities (95% HPD) are shown as branch labels. The values for nodes (i) through (ix) are listed in Table S2.

The central clade is relatively diverse, and its diversification began about 1.1 Mya (0.6-1.8, 95% HPD). At that point the *C. chadwicki* associated with M. *platyrhachi*s separated from all other central clade thrips. Further diversification events within the central clade are not as distinct and separate the lineage (labelled (ii) in Figure 4.3) into a further two or three lineages beginning about 0.7 Mya (0.4-1.1, 95% HPD). Of these lineages, one is comprised of the thrips from *M. lucida* and *M. macleayi* C (labelled (v) in Figure 4.3), another of the thrips from *M. longispina* and the last of thrips from *M. mountperriensis* and *M. macleayi* (B). These shallow divergence events were further assessed by analysing gene flow across central clade thrips populations with microsatellites.

4.3.2 Microsatellite Gene Flow Analyses

The following results refer to thrips populations from the central clade. Two loci, CHAD02 and CHAD10, showed significant deviations from HWE (Table 4.1), each in one thrips population, after Bonferroni correction. The frequency of null alleles was low across all 10 loci, with the exceptions of locus CHAD09, which had an estimated frequency of 0.16 in population *M. macleayi* C2, and locus CHAD10 with an estimated frequency of 0.19 in population *M. mountperriensis* (data not shown). The remaining eight loci used across populations from the ‘central’ clade were highly variable with a total of 148 alleles scored across all loci (mean alleles per locus = 18.5) (Table 4.1).

Clustering analysis in STRUCTURE indicated two levels of hierarchical structuring. Specifically, the Delta K method (Evanno *et al.* 2005) indicated that K=2 is the most likely number of clusters but K=3 also had good support (Figures S4.4 and S4.5), thus revealing two levels of population subdivision with three populations total. The results of a PCoA (Figure S4.6) and pairwise F_{ST} analysis (Table S4.3) also revealed two levels of differentiation in these populations. These analyses detected no gene flow, or extremely low amounts of gene flow, between the thrips making up these three clades. The first hierarchical cluster, K=2, organised the thrips in lineage (ii) (Figure 4.3), namely those from *M. lucida* and *M. macleayi* (C) into one cluster, and those thrips from *M. mountperriensis*, *M. macleayi* (B) and *M. longispina* made up the second cluster (Figure S4.5), this grouping of samples returned a mean pairwise F_{ST} of 0.18. The K=3 hierarchical clustering further separated the thrips from *M. longispina* from those associated with *M. mountperriensis* and *M. macleayi* B (Figure 4.4). Mean pairwise F_{ST}s within the *M. lucida*/*M. macleayi* (C) cluster were low (0.03), between this group and the thrips from *M. longispina* the mean pairwise F_{ST} was 0.18, whereas between *M. longispina* and *M. mountperriensis*/*M. macleayi* B the mean was 0.13. All three analyses thus support the presence of three distinct gene pools in this region.

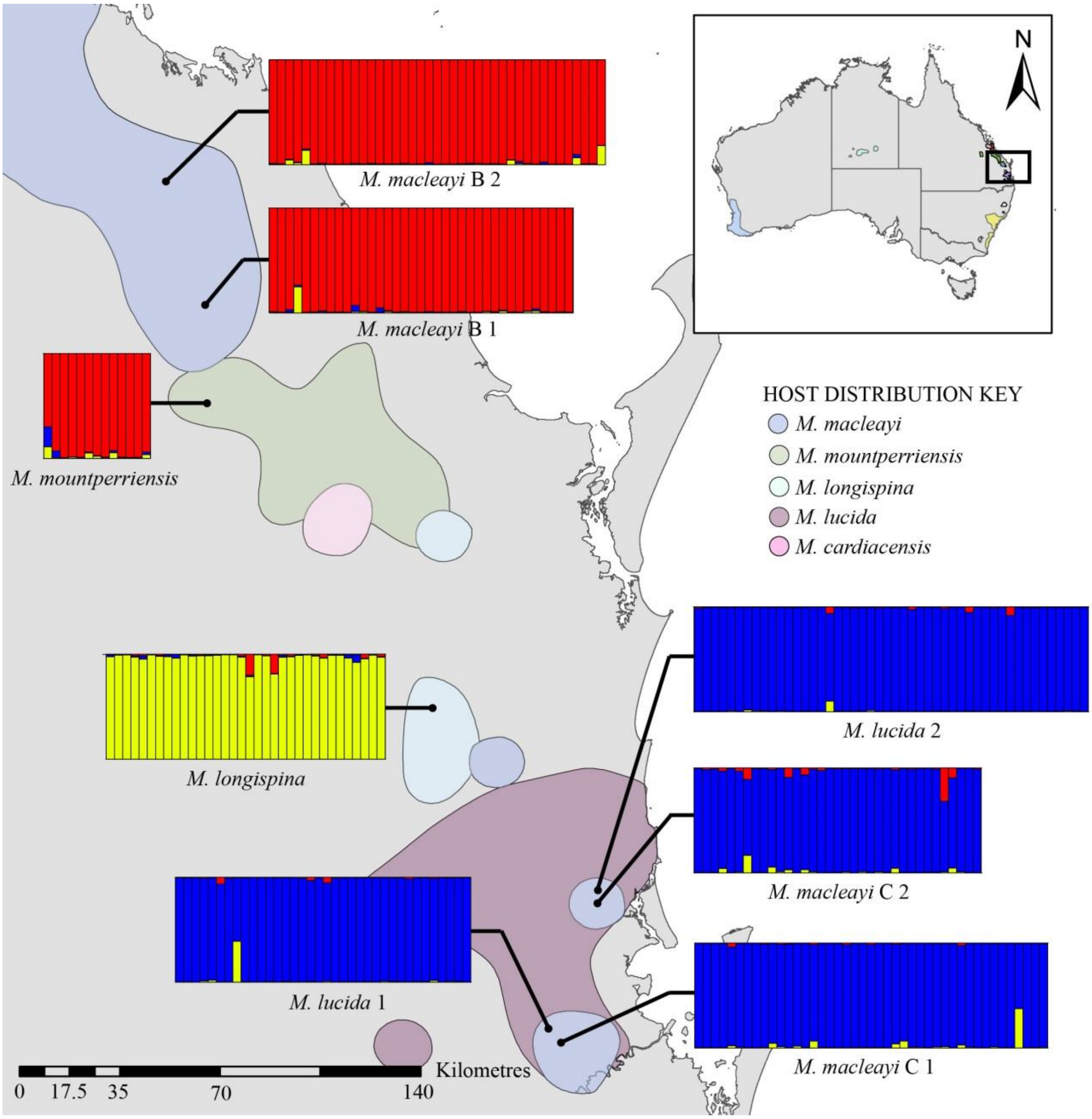


Figure 4.4. Map showing the distribution of the *Macrozamia* populations represented in Figure S4.5. The distribution of each *Macrozamia* species is represented by a different colour (see key on right of figure). The distribution of *M. lucida* overlaps with much of the distribution of *M. macleayi* and hybridization is thought to occur in some locations. Each box shown represents the individuals included in the K3 STRUCTURE analysis and they are presented separately to show the origin of the insects relative to a particular host plant species (Figure S4.5).

4.4 Discussion

4.4.1 Species Diversity and Host Relationships

Analysis of the *Macrozamia-Cycadothrips* mutualism has revealed waves of diversification and a complex pattern of host association within this system. Significantly, the pollinating thrips are far more diverse than previously interpreted, with these results showing at least five host-associated *Cycadothrips* species within *C. chadwicki* (Table 4.2). The species diversity uncovered within *Cycadothrips* also exhibits strong geographical structure and host plant associations. We have been conservative in interpreting these lineages as cryptic species, and have based this on molecular evidence (Table 4.2).

Table 4.2. All known *Cycadothrips* species, including the cryptic species revealed in the current study and the evidence used to designate them as such. The *Cycadothrips* populations not sampled are indicated as such in the left hand column.

***Cycadothrips* species**	**Host species**	**Evidence**
Eastern Australia (northern clade)		
C. chadwicki 1	*M. serpentina, M. miquelii, M. macleayi* (A)	28S, COI (Figure 4.2), lack of microsatellite amplification
Eastern Australia (central clade)		
C. chadwicki 2	*M. platyrhachis*	COI (Figure 4.2)
C. chadwicki 3	*M. longispina, M. mountperriensis, M. macleayi* (B)	COI (Figure 4.2), microsatellite (Figure 4.4)
C. chadwicki 4	*M. lucida, M. macleayi* (C)	COI (Figure 4.2), microsatellite (Figure 4.4)
C. chadwicki (not sampled)	*M. cardiacensis*	-
Eastern Australia (southern clade)		
C. chadwicki 5	*M. communis, M. montana*	28S, COI (Figure 4.2), lack of microsatellite amplification
Northern Territory		
C. albrechti	*M. macdonnellii*	28S, COI (Figure 4.2), morphology[1]
Western Australia		
C. emmaliami (not sampled)	*M. riedlei*	Morphology[2]

[1](Mound and Terry, 2001), [2](Mound and Marullo, 1998)

The deepest level of molecular divergence within *C. chadwicki sensu lato* is evident across the geographically separated northern, southern and central clades. The COI analyses indicate millions of years of each one having been independently associated with different subsets of host species. Further, few of the 18 microsatellites developed from the *M. lucida* thrips population could be PCR amplified in the northern or southern clades, and those that did amplify produced non-specific products (Brookes and Hereward, 2013). Genomic divergence is therefore deep between these three lineages, and the northern clade thrips on *M. serpentina*, *M. miquelii* and M. *macleayi* (A) represent a separate species (*C. chadwicki* species 1) and so do the southern clade thrips (from *M. communis* and *M. montana*) (*C. chadwicki* species 5) (Table 4.2).

With regard to the third lineage, represented by the central clade, the STRUCTURE analysis indicates that the thrips from *M. longispina* and those from *M. lucida*/*M. macleayi* C are separate gene pools from one another. No indication of gene flow between them is evident despite gene flow occurring over similar distances between populations on *M. lucida* and *M. macleayi* C (Figures S4.4 and S4.5). The central clade populations thus represent two separate species as indicated in the K=2 STRUCTURE analysis (Figure S4.5). The population structure evident between *M. longispina* and *M. mountperriensis*/*M. macleayi* B (Figures 4.4 and S4.6) might indicate an additional species division. Support comes from the extent of their genomic divergence, with only 10 loci successfully amplifying for *M. mountperriensis*/*M. macleayi* B, and 13 for *M. longispina*, (Brookes and Hereward 2013), which accords with the mitochondrial results in Figure 4.2. However, a 50km gap in cycad distribution occurs between the populations, so they could simply be geographically isolated from one another. Finally, with respect to the central clade, we have designated the *Cycadothrips* collected from the geographically isolated *M. platyrhachis* as *C. chadwicki* cryptic species 2 (Table 4.2), based on it being the most divergent for COI in the central group. Further tests of species status, perhaps with cross-mating and host association tests, if possible, as well as morphological analyses would be useful.

Each of the five species that make up *C. chadwicki* is found only on one to three *Macrozamia* species, and each host species is associated, in turn, with only a single pollinator species (Table 4.2). The *Cycadothrips* species maintain these host associations where different *Macrozamia* species (with different species in the *C. chadwicki* complex) have adjacent distributions (close enough to expect thrips movement), as is the case of thrips from *M. longispina* and those from *M. lucida* and *M. macleayi* C. These results demonstrate that the cryptic species uncovered in the study under the single name *C. chadwicki* are host specific, though usually to subsets of *Macrozamia* species rather than to single species. The phylogenetic analysis of *Cycadothrips* indicates that this degree of specificity has been generated by diverse processes as discussed in Section 4.4.2.

Despite this specificity there are three instances where a single pollinator species in the *C. chadwicki* complex is associated with more than one *Macrozamia* species. Although this occurs only when the host populations overlap or are geographically close, in all cases the seasonal coning phenology of the host species overlaps (Jones *et al.* 2001; Terry *et al.* 2008; Terry *et al.* 2004), unlike that of some cycad species (Griffith *et al.* 2012). How then do these cycad species persist when they share a pollinator, and how have these host associations arisen? The problem is best illustrated with reference to *M. lucida* and *M. macleayi* (C, Figure 4.2), which occur in sympatry, have similar phenology and volatile profiles, and share a pollinator. Hybrid individuals have been reported in some areas but the populations used in this study remain morphologically distinct (Terry *et al.* 2008). Our analysis shows that the pollinators are a single freely mixing population (*M. lucida* 1 and *M. macleayi* C1; Figure 4.4). Targeted research on these situations is clearly needed to determine the extent to which gene flow occurs between the cycad populations.

The specificity in the host associations of the eastern Australian *Cycadothrips* on *Macrozamia*, as now understood, can perhaps also provide insight on aspects of *Macrozamia* taxonomy, particularly that involving *M. macleayi*. The northern populations of *M. macleayi* (A and B; Figure 4.2) are each associated with one of two *C. chadwicki* species that diverged ~7.3 Mya (4.4-11.1, 95% HPD). Given the pattern of host association in the rest of the mutualism, and the millions of years that separate the two thrips species on *M. macleayi* A and *M. macleayi* B, it seems highly likely that these separate populations of *M. macleayi* represent two distinct species. Each population is only associated with a single thrips, so a taxonomic and genetic assessment of these populations is required to determine whether a host switch or co-diversification is more likely to have occurred. Taxonomic assessment of these cycad populations is currently being conducted.

4.4.2 Timing and Patterns of Diversification

As with *Macrozamia*, the extant *Cycadothrips* appear to have evolved relatively recently and the estimates of their divergence times (10.8 Mya, 6.0-17.1, 95% HPD) overlap with most estimates for the extant *Macrozamia* species (5.9 Mya, 1.0-17.0, 95% HPD (Crisp and Cook 2011); 3.7 Mya, 3.7-11.8, 95% HPD (Nagalingum *et al.* 2011); 1.8 Mya, (Salas-Leiva *et al.* 2013)). This result supports the recent diversification of extant *Macrozamia*. These estimates coincide with the fragmentation of mesic environments as a result of the increasing aridification of the continent in the mid to late Miocene (Byrne *et al.* 2011; Kershaw 1997; Kershaw *et al*, 1994; McGowran *et al.* 2004; Moussalli *et al.* 2005) and the most recent common ancestor for both host and pollinator derive from this period. Within each of the three main lineages of *C. chadwicki*, diversification has occurred no earlier than about 1.1 Mya (0.6-1.8, 95% HPD) and usually much more recently (Figure 4.3) likely due to further fragmentation of their hosts' distributions as a result of increasing climatic variability and seasonality during the Quaternary (Byrne *et al.* 2011; McGowran *et al.* 2004). This correlation with the fragmentation of mesic habitats, and known geographic breaks between the regions (i.e. the Border Ranges (Joseph and Omland 2009; Schodde and Mason 1999) and the Dawson-Mackenzie Gap (Edwards and Melville 2010)), indicate that divergence in allopatry has been important in the evolution of extant *Cycadothrips* and *Macrozamia*.

The similar dates estimated for the diversification of the extant *Macrozamia* and *Cycadothrips* indicate consistent association between them. This provides further evidence for the recent diversification of *Macrozamia*, simply because of the obligate nature of the interaction. Given the relatively recent last common ancestor of *Macrozamia* species compared to the age of the genus, there has probably been a large amount of co-diversification as well as extinction of both thrips and cycads through this period of their evolutionary history. However, the ratio of host species to pollinator species suggests that speciation and extinction do not always occur in a strictly one-to-one fashion. The relatively greater number of *Macrozamia* species suggests that speciation has been more frequent in *Macrozamia* than in *Cycadothrips*, and that extinction and subsequent host switching of the pollinators has been common, or perhaps both sets of processes have prevailed.

Strong evidence supports the view that host switching took place within the older lineages of the mutualism. A single common *Cycadothrips* ancestor diverged into *C. albrechti* and the *C. chadwicki* lineage (10.8 Mya, 6.0-17.1, 95% HPD) (Figure 4.3). This took place millions of years before the most recent common ancestor of their corresponding *Macrozamia* host species had diverged, about 1.1 Mya (*M. macdonnellii* and the eastern *Macrozamia*; 0.2-3.4 Mya, 95% HPD) (Ingham *et al.* 2013). Therefore the thrips representing these two lineages must have switched hosts from a now extinct *Macrozamia*. Host switching may also have occurred within the *C. chadwicki* complex, with the thrips species (species 1 in Table 4.2) that pollinates *M. miquelii*, *M. serpentina* and *M. macleayi* A having undergone a host switch to these species of *Macrozamia*, perhaps from an extinct cycad or from one of the current host species to the others. These *Macrozamia* species are evidently closely related to one another and to the various *Macrozamia* species on which the central clade thrips are found (Jones *et al.* 2001) despite the large genetic divergence between *C. chadwicki* species 1, on the one hand, and species 2, 3, 4 and 5, on the other (see data in Table 4.2).

Determining the relative importance of host switching, host specialization and co-diversification, especially in the mutualism's younger lineages, is hampered by the low levels of genetic differentiation and incomplete lineage sorting between many of the *Macrozamia* species (Ingham *et al.* 2013). Clearly, further research into the pollination biology of each host-pollinator group and a full comparison of the *Cycadothrips* phylogeny and a *Macrozamia* phylogeny is required, and the following questions require particular attention. (1) What are the evolutionary relationships among *Macrozamia* species and what does this say about the relative role of host switching? (2) What are the similarities and differences in coning characteristics (including volatile profiles) across these *Macrozamia* species? (3) How do the patterns of *Cycadothrips* host association compare with these coning characteristics of *Macrozamia*? (4) What are the spatiotemporal dynamics of gene flow in *Cycadothrips* and *Macrozamia* relative to one another? Resolution of these points will help reveal the relative role of the processes that have shaped the evolution of this mutualism. In short, more biological detail is required across all species combinations of *Macrozamia* and *Cycadothrips,* but our results set the direction for this research, as outlined above.

4.4.3 General Consequences for Obligate Pollination Mutualisms

Co-diversification in allopatry and host switching both appear to have played a role in generating the patterns of species association recorded in the *Macrozamia-Cycadothrips* mutualism, as detailed above, and these processes have been shown to be important in other obligate pollination mutualisms (Althoff *et al.* 2012; Hembry *et al.* 2013; Machado *et al.* 2005). The fewer pollinator species than host plant species suggests that strict coevolution has not played a primary role in generating the extant species associations. If coevolution does not structure the extant associations, even in the highly specialized systems mentioned above, its general importance as an evolutionary process is called into question.

The evolutionary history of cycad mutualisms is potentially very long, and probably precedes the origin of angiosperms. These interactions have endured despite their specificity and their obligate nature. The data presented here, and elsewhere in the literature, indicate that these species associations can, nevertheless, change through evolutionary time. We cannot be certain which insects pollinated cycads through their entire evolutionary history (Labandeira *et al.* 2007), but pollination mutualisms typify all extant cycad genera, and usually involve weevils (Downie *et al.* 2008; Hall *et al.* 2004; Norstog *et al.* 1986; Schneider *et al.* 2002; Wilson 2002). That different beetle taxa pollinate cycads across the different continents that have cycads, from ancestral forms of beetles that were originally associated with angiosperms (Anderson 1993; Marvaldi *et al.* 2006; Oberprieler *et al.* 2007; Stevenson *et al.* 1998), reveals that new associations have evolved periodically, but it is not clear what drives such adjustments. These highly specific systems of mutual reliance have, therefore, apparently persisted over millions of years by being able to incorporate new pollinators under some (as yet unknown) circumstances.

4.5 Conclusions

We analysed the obligate mutualistic pollinators of a subset of endemic Australian *Macrozamia* cycads, phylogenetically and with population genetics data, and demonstrated a pattern of recent species diversification within these *Cycadothrips* pollinators, at ~10.8 Mya for *C. chadwicki* and *C. albrechti* and ~7.3 Mya for the three main *C. chadwicki* clades. The species within each of the three *C. chadwicki* clades originated at most ~1.1 Mya, and as recently as ~0.15 Mya. Relating the phylogeny of these insects to their host species associations revealed that host switching alone accounts for the host association of the oldest and most divergent *Cycadothrips* species, because of the older estimate for the *Cycadothrips* divergence (~10.8 Mya) relative to that of their hosts, *M. macdonnellii* and the eastern *Macrozamia*; (~1.1 Mya (Ingham *et al.* 2013)). This means that pre-adaptation must have been important in *Cycadothrips* diversification at this particular level and, correspondingly, coevolution could not have played a significant role. In the younger *Cycadothrips* lineages co-diversification with their hosts in allopatry appears to have been the dominant process, which again suggests that coevolution was not important. The *Cycadothrips* phylogeny provides the basis from which a comparative study of the structure and operation of this mutualism can be designed, to help determine which aspects have changed and how the extant mutualistic associations arrived at their specific structure. Mapping these functional traits onto more comprehensive phylogenies of the plants and the pollinator phylogeny presented here would allow further inference into the processes responsible for the diversification of these tightly associated mutualistic organisms.

4.6 Supporting information

Chapter Appendix S4.6.1 PCR conditions and primer information for the amplification of the 28S and COI genes from *Cycadothrips chadwicki* and *C. albrechti*.

For 28S, PCR conditions for both *C. chadwicki* and *C. albrechti* were 1x Buffer, 3.0mM $MgCl_2$, 1.0µM dNTPs, 0.4µM of each primer – A335 (Whiting *et al.* 1997) and S3660 (Dowton and Austin 1998), 1 unit of myTAQ DNA polymerase and 2.0µl of DNA template. Cycling conditions were 10min at 95°C followed by 35 cycles of 25s at 95°C, 45s at 54°C and 45s at 72°C followed by extension of 10min at 72°C. PCR product was cleaned using Exo-Ap (adapted from Werle (1994), see below) at 2µl per sample and 15min at 37°C followed by 15min at 80°C. PCR product was sequenced bi-directionally on an ABI3730 capillary sequencer (Macrogen Inc., Seoul, Korea). Exo-Ap used 5µl Exonuclease I (20U/µl), 10µl buffer and 85µl H_2O combined with 20µl Antarctic Phosphatase (5U/µl), 10µl buffer and 70µl H_2O.

PCR Conditions for *C. chadwicki* COI were 1x Buffer, 2.5mM $MgCl_2$, 0.2µM dNTPs, 0.4µM of each primer (HCO2198 and LCOI460 (Folmer *et al.* 1994)(Folmer *et al.* 1994)), 1.25 units of mangoTAQ (0.375 units of PlatinumTAQ for *C. albrechti*) DNA polymerase and 3.6µl of DNA template with a total reaction volume of 30µl. Cycling conditions were 10min at 95°C followed by 5 cycles of 95°C for 45s, 45s at 45°C and 1min at 72°C. Following this was a second series of 30 cycles of 95°C for 45s, 45s at 60°C and 1min at 72°C, ending with a final extension of 5min at 72°C.

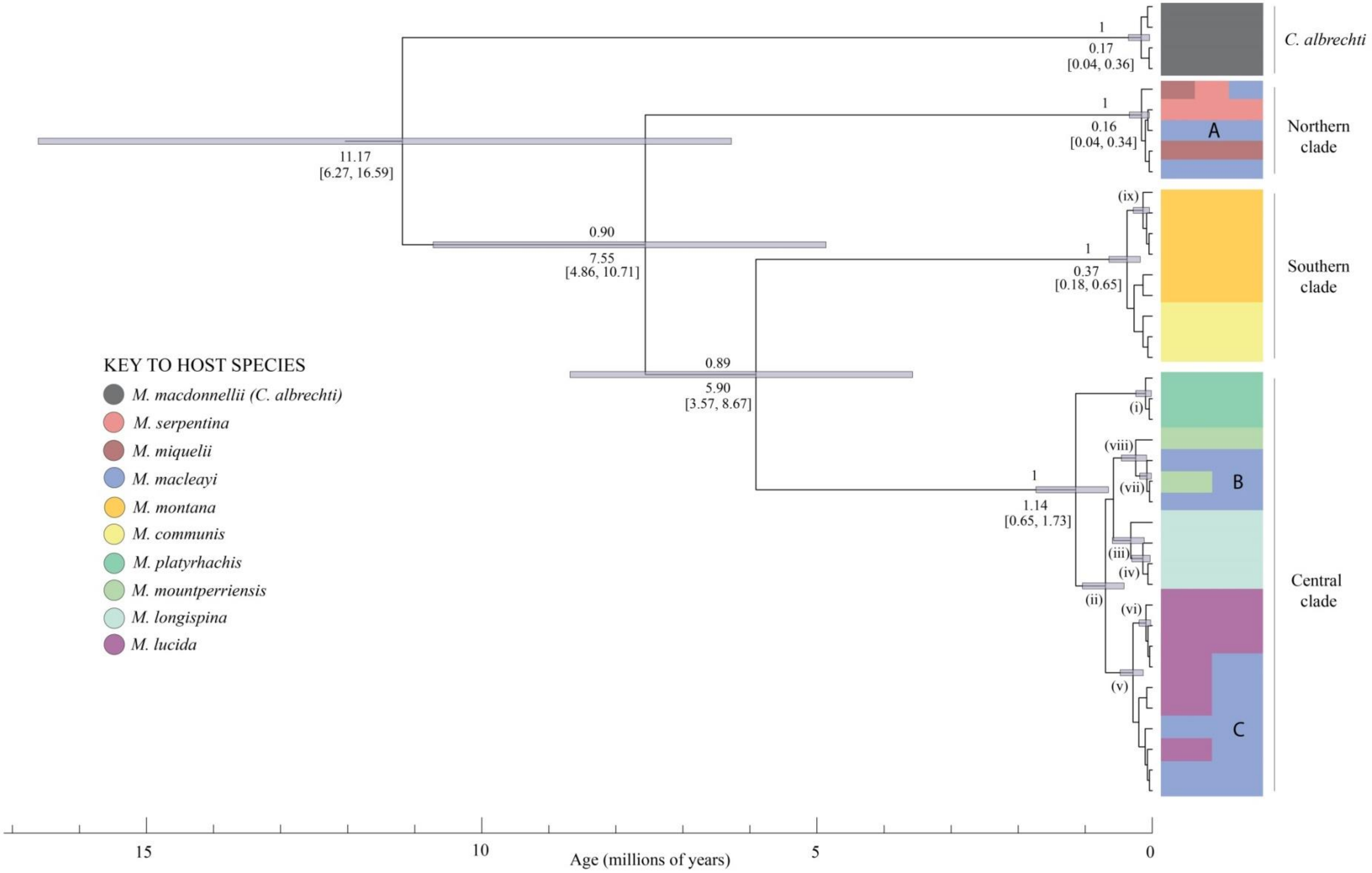


Figure S4.1. Chronogram from a BEAST analysis using a COI substitution rate of 0.0145 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

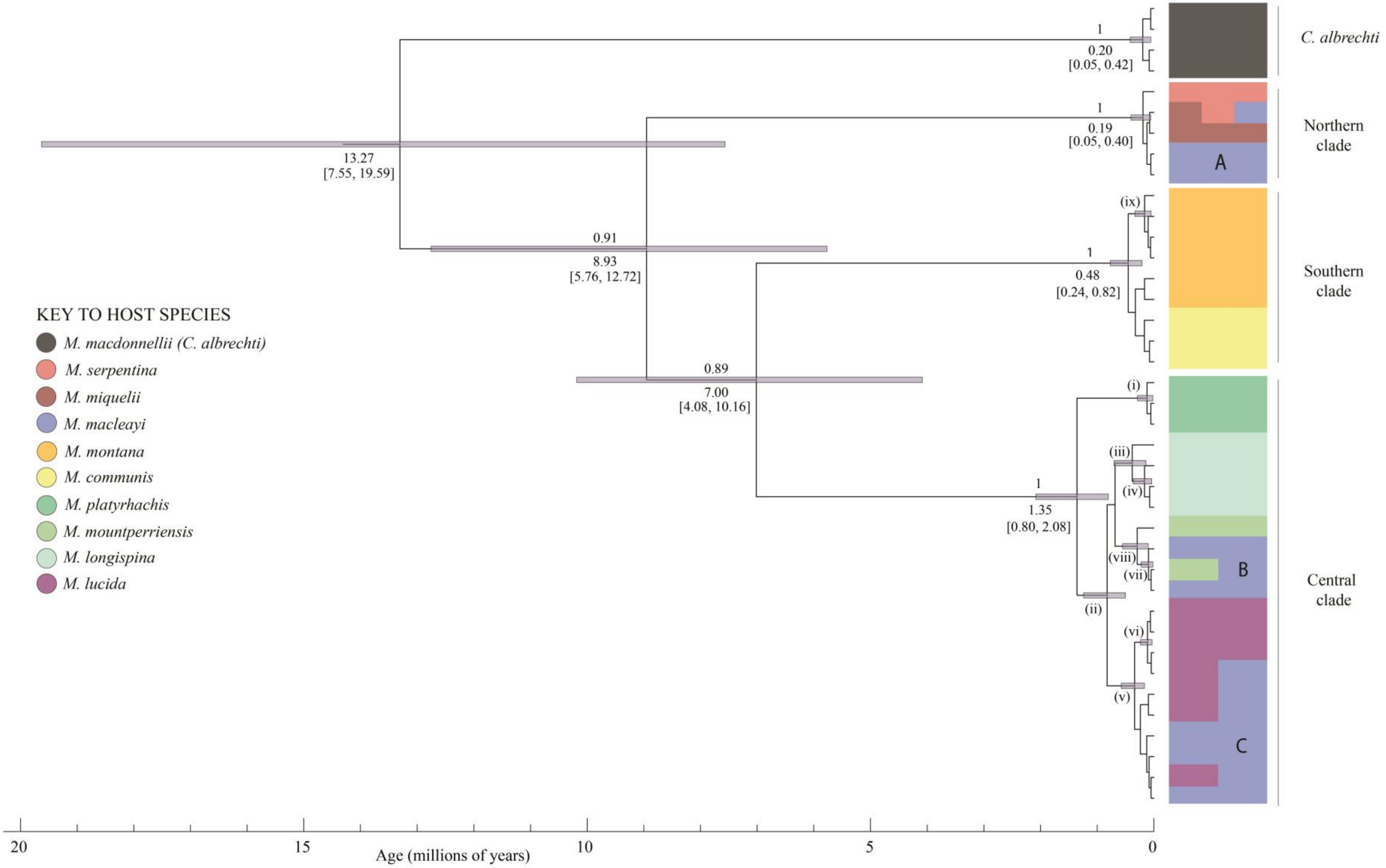


Figure S4.2. Chronogram from a BEAST analysis using a COI substitution rate of 0.0112 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

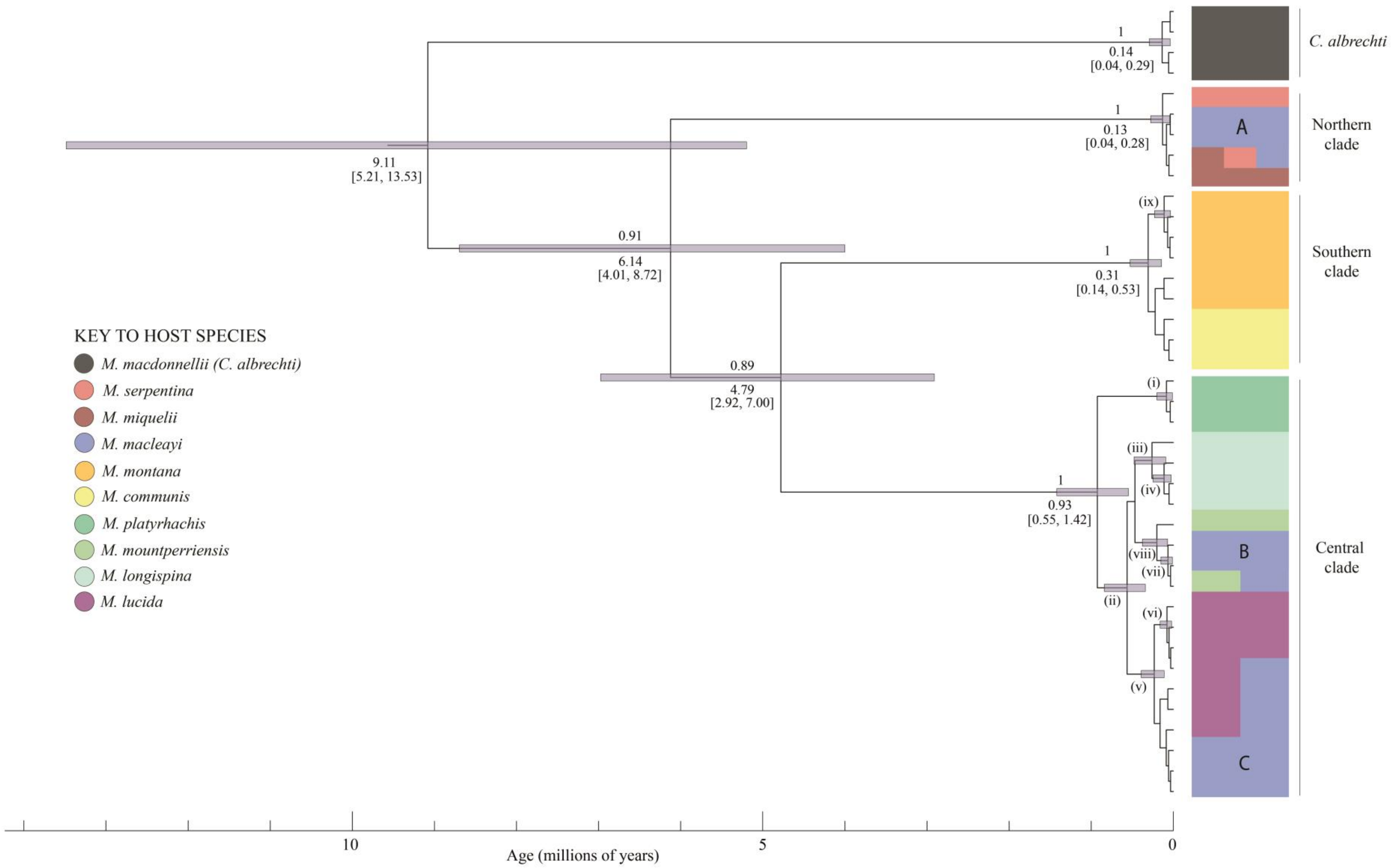


Figure S4.3. Chronogram from a BEAST analysis using a COI substitution rate of 0.0178 sites per My adapted from a FigTree 1.4 output. Node ages represent median ages and node bars represent 95% highest posterior density (HPD). The values for nodes (i) through (ix) are listed in Table S4.2.

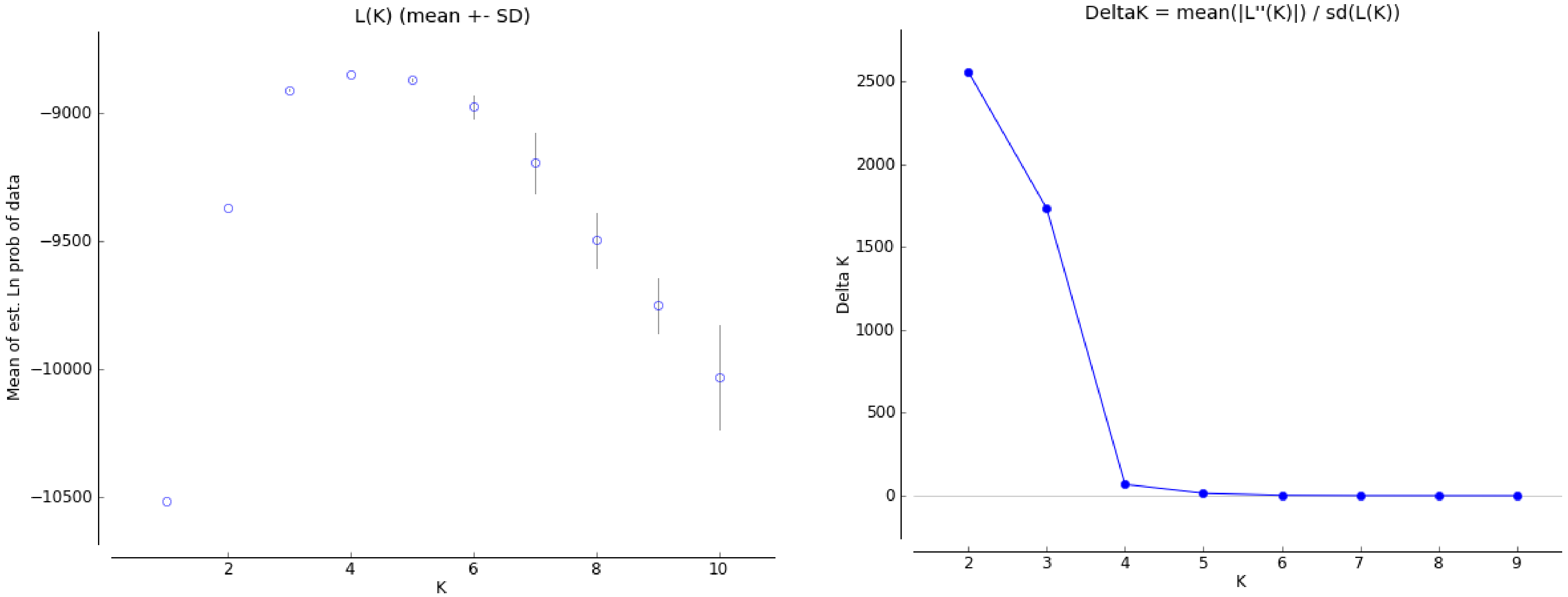


Figure S4.4. STRUCTURE HARVESTER results showing mean maximum likelihood (left) and Delta K (right). Delta K indicates a hierarchical structure with K=2 as the most likely value of K, although K=3 also has considerable support.

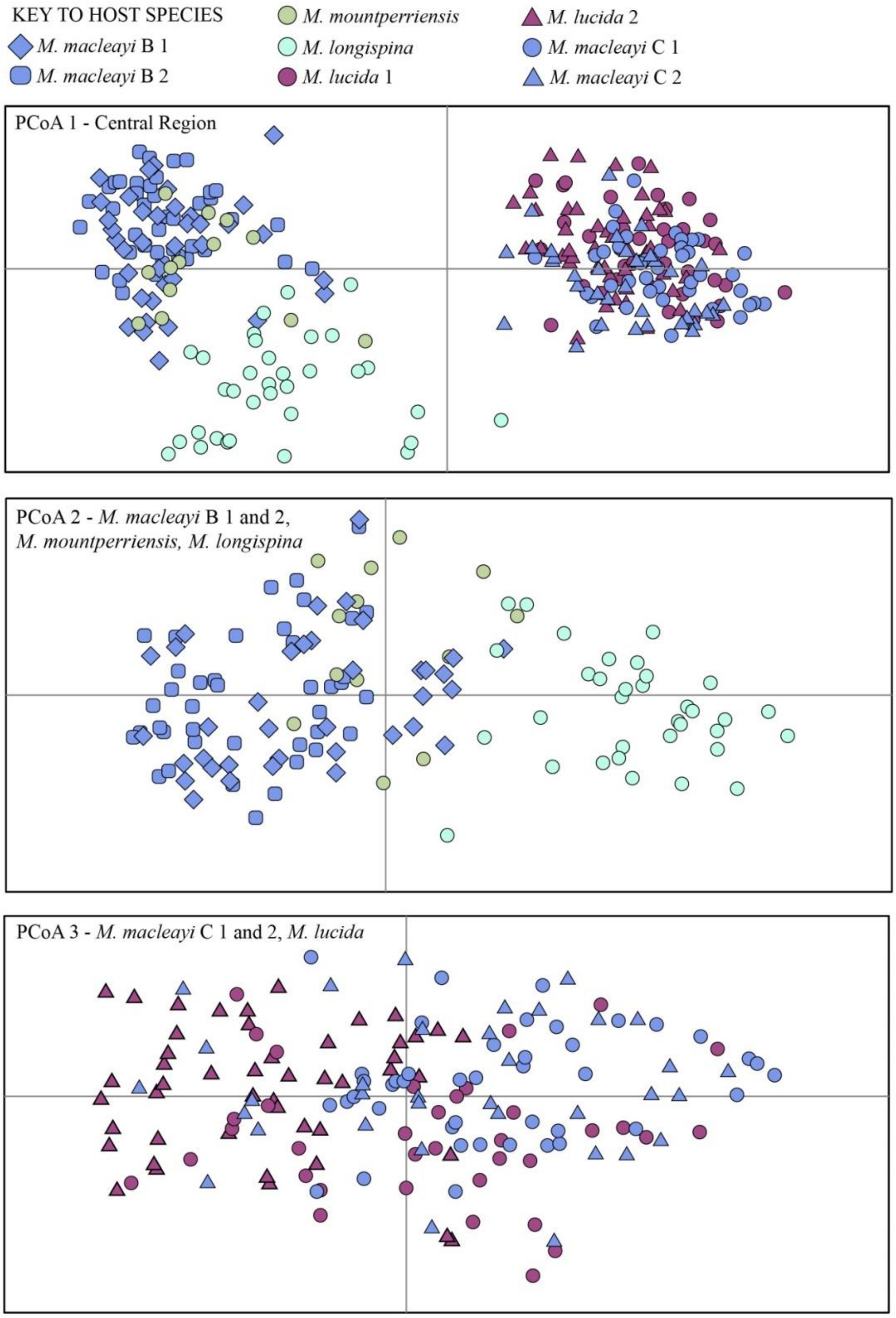


Figure S4.5. This Principle Coordinate Analysis (PCoA) shows ordination of genetic identity between different populations of *Cycadothrips* individuals from the central clade (Figures 4.2 and 4.3) and the *Macrozamia* species from which they were collected. The same populations are included in this analysis as those that were used in the STRUCTURE analysis displayed in Figures 4.4 and S4.6. Thrips from different host species are coded by colour and different populations of those host species are represented by different symbols (see key above the plots). In PCoA 1 (top), the first and second axes explain 41.4% and 15.7% (12.5% third) of the variation respectively. In PCoA 2 (middle), the first and second axes explain 34.1% and 17.1% (14.7% third) of the variation respectively. In PCoA 3 (bottom), the first and second axes explain 23.8% and 18.9% (16.2% third) of the variation respectively.

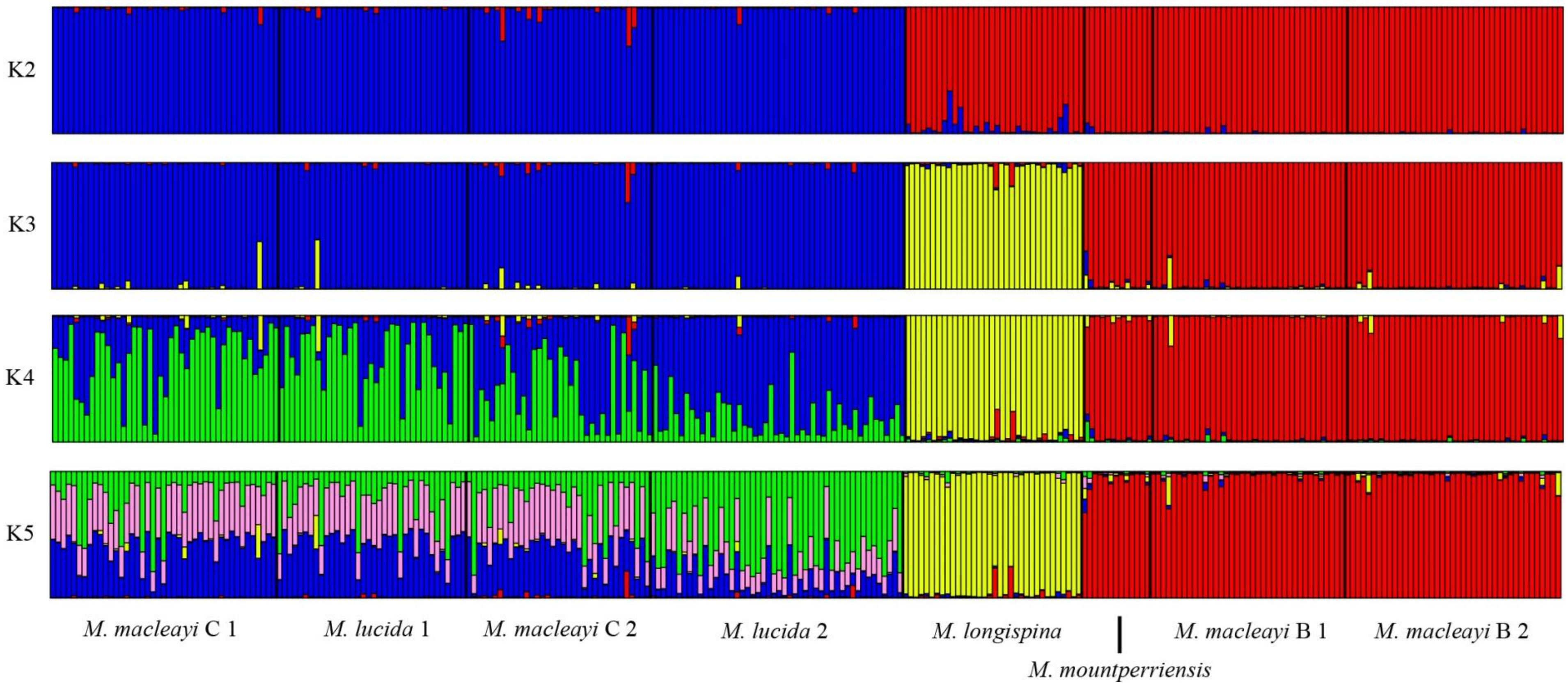


Figure S4.6. Results of the STRUCTURE algorithm illustrate the posterior probability of an individual's assignment to one of K clusters (given on the left of each diagram) based on eight microsatellite loci screened across 287 *C. chadwicki* individuals from eight populations and four *Macrozamia* species. Analyses were performed using the admixture model with allele frequencies correlated, a burn-in of 100,000 iterations and a further 1,000,000 iterations after burn-in. Each vertical bar represents one individual, comprised of one or more colours, representing the posterior probability that each individual belongs to each of the sampled populations. The K-value associated with each iteration represents the number of hypothetical populations tested for that STRUCTURE run.

Table S4.1. Host plant species, their pollinators and sampling localities for *Cycadothrips* samples included in this study. Those species pollinated by both thrips and *Tranes* weevils are indicated. The exact localities of many *Macrozamia* species are considered sensitive information in Queensland and so have not been specified precisely. *Macrozamia macleayi* is associated with different thrips clades in different areas (see results), so it is labelled A, B and C for ease of cross-referencing to the text and Figure 4.2. NP = National Park; SF = State Forest. The QDPC and Genbank code identifies the host population that each of the individuals is associated with.

***Macrozamia* species**	**Pollinator species**	**Sample district**	**QDPC and Genbank ID**
M. communis	*C. chadwicki*[1] and *Tranes*[2]	Paterson, NSW	DCR
		Smiths Lake, NSW	SML
M. longispina	*C. chadwicki*[3]	Wrattens NP, QLD	WSF
M. lucida	*C. chadwicki*[3]	D’Aguilar NP, QLD	MTN
M. macdonnellii	*C. albrechti*[4]	King’s Valley Canyon, NT	KVC
M. macleayi (A)	*C. chadwicki*[3]	Dawson Highway, QLD	DHW
M. macleayi (B)	*C. chadwicki*[3]	Gaeta, QLD	GAE
		Blackman’s Gap, QLD	BMG
M. macleayi (C)	*C. chadwicki*[3]	Beerburum SF, QLD	BEE
		D’Aguilar NP, QLD	BF7
M. miquelii	*C. chadwicki*[3]	Mount Archer, QLD	MTA
		Werriba, QLD	WER
M. mountperriensis	*C. chadwicki*[3]	Mount Perry, QLD	MPE
M. montana	*C. chadwicki*[5] and *Tranes*[5]	Styx River NP, NSW	MON
M. platyrhachis	*C. chadwicki*[5]and possibly *Tranes*[5]	Expedition Range, QLD	BDT
M. serpentina	*C. chadwicki*[3]	Milman, QLD	MIL

[1] (Jones *et al.* 2001), [1] (Mound 1991), [2] (Ornduff 1990), [3] (Forster *et al.* 1994), [4] (Mound and Terry 2001), [5] (Terry *et al.* 2008)

Table S4.2. Posterior probability, date estimates and 95% HPD for nodes (i) through (ix) for the chronograms shown in Figures 4.3, S4.1, S4.2 and S4.3.

Node	Substitution Rate	Node Posterior Probability	Date Estimate (My)	Date 95% HPD (My)
(i)	0.0112-0.0178	1	0.10	[0.01, 0.25]
	0.0115	1	0.08	[0.01, 0.20]
	0.0145	1	0.10	[0.01, 0.24]
	0.0175	1	0.12	[0.02, 0.29]
(ii)	0.0112-0.0178	0.92	0.67	[0.38, 1.07]
	0.0115	0.93	0.56	[0.34, 0.84]
	0.0145	0.93	0.70	[0.42, 1.04]
	0.0175	0.93	0.82	[0.50, 1.24]
(iii)	0.0112-0.0178	0.93	0.31	[0.11, 0.60]
	0.0115	0.92	0.26	[0.09, 0.48]
	0.0145	0.92	0.32	[0.12, 0.59]
	0.0175	0.92	0.38	[0.14, 0.70]
(iv)	0.0112-0.0178	0.95	0.13	[0.03, 0.30]
	0.0115	0.95	0.11	[0.03, 0.25]
	0.0145	0.95	0.14	[0.03, 0.30]
	0.0175	0.95	0.16	[0.04, 0.36]
(v)	0.0112-0.0178	1	0.28	[0.12, 0.48]
	0.0115	1	0.23	[0.11, 0.39]
	0.0145	1	0.29	[0.13, 0.48]
	0.0175	1	0.34	[0.16, 0.57]
(vi)	0.0112-0.0178	1	0.09	[0.02, 0.20]
	0.0115	1	0.08	[0.02, 0.16]
	0.0145	1	0.11	[0.03, 0.22]
	0.0175	1	0.11	[0.03, 0.23]
(vii)	0.0112-0.0178	1	0.08	[0.01, 0.18]
	0.0115	1	0.06	[0.01, 0.15]
	0.0145	1	0.08	[0.01, 0.19]
	0.0175	1	0.09	[0.02, 0.22]
(viii)	0.0112-0.0178	1	0.24	[0.08, 0.47]
	0.0115	1	0.20	[0.07, 0.038]
	0.0145	1	0.25	[0.08, 0.46]
	0.0175	1	0.29	[0.10, 0.55]
(ix)	0.0112-0.0178	0.99	0.13	[0.04, 0.28]
	0.0115	0.99	0.11	[0.03, 0.23]
	0.0145	0.99	0.14	[0.04, 0.28]
	0.0175	0.99	0.16	[0.05, 0.33]

Table S4.3. Pairwise F_{ST}s calculated using microsatellite data in Genepop 4.2 between central clade thrips populations as shown in Figures 4.4 and 4.5.

	***M. macleayi* C 1**	***M. lucida* 1**	***M. macleayi* C 2**	***M. lucida* 2**	***M. longispina***	***M. mountperriensis***	***M. macleayi* B 2**	***M. macleayi* B 1**
***M. macleayi* C 1**	0.0000							
***M. lucida* 1**	0.0239	0.0000						
***M. macleayi* C 2**	0.0206	0.0431	0.0000					
***M. lucida* 2**	0.0468	0.0376	0.0363	0.0000				
M. longispina	0.1833	0.1884	0.1694	0.1715	0.0000			
M. mountperriensis	0.1749	0.1781	0.1581	0.1552	0.1031	0.0000		
***M. macleayi* B 2**	0.2015	0.2004	0.1894	0.1784	0.1547	0.0354	0.0000	
***M. macleayi* B 1**	0.1985	0.183	0.1777	0.1687	0.1316	0.0195	0.0163	0.0000

Chapter Five – Spatiotemporal analysis of the population structure of a specialist thrips herbivore and pollinator across its *Macrozamia* cycad hosts

5.1 Introduction

The life history of phytophagous insects is necessarily tightly linked to that of their host plants, but how host plant interactions might lead to speciation in the insects that feed on them is unclear (Janz 2011) and sometimes the source of controversy (Berlocher & Feder 2002; Matsubayashi *et al.* 2010; Nuismer *et al.* 2010). For highly specialized insect species, gene flow among its constituent populations will be heavily influenced by the distribution of their host plants. For those herbivorous insects that are also specialized pollinators of their host plant, such as in brood-site mutualisms, gene flow among host plant populations will in turn be limited by pollinator movement and its timing.

That geographical isolation and host switching are the primary drivers of species diversification in brood-site mutualisms is clear (Althoff *et al.* 2012; Brookes *et al.* 2015; Hembry & Althoff 2016). Characterising the interaction between geography and gene flow in such systems is thus necessary to understand better how speciation might proceed under different biogeographical circumstances. Host switching, for example, is more likely to occur when gene flow is more restricted spatially through host populations being fragmented, dispersal of the insect being limited for some reason, or some combination of these two influences. I investigated these issues, as detailed below, with reference to a single *Cycadothrips* species from the *Macrozamia-Cycadothrips* brood-site mutualism.

Each *Cycadothrips* species pollinates one or two *Macrozamia* cycad species (though in one case possibly three). The species diversity of *Cycadothrips*, in other words, is lower than that of the *Macrozamia* species they pollinate (Brookes *et al.* 2015), and so whatever evolutionary processes have been responsible for this diversity have acted independently on the plant and insect, at least to some extent. Both partners in the *Macrozamia-Cycadothrips* mutualism rely on one another to reproduce. All life stages of the thrips feed on pollen of the male *Macrozamia* cones and pollination of the cycads is achieved through a push-pull mechanism that is best known with respect to *M. lucida* and the southern populations of *M. macleayi* in the eastern subtropics of Australia (Terry *et al.* 2007; Terry *et al.* 2016; Terry *et al.* 2014; Terry *et al.* 2005). The thrips are attracted to the volatile compounds released by male and female cones, but during daily thermogenic events, which last a few hours, the male cones reach temperatures up to 12°C above ambient and simultaneously increase their volatile emissions dramatically (Terry *et al.* 2007; Terry *et al.* 2014). The increased temperatures and volatile emissions, accompanied by highly elevated relative humidity levels, drive the pollen-bearing *Cycadothrips* to fly from the male cones (Roemer *et al.* 2005; Terry 2001; Terry *et al.* 2014). Female *Macrozamia* cones emit similar volatiles to those of males and so pollination is achieved when some of the free flying pollen-covered thrips are attracted to female cones. Most thrips are presumably attracted back to male cones when they cool because numbers in female cones are relatively low.

The thrips *C. chadwicki* sp. 4 (Brookes *et al.* 2015) is associated with two cycad species, *M. lucida* and *M. macleayi* (but only southern populations of this latter species, as the northern ones are likely to represent an undescribed species (Brookes *et al.* 2015)). These cycads overlap geographically and do produce hybrids in at least one locality where they overlap (Terry *et al.* 2004). Both species have a patchy distribution, and in *M. lucida* this is known to reflect its specific environmental requirements (Kaye *et al.* 2016). That means that the fragmented spatial distribution of *Macrozamia* is likely to influence the genetic structure of this thrips species, spatially and temporally. Of particular note is the question of how limited thrips dispersal is. If dispersal is limited, then thrips populations may be highly structured genetically across even short distances. Possibly, though, *Cycadothrips* individuals may move regularly between cycad populations that are separated by large distances. They may disperse when they leave male cones *en mass* during daily thermogenic events, with wind for example, or they may disperse when cycad pollen is unavailable locally, or both these processes could occur. *Cycadothrips* may diapause annually (Forster *et al.* 1994), when *Macrozamia* pollen is not available (outside of only a couple of months (Jones *et al.* 2001)), but this is not certain. If they do not diapause for most of the year, then they must disperse away from *Macrozamia* to find an alternative host and gene flow between their populations is likely to be high.

The movement of *Cycadothrips* also has direct relevance to the ecology and species diversity of *Macrozamia*. Gene flow in these cycads is mediated almost entirely by their pollinators because of the mostly limited dispersal of *Macrozamia* seeds (Hall & Walter 2014; Ornduff 1990; Snow & Walter 2007). *Macrozamia* has a short window of pollen receptivity, probably similar to that of *Encephalartos* (in which it is up to 8 days (Crosiers & Malaisse 1995)). The timing of the onset of the pollination window differs across *Macrozamia* populations that are found at different altitudes, even within a single *Macrozamia* species, but populations generally overlap to some extent in this regard. Pollination may be limited across *Macrozamia* populations if thrips populations are localised, perhaps allowing for divergence in the thrips or the cycads through local adaptation and genetic drift. Alternatively, the sequential onset and end of pollen availability across cycad populations may cause thrips to move sequentially across those populations as the (reproductive) coning season progresses, perhaps moving pollen between them.

Geographical isolation of *Macrozamia* or *Cycadothrips* populations appears to be important to speciation in the *Macrozamia-Cycadothrips* mutualism (Brookes *et al.* 2015). Investigating gene flow among populations of a single *Cycadothrips*, relative to the distribution of their host cycads, will provide insights into how speciation might proceed in this mutualism. Here I characterized the population genetic structure of the thrips *C. chadwicki* sp. 4 across its two hosts (in south-east Queensland), spatially, and through time (across two seasons, 2011 and 2015) using microsatellite markers (Brookes *et al.* 2015). I predict that within *C. chadwicki* sp. 4, geography will have the most significant influence on population genetic structure. *Cycadothrips* populations should be geographically isolated from one another. Spatially outlying populations should differ more from populations within the centre of the species distribution than will central populations from one another. Host plant species, in the absence of any other influences, should have no effect on the genetic structure of thrips populations because the data presented in Chapter 4 demonstrate that *C. chadwicki* sp. 4 are a single species that is associated with both *M. lucida* and the southern populations of *M. macleayi*. Finally, we expect that thrips populations are relatively localised (under the scenario of diapause) and that their populations should remain localised through time and not undergo regular bottleneck events.

5.2 Methods and materials

5.2.1 Sampling and DNA extraction

Cycadothrips chadwicki sp.4 (Brookes *et al.* 2015) were collected from the cones of *M. lucida* and southern *M. macleayi* cycads at 13 locations during 2015, one location during 2013, and previously from four of these same locations in 2011 (Brookes *et al.* 2015). All sample information can be found in Table 5.1. Insects were collected using a fine brush and placed in 100% ethanol. The thrips collected in 2011 were included to assess the temporal persistence of any genetic structure detected. Sampling sites covered the known geographical distribution of this insect relatively evenly. For DNA extraction, thrips were pipetted onto paper towel to evaporate the ethanol on them and each was placed individually into one well of a 96-well plate. To each well was added, 110µl of 20% Chelex (100 Resin

– BioRad) in TE (10 mM Tris HCl and 1 mM EDTA (pH 8.0)). The plate was then heated to 98°C in a PCR machine for 15 minutes.

The taxonomy of the cycads from which the thrips were sampled from also needs to be considered. The cycads currently named *M. longispina* (YFO13, Table 5.1) illustrate the current difficulty. These cycads were previously recorded as hybrids between that species and *M. macleayi.* For this study the cycads at YFO13 will be tentatively treated as *M. macleayi* given that their morphology is obviously more similar to that of *M. macleayi* than to that of *M. lucida*.

Table 5.1. Summary of sample information for the 583 *Cycadothrips chadwicki* included in this study. Some entries represent two independent collections from the same location made in different years (see text). *Macrozamia* cycads are protected species in Queensland and so specific site information has been withheld, but see map (Figure 5.1) for the distribution of the sites sampled. Pairwise geographic distance can be found in Table S5.2.

Site Code	Host Plant	Years Sampled	No. Individuals
BCN	*Macrozamia lucida*	2015	31
CNP	*Macrozamia lucida*	2015	32
BTW	*Macrozamia lucida*	2015	31
MME	*Macrozamia lucida*	2015	32
CMR	*Macrozamia lucida*	2015	30
LSF	*Macrozamia lucida*	2015	30
MFR	*Macrozamia macleayi*	2015	32
LBT	*Macrozamia lucida*	2015	32
BSF	*Macrozamia macleayi*	2011/2015	35/31
MNM	*Macrozamia macleayi*	2011/2015	43/27
MNL	*Macrozamia lucida*	2011/2015	36/26
MNG	*Macrozamia lucida*	2011/2015	48/31
YFL	*Macrozamia lucida*	2015	31
YFO	*Macrozamia macleayi/longispina*	2013	25

5.2.2 Microsatellite genotyping and summary statistics

A total of 583 female *Cycadothrips* from the 14 locations was genotyped at eight tetra-nucleotide microsatellite loci (CHAD01 to CHAD08 (Brookes & Hereward 2013)) all previously shown to have high allelic diversity (Brookes *et al.* 2015). Some of the thrips (n = 162) had been genotyped earlier

(Brookes *et al.* 2015). Between 25 and 32 thrips were genotyped from each location. Only female thrips were genotype because sex determination in *Cycadothrips* is haplo-diploid, with the females being diploid. Microsatellites were confirmed and binned manually using Geneious 9.1.6 microsatellite plugin 1.4.3 and summary microsatellite statistics can be found in Table 5.2.

Table 5.2. Summary statistics for the microsatellite loci genotyped for *Cycadothrips chadwicki* sp. 4 individuals (Table 5.1). All loci are tetra-nucleotides (Brookes & Hereward 2013).

Locus and Dye	**$\sum Na$**[a]	**$\hat{N}a$**[b]	***Ho***[c]	***He***[d]	**HWE**[e]	**Est. Null Alleles**	**gF_{ST}**	**gF_{ST} ENA**
CHAD01 (PET)	12	7.8	0.586	0.695	5 (2)	0.07	0.067	0.067
CHAD02 (NED)	17	10.7	0.739	0.771	3 (1)	0.02	0.044	0.044
CHAD03 (VIC)	14	8.8	0.807	0.825	0 (0)	0.01	0.021	0.020
CHAD04 (NED)	36	17.9	0.855	0.885	1 (0)	0.01	0.013	0.013
CHAD05 (FAM)	9	4.4	0.639	0.652	1 (0)	0.02	0.070	0.069
CHAD06 (FAM)	27	14.9	0.904	0.894	3 (2)	0.01	0.008	0.008
CHAD07 (PET)	26	15.7	0.874	0.897	2 (0)	0.01	0.008	0.008
CHAD08 (VIC)	25	12.2	0.784	0.794	3 (0)	0.02	0.027	0.026

[a] Total number of alleles.

[b] Average number of alleles.

[c] Average observed heterozygosity.

[d] Average expected heterozygosity.

[e] Deviations from Hardy-Weinberg equilibrium and their within-loci sequential Bonferroni corrected values (Holm 1979) shown in brackets.

5.2.3 Microsatellite analyses

Hardy-Weinberg exact tests and a probability test for linkage disequilibrium were carried out using Genepop 4.2 (Raymond & Rousset 1995; Rousset 2008) and multiple significance tests were accounted for using sequential Bonferroni correction (Holm 1979) within loci. Null allele estimates and global F_{ST} calculations, with and without ENA correction, were made in FreeNA (Chapuis & Estoup 2007). The individual based clustering algorithm STRUCTURE 2.3.4 (Falush *et al.* 2003; Falush *et al.* 2007;

Pritchard *et al.* 2000) was used to investigate population structure. K values 1 through 10 were calculated with 2,000,000 iterations using a burn-in of 200,000 iterations and with 20 replicates each. STRUCTURE Harvester (Earl & VonHoldt 2012) was used to estimate the most likely value of K using the Evanno method (Evanno *et al.* 2005) and Clumpak was used to cluster replicates of K together (Kopelman *et al.* 2015). Genepop 4.2 (Raymond & Rousset 1995; Rousset 2008) was used to test for significant genic differentiation using exact G-tests (dememorisation 10,000, batches 500, iterations per batch 10,000) as well as to calculate pairwise F_{ST}. Genic rather than genotypic differentiation was used because haploid males will break the association between alleles (genotype) at each locus.

Analysis of Molecular Variance (AMOVA (Excoffier *et al.* 1992)) was used to examine the effects of time and host plant species on genetic structure using Arlequin 3.5 (Excoffier & Lischer 2010) with 1000 permutations. Four locus by locus AMOVA were performed, with each separately using number of different alleles (F_{ST}-like) and sum of square differences (R_{ST}-like). R_{ST}-like analyses are appropriate to include because the mutation rate of these microsatellite loci is probably high, given large *Cycadothrips* population sizes and the high allelic diversity of microsatellite loci (Gaggiotti *et al.*1999; Balloux and Lugon-Moulin 2002; Putman and Carbone 2014). The first AMOVA included all populations without any further groupings. The second AMOVA had two host plant groups, but included only geographically paired populations that had different host plants, either *M. lucida* or *M. macleayi* (with YFO13 (Table 5.1) being treated as *M. macleayi*). The third AMOVA was performed only on the four populations sampled earlier (2011) and included two groups, each of which represented the different sampling years 2011 and 2015. The fourth AMOVA included two regional groups, a northern and a southern one, and only the populations within them (i.e. BCN, CNP, BTW, LBT, MNG, and YFO13 (northern) and CRM, MME, MNL, MNM, and MFR (southern)). Discriminatory Analysis of Principle Coordinates (DAPC) was performed using the *adegenet* package (Jombart *et al.* 2010) for R (R 2016), and using 100 cross-validation replicates to choose the number of principle and discriminatory coordinates retained. Populations for DAPC were grouped according to their non-significant pairwise interactions (Table S5.1) as follows, a northern group (as above, and YFL), a southern group (as above, and LSF), and then separately, BSF (and these comparisons will be justified further in the results section).

Pairwise F_{ST} was used in Mantel tests to estimate isolation by distance (IBD) using IBD Web Service 3.23 (Jensen *et al.* 2005) with 10,000 randomizations. One IBD analysis included all populations sampled during 2015 and a second one excluded populations from the first analysis that were genetic outliers despite their close spatial proximity to other sites. I tested for population bottlenecks with Bottleneck 1.2.02 (Piry *et al.* 1999) using Wilcoxon tests (because they are more appropriate with a lower number of loci (Piry *et al.* 1999)) and the two-phase model (TPM) with 0.36 variance and three separate single-step mutation proportions (i.e. 70%, 80% and 95%). Although 95% is considered to be the most generally applicable proportion for microsatellites (Piry *et al.* 1999) the high number of alleles per locus in this study (Table 5.2) means that lower proportions are likely to be more suitable (Ellegren 2004; Lai & Sun 2003).

5.3 Results

5.3.1 Microsatellite genotyping

The allelic diversity across the eight loci was moderate to high, with each locus having between nine and 36 alleles, and the total number of alleles across all loci being 166. The average number of alleles found in each population was 92.4, slightly more than half the number of alleles found across all populations. Three loci showed evidence of deviation from HWE after sequential Bonferroni correction, but only in five of the populations sampled. No locus pairs showed evidence of linkage disequilibrium. Average null allele estimates for all loci were below 10% in all populations, although these estimates were higher for CHAD01 relative to the other seven loci. All microsatellite summary statistics can be found in Table 5.2.

5.3.2 Individual population assignment and pairwise differentiation

STRUCTURE analysis indicated that K=4 and (Delta K = 16.5) and K=2 (Delta K = 14.4) were the most likely values of K but the absolute likelihood values are similar (Figure S5.1). Although K=4 had the highest Delta K, a high number of subpopulations is unlikely as few individuals, if any, have a membership coefficient above 0.5 for most clusters in the K=4 analysis (Figure 5.1). Clusters such as these are considered to be ‘ghost’ clusters by other authors (Guillot *et al.* 2005; Puechmaille 2016). The STRUCTURE plots for samples from the same localities, but with collections separated by four years, did not differ across sample dates (Figure 5.1). Differentiation between populations appears to be low overall but some geographically-organised population structure can be identified when the output is visualized (Figure 5.1). Although some population genetic structure is present (see below) all

Cycadothrips populations are broadly similar. K=1 and K=2 are probably the most biologically meaningful results and so the usefulness of STRUCTURE for exploring this data appears to be low.

DAPC only assesses between group genetic differences, ignoring within population variation (Jombart *et al.* 2010), and so this method has more effectively recovered the subtle genetic structure present among *C. chadwicki* sp. 4 populations (Figure 5.2). The group designations chosen for DAPC here (Figure 5.2) were chosen based on the results from exact G-tests and STRUCTURE results (Figure 5.1).DAPC indicates the presence of two groups that are somewhat differentiated (Figure 5.2). One group contains all but one population, and is further subdivided into a northern cluster (blue - YFL, CNP, BTW, MNG, LBT, BCN) and a southern cluster (grey - MNL, MNM, MFR, and LSF). The second group comprises just one population, (orange – BSF).

Pairwise exact G-tests for genic differentiation showed that most populations were significantly different from one another in pairwise comparisons (Table S5.1). By mapping the pairwise connections that were non-significant (green lines in Figure 5.1) and moderately significant to slightly non-significant ($p = >0.01$ and <0.06; yellow lines in Figure 5.1) the connectivity between populations can be visualized. Two major groups can be observed to have non-significant pairwise differences (green) within them, namely the highly inter-connected northern populations (BCN, BTW, CNP, LBT, MNG, and YFO13) and the southern populations, which exhibited step-wise connectivity (MNM > MNL > MME > CMR). Moderately significant to slightly significant interactions (yellow) represent connections between these two groups as well as links with some other populations.

Two populations were significantly different from all other populations, BSF and LSF (Table S5.1 and Figure 5.1). LSF is a spatial outlier on the western edge of the distribution of *M. lucida*, so this relationship can be expected. Nevertheless, the pattern from the STRUCTURE output and DAPC for LSF is broadly similar to that of the other southern populations (Figures 5.1 and 5.2). BSF is geographically close to a number of other eastern populations and so spatial separation cannot explain why it is genetically distinct from all other populations. Although BSF did differ moderately significantly ($p = 0.022$) after four years (Table S5.1) the pattern in the output of STRUCTURE was not dramatically different across years and remained distinct from the surrounding populations at both times (Figure 5.1). The population YFL was also somewhat differentiated from the populations around it and these, in turn, were generally highly similar to one another (Figure 5.1). Some variation is also expected given the high number of alleles relative to the number of individuals sampled.

5.3.3 Influence of population, host species, and time

Populations were significantly different from one another in both the F_{ST}-like (3.0% var. and p=0.000) and R_{ST}-like (4.0% var. and p=0.000) analyses (Table 5.3). Paired host plant populations were significantly different from one another in the F_{ST}-like analysis (1.9% var. and p=0.001) but not the R_{ST}-like analysis (0.2% var. and p=0.289), and populations within each host plant group were significantly different from one another in both analyses (respectively 2.1% and 5.0% var., and p=0.000 for both). Sampling year had no significant effect on temporally sampled populations in either analysis (-0.8% var. for both, and p=0.991 and 0.674 respectively). The northern and southern regions, determined by non-significant differences between populations within each (Table S5.1), were significantly different from one another in both analyses (respectively 2.0% and 4.6% var., and p=0.000 for both). Of the three pairs of populations that had different host species in the same locality, two were significantly different from one another in exact G-tests (Table S5.1). Also in exact G-tests, three of the four temporally sampled populations (MNM, MNL, and MNG) showed no significant difference after four years and the fourth (BSF) showed only a moderately significant difference (p = 0.022) (Table S5.1).

Although there is spatial genetic structure in these thrips, they are generally similar to one another when compared with other members of the *C. chadwicki* species complex (Brookes *et al.* 2015). The only population that represents an entirely separate group in DAPC are the thrips collected from site BSF (Figure 5.2). The geographical outlier site, LSF, fits entirely within the same cluster as the other southern populations despite the geographical distance that separates them.

5.3.4 Isolation by distance

A significant isolation by distance relationship was found although the relationship was weak (p = 0.013, R^2 0.116; Figure 5.3A). The relationship was stronger when a single genetic outlier population, BSF, was excluded from the analysis (p = 0.000, $R^2 = 0.288$; Figure 5.3B). BSF was a genetic outlier despite its geographical proximity to other populations (and less than five kilometres from the nearest), and also in being significantly different from all other populations (Table S5.1) and forming a distinct group in the DAPC analysis (Figure 5.1). A correlation between geographic and genetic distance is clearly evident in *C. chadwicki* sp. 4, even though the genetic structure of populations can vary considerably over smaller geographical scales as in population BSF, and even though other factors may play a more significant role.

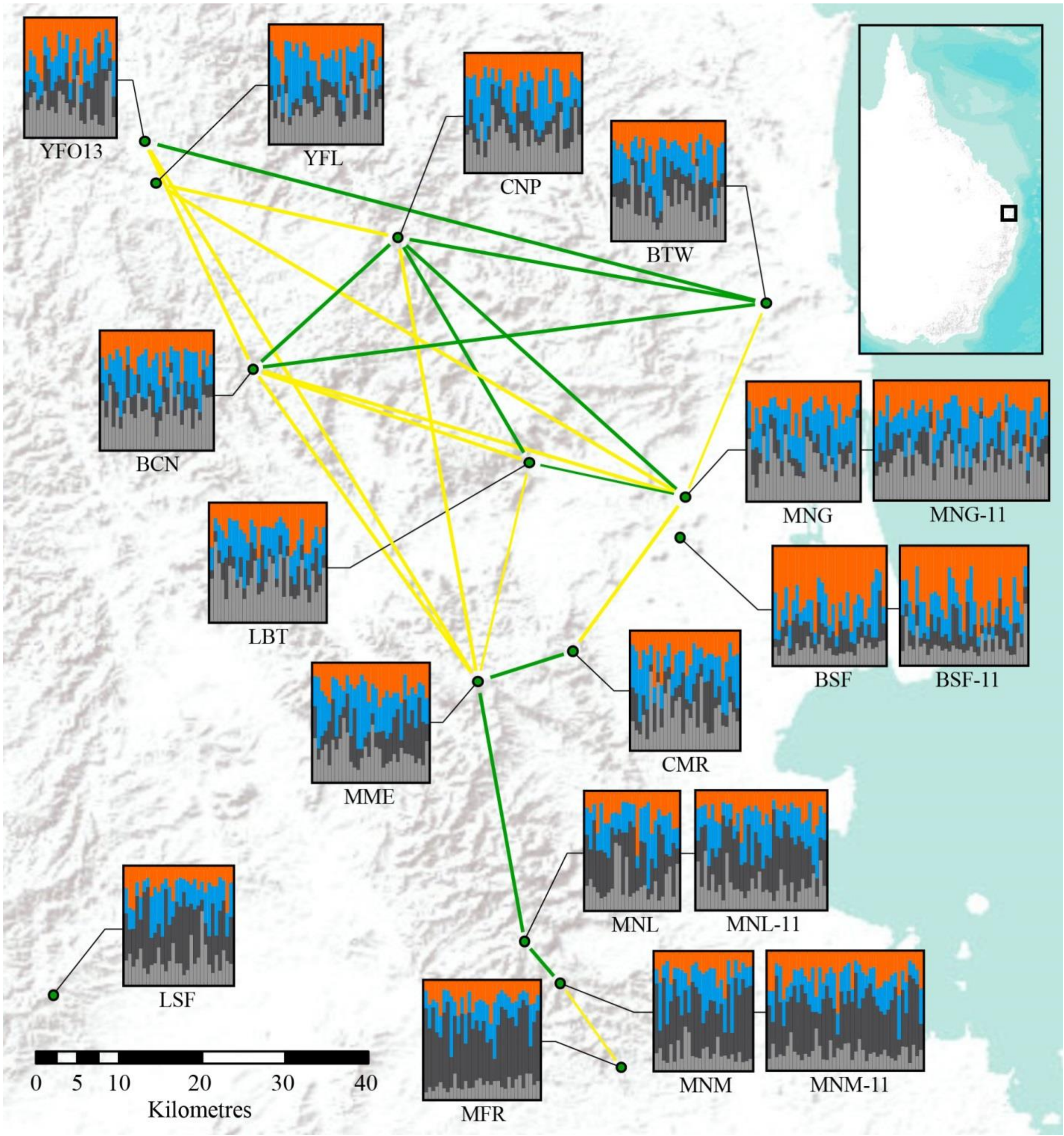


Figure 5.1. A topographic map of south-east Queensland showing the locations in which individuals of *Cycadothrips chadwicki* sp. 4 were collected (Table 5.1) together with the results of a K=4 STRUCTURE analysis (Figure S5.2). The latter is displayed in coloured boxes associated with each collection site. The results of a pairwise exact G-test to assess genic population differentiation (Table S5.1) are represented as lines connecting populations. The green lines represent pairwise comparisons that were not significantly different across populations and the yellow lines represent ones that were moderately or nearly significant ($p = >0.01$ and <0.06). Pairwise interactions of populations sampled in 2011 are not shown. Boxes alongside one another represent two independent collections from the same locations but made in different years (Table 5.1). All thrips were collected in 2015 except for those labelled with the numbers 11 (2011) and 13 (2013) in their site codes.

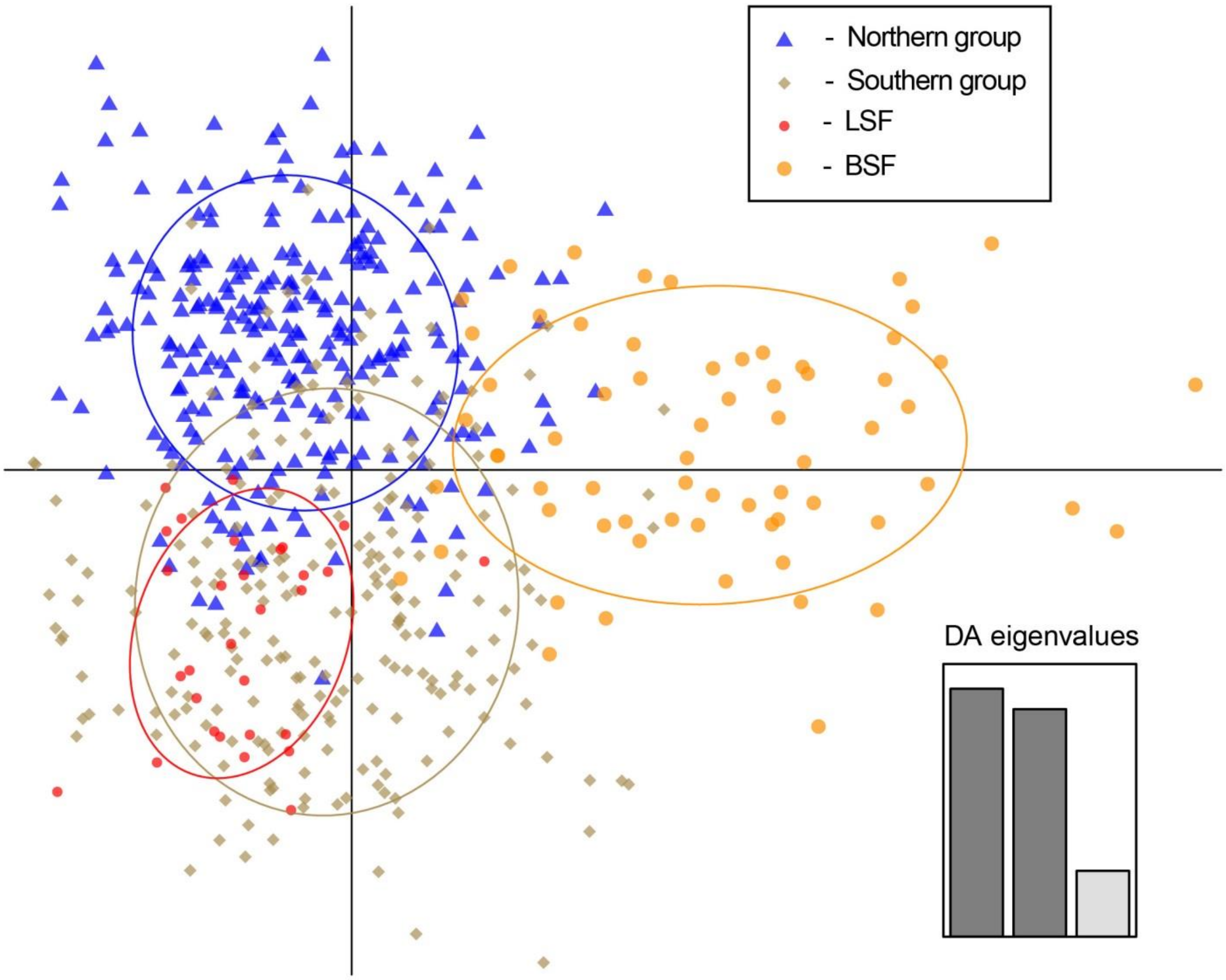


Figure 5.2. Discriminatory Analysis of Principle Components (DAPC) using microsatellite data of all genotyped *Cycadothrips chadwicki* sp. 4 (Table 5.1). Individuals are coloured to indicate the northern and southern groups, as well as site BSF, all of which were identified by STRUCTURE analysis and exact G-tests (Figure 5.1). Site LSF was grouped separately because it a geographical outlier, and to observe its relationship to other populations. The number of discriminatory axes that were retained from the analysis was 3.

5.3.5 Population bottlenecks

The results of Wilcoxon tests for the presence of genetic bottlenecks showed low to no evidence for such events (Table 5.4). At single-step mutation proportions of 80% and 95%, no populations showed evidence of genetic bottlenecks, with non-significant results for all heterozygosity-excess tests (Table 5.4). In these same tests six and seven populations respectively showed evidence for heterozygosity deficiency (Table 5.4). At the lowest proportion of single-step mutations used, 70%, one population showed evidence for a genetic bottleneck, YFL (70% one-sided p; Table 5.4).

Table 5.3. Locus by locus AMOVA results for four different analyses conducted in Arlequin (Excoffier & Lischer 2010), each with F_{ST}-like and R_{ST}-like analyses. AMOVA a) includes all populations. AMOVA b) includes two host plant groups, population pairs of *M. lucida* or *M. macleayi*. AMOVA c) includes only the four populations that were sampled in both 2011 and 2015, and so year sampled is included. AMOVA d) includes only the two groups of populations that had no significant pairwise differences, the northern and southern groups (Figure 5.1).

	No. of different alleles – (F_{ST}-like)			**Sum of square differences – (R_{ST}-like)**		
Source of variation	**Variance explained**	**F-statistics**	**P val.**	**Variance explained**	**F-statistics**	**P val.**
a) all populations						
Among populations	3.0%	0.030	0.000	4.4%	0.044	0.000
Among individuals within populations	4.7%	0.050	0.000	5.7%	0.059	0.003
Within individuals	92.1%	0.079	0.000	89.9%	0.101	0.000
b) all populations, across hosts						
Among host plants	1.9%	0.019	0.001	0.2%	0.002	0.289
Among populations	2.1%	0.022	0.000	5.0%	0.050	0.000
Among individuals within populations	4.6%	0.048	0.000	7.7%	0.081	0.000
Within individuals	91.3%	0.087	0.000	87.1%	0.129	0.000
c) temporal populations only						
Among years	-0.8%	-0.008	0.991	-0.8%	-0.008	0.674
Among populations	3.8%	0.037	0.000	6.0%	0.060	0.000
Among individuals within populations	4.6%	0.047	0.000	8.6%	0.090	0.001
Within individuals	92.5%	0.075	0.000	86.2%	0.138	0.000
d) northern and southern regions						
Among regions	2.0%	0.020	0.000	4.6%	0.046	0.000
Among populations	1.4%	0.015	0.000	1.3%	0.014	0.001
Among individuals within populations	4.9%	0.051	0.000	5.3%	0.056	0.011
Within individuals	91.6%	0.084	0.000	88.9%	0.111	0.001

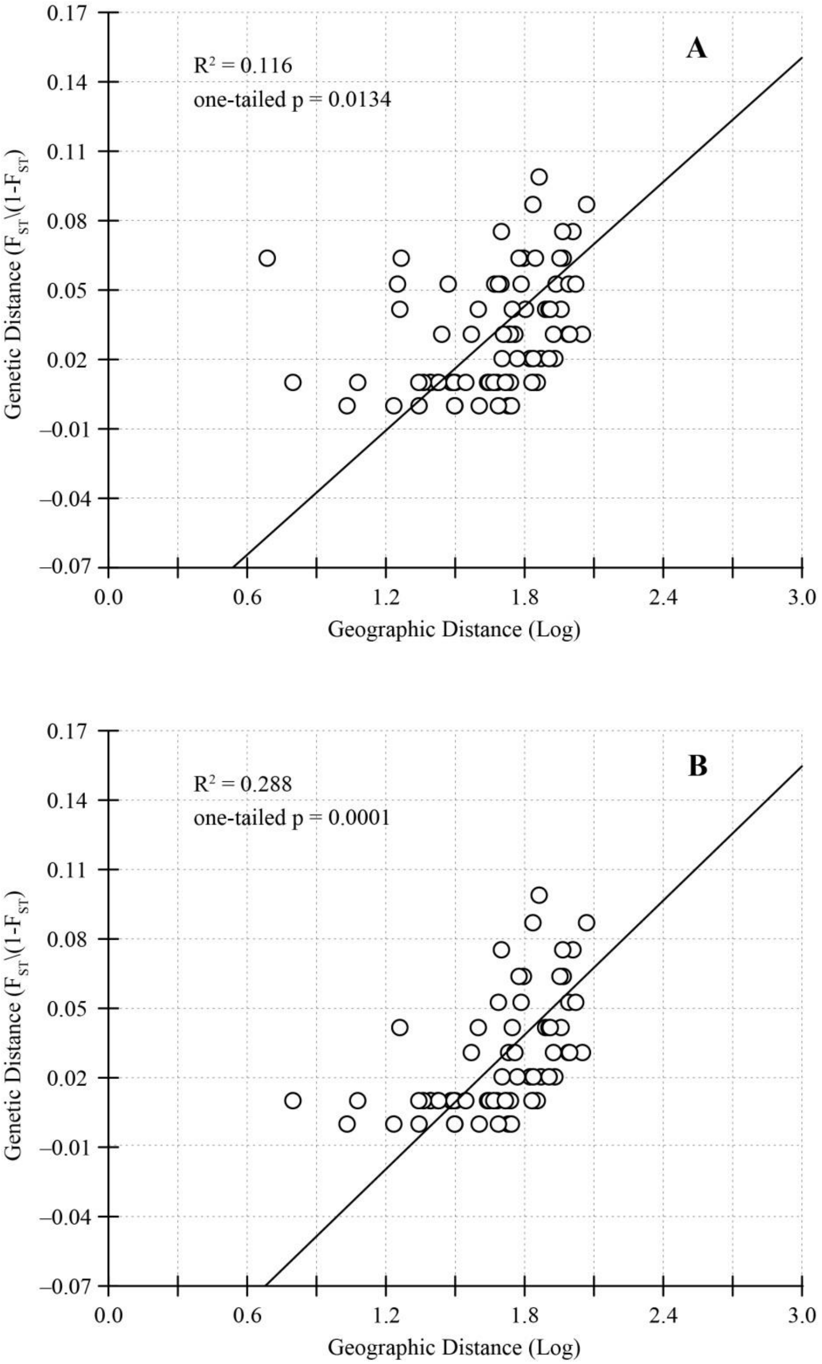


Figure 5.3. The results of two isolation by distance analyses (IBD) using Mantel tests performed on the *Cycadothrips chadwicki* sp. 4 genotype data. Analysis A includes all populations that were sampled during 2015 and analysis B excludes the single population that was a genetic outlier with respect to its geographical location, BSF (Figure 5.1). In both cases significant isolation by distance was detected with low to moderately high R^2 values.

Table 5.4. The results of analyses using the software Bottleneck 1.2.02 to perform Wilcoxon tests for heterozygosity excess (one-sided p) and for both heterozygosity excess and heterozygosity deficiency (two-sided p) in *Cycadothrips chadwicki* sp. 4 genotype data. These values were computed using a two-phase mutation model (TPM) with 0.36 variance and using three different single-step mutation proportions; 70%, 80%, and 95%.

Site Code	70% one-sided p	70% two-sided p	80% one-sided p	80% two-sided p	95% one-sided p	95% two-sided p
BCN	0.422	0.844	0.963	0.195	0.963	0.195
CNP	0.273	0.547	0.980	0.055	0.990	0.027
BTW	0.371	0.742	0.527	1.000	0.629	0.844
MME	0.727	0.641	0.986	0.039	0.990	0.027
CMR	0.156	0.313	0.875	0.313	0.875	0.313
LSF	0.727	0.641	0.963	0.195	0.973	0.074
MFR	0.809	0.461	0.980	0.055	0.980	0.055
LBT	0.629	0.844	0.996	0.012	0.998	0.008
BSF	0.098	0.195	0.844	0.383	0.875	0.313
MNM	0.680	0.742	0.973	0.074	0.973	0.074
MNL	0.809	0.461	0.990	0.027	0.990	0.027
MNG	0.320	0.641	0.809	0.461	0.875	0.313
YFL	0.002	0.004	0.727	0.641	0.809	0.461
YFO13	0.125	0.250	0.629	0.844	0.727	0.641
BSF11	0.190	0.382	0.844	0.383	0.844	0.383
MNM11	0.629	0.844	0.986	0.039	0.986	0.039
MNL11	0.902	0.250	0.986	0.039	0.994	0.020
MNG11	0.037	0.074	0.986	0.039	0.996	0.012

5.4 Discussion

5.4.1 Population connectivity and consequences for local adaptation and speciation

The spatial and temporal genetic structure of *C. chadwicki* sp. 4 populations (Figure 5.1) indicates that a significant amount of gene flow occurs between all thrips populations. Determining the extent of gene flow between *Macrozamia* populations cannot be done with the data presented here, but it can be inferred only from what is known about the morphology of these cycads. Populations of *M. lucida* and southern *M. macleayi* can persist near one another despite hybridization between them possibly occurring on occasion (Terry *et al.* 2004)). The persistence of morphologically distinct *Macrozamia* populations must be explained because they share pollinator populations that are included in the analyses presented here. Gene flow among *Cycadothrips* populations must occur at times when corresponding gene flow does not occur in *Macrozamia*, but the results presented here cannot indicate when gene flow occurs relative to the timing of pollination because gene flow is high in the long term. Gene flow among *Macrozamia* populations must be lower than among *Cycadothrips* populations over the same distribution, and probably occurs outside of the short pollinator window of *Macrozamia.* Cycad populations can thus remain, or become, differentiated without the same happening to the *Cycadothrips* populations they are associated with.

Population connectivity across *C. chadwicki* sp. 4 populations may be maintained by the mostly regular distribution of its *Macrozamia* hosts. The pattern of connectivity that is indicated by the exact G-tests (Figure 5.1) and the genetic distinctiveness of population BSF (Figure 5.2) suggest that the movement of *Cycadothrips* between *Macrozamia* populations may occur in a step-wise manner. The distribution of *Macrozamia* may be the relevant factor because the correlation between genetic and geographic distance alone is relatively low (Figure 5.3). The population BSF (Figures 5.1 and 5.2) also highlights this relationship, because it is the most genetically distinct population (Figure 5.2) despite being only five km from another population (MNG) that clusters with all other populations. In the long term high amounts of gene flow clearly occur between most *C. chadwicki* sp. 4 populations (Figure 5.1), but those few populations that are genetically distinct remain so, even across years (Tables 5.2 and S5.2). This suggests, presumably, that the movement of thrips is limited in some respects through their close association with *Macrozamia*. Clearly the next step for understanding how evolution occurs in this system is to investigate gene flow in the *Macrozamia* populations that correspond to *C. chadwicki* sp. 4, and *Cycadothrips* populations more generally.

The high similarity between *C. chadwicki* sp. 4 populations shown here contrasts strongly with the population genetic structure of the several putative species found across the central clade of *C. chadwicki* as a whole (Brookes *et al.* 2015). The results presented here thus support the view that the central clade *C. chadwicki* is a species complex, as *C. chadwicki* sp. 4 shows little to no population fragmentation across its entire distribution. These results also have implications for understanding how speciation occurs in *Cycadothrips.* With the high amount of gene flow observed among *C. chadwicki* sp. 4 populations in different localities and on different host species, stabilizing selection should also be acting on the push-pull pollination mechanism (Terry *et al.* 2007) as this pollination mechanism is the primary interactive aspect of the shared biology of the thrips and cycads. Adaptation of *C. chadwicki* sp. 4 populations to geographically isolated *Macrozamia* populations is thus unlikely without further disruption to the distribution of *Macrozamia*. Geography was determined here to be important for speciation in *Cycadothrips* and *Macrozamia* (Brookes *et al.* 2015), and so the conditions that lead to speciation may be the product of rare events that isolate both thrips and cycad populations.

5.4.2 *Cycadothrips* and *Macrozamia* population fragmentation and conservation

The high amount of gene flow between geographically disparate sites was surprising. In particular, the small and isolated site LSF is not genetically distinct from other southern populations relative to the large geographical distance that separates it from all other populations. Comparatively, the site BSF differs more genetically from all other population pairs despite being separated by such a short geographic distance. The historical versus present day distribution of *Macrozamia* may be affecting the results here. Determining the relative effect of the environmental requirements of *Macrozamia* and habitat loss on the present day distribution of *Macrozamia* will be important for understanding the results shown here for *Cycadothrips*. The western distribution of *Macrozamia* was more regular before significant habitat loss occurred, connecting LSF to the other southern populations that the *Cycadothrips* associated with LSF are genetically similar to (Figures 5.1, 5.2, and S5.3). This may mean that the site LSF has become isolated only recently and so there is not yet a strong genetic signal. Site BSF has very few *Macrozamia* plants which are restricted to only the highest point in the immediate vicinity (pers. obs.). It may be that suitable environmental conditions for these cycads (Kaye *et al.* 2016) are rare in this area, and so their distribution has become restricted over time due to climactic change. These factors will need to be investigated specifically in future.

Any further fragmentation of distribution of these cycads could result in the isolation of *Cycadothrips* and *Macrozamia* populations over even small distances. Smaller isolated *Macrozamia* populations have been observed with no *Cycadothrips* present (pers. obs.), and so small populations of cycads may not produce pollen cones regularly enough to support *Cycadothrips* populations in the long term. Cycads are among the most threatened of plant groups, and of high conservation priority (Donaldson 2003). It is already well known that cycad populations are vulnerable to habitat fragmentation, but the inferences made here suggest that even cycad populations that look viable may be vulnerable to loss of their pollinator, followed by the gradual decline of the cycads themselves. Gene flow between *C. chadwicki* populations is clearly high, but if movement occurs in a step-wise across *Macrozamia* populations, then recolonization of thrips following their local extinction may be unlikely if habitat loss in an area has been significant.

5.5 Supporting information

Table S5.1. Pairwise exact G-test for genic differentiation. For samples taken during 2015, non-significant pairwise associations are highlighted green, moderately or nearly significant associations (p = >0.01 and <0.06) are highlighted in yellow, and both are displayed in Figure 5.1 using lines of the same colour to connect populations, so as to display the relationship between populations graphically. Highly significant associations are indicated with an asterisk. Samples found below the horizontal line are from the temporal dataset only and are not shown in Figure 5.1.

	BCN	CNP	BTW	MME	CMR	LSF	MFR	LBT	BSF	MNM	MNL	MNG	YFL	BSF11	MNM11	MNL11	MNG11	YF013
BCN	-																	
CNP	0.648	-																
BTW	0.702	0.545	-															
MME	0.055	0.019	0.002	-														
CMR	0.007	0.004	0.005	0.089	-													
LSF	*	*	0.000	0.001	0.000	-												
MFR	*	*	*	*	*	*	-											
LBT	0.025	0.232	0.005	0.007	0.000	*	*	-										
BSF		*	*	0.000	*	*			-									
MNM	0.000	0.000	0.000	0.000	0.000	0.000	0.034		0.000	-								
MNL	0.004	0.000	0.004	0.390	0.005	0.008	0.000	*	0.000	0.239	-							
MNG	0.043	0.261	0.024	0.003	0.057	*	*	0.081	*	*	*	-						
YFL	0.002	0.012	0.000	0.002	0.001	*	*	0.000	*	0.000	*	0.053	-					
BSF11	*	0.000	*	0.000		*	*		0.022	*	0.000		*	-				
MNM11	0.000	*	0.000	0.000	0.000	0.000	0.007		0.000	0.244	0.473	*	*	0.000	-			
MNL11	0.000	0.000	0.000	0.001	0.000	*	*	*	*	0.000	0.466	*	*	*	0.000	-		
MNG11	0.007	0.391	0.001	0.000	0.000	*	*	0.002	*	*	0.000	0.420	0.536		*	*	-	
YF013	0.055	0.002	0.320	0.030	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	-

Table S5.2. Pairwise F_{ST}s are presented below the diagonal and above it is the pairwise geographic distance between populations in kilometres values (km). All thrips were collected in 2015 except for those labelled with the numbers 11 (2011) and 13 (2013) in their site codes.

	BCN	**CNP**	**BTW**	**MME**	**CMR**	**LSF**	**MFR**	**LBT**	**BSF**	**MNM**	**MNL**	**MNG**	**YFL**	**BSF11**	**MNM11**	**MNL11**	**MNG11**	**YFO13**
BCN	-	22.04	40.13	53.8	52.84	97.73	102.05	30.33	46.87	90.84	85.25	43.64	26.59	46.87	90.84	85.25	43.64	29.4
CNP	0	-	55.34	44.39	47.92	77.83	92.24	31.48	49.82	80.4	74.29	48.56	24.66	49.82	80.4	74.29	48.56	29.62
BTW	0	0	-	54.74	46.43	112.29	92.7	31.68	29.48	84.24	80.55	24.75	66.7	29.48	84.24	80.55	24.75	69.18
MME	0	0.01	0.01	-	10.77	58.64	48.58	26.79	27.64	37.11	31.45	31.26	68.9	27.64	37.11	31.45	31.26	73.77
CMR	0.01	0.01	0.01	0	-	68.87	50.07	23.02	17.77	39.76	35.11	21.97	71.54	17.77	39.76	35.11	21.97	76.17
LSF	0.03	0.04	0.03	0.02	0.02	-	60.87	81.36	86.28	53.84	50.46	89.82	97.91	86.28	53.84	50.46	89.82	102.59
MFR	0.07	0.07	0.06	0.05	0.07	0.05	-	73.01	63.69	11.96	18.22	68.56	116.9	63.69	11.96	18.22	68.56	121.85
LBT	0.01	0	0.01	0.01	0.01	0.04	0.09	-	18.46	62.37	57.28	17.16	52.03	18.46	62.37	57.28	17.16	56.22
BSF	0.05	0.05	0.05	0.03	0.05	0.05	0.04	0.06	-	54.82	51.07	4.86	70.24	0	54.82	51.07	4.86	74.28
MNM	0.04	0.04	0.03	0.03	0.04	0.03	0.01	0.06	0.03	-	6.28	59.65	105.06	54.82	0	6.28	59.65	110.01
MNL	0.02	0.02	0.02	0	0.01	0.02	0.04	0.03	0.03	0.01	-	55.82	98.95	51.07	6.28	0	55.82	103.91
MNG	0.01	0	0.01	0.01	0.01	0.06	0.08	0	0.06	0.06	0.04	-	67.85	4.86	59.65	55.82	0	71.71
YFL	0.01	0.01	0.02	0.01	0.01	0.05	0.08	0.01	0.06	0.05	0.03	0.01	-	70.24	105.06	98.95	67.85	5.02
BEE11	0.04	0.04	0.04	0.02	0.03	0.05	0.05	0.05	0	0.03	0.02	0.05	0.04	-	54.82	51.07	4.86	74.28
BF711	0.04	0.04	0.03	0.02	0.04	0.03	0.01	0.06	0.02	0	0	0.06	0.06	0.02	-	6.28	59.65	110.01
MTN11	0.03	0.04	0.03	0.02	0.02	0.03	0.07	0.05	0.05	0.03	0.01	0.05	0.04	0.04	0.02	-	55.82	103.91
NGU11	0	0	0.01	0.01	0.01	0.05	0.07	0	0.05	0.05	0.03	0	0	0.04	0.05	0.04	-	71.71
YFO13	0.01	0.02	0	0.01	0.02	0.02	0.03	0.02	0.03	0.02	0.02	0.02	0.03	0.03	0.02	0.04	0.02	-

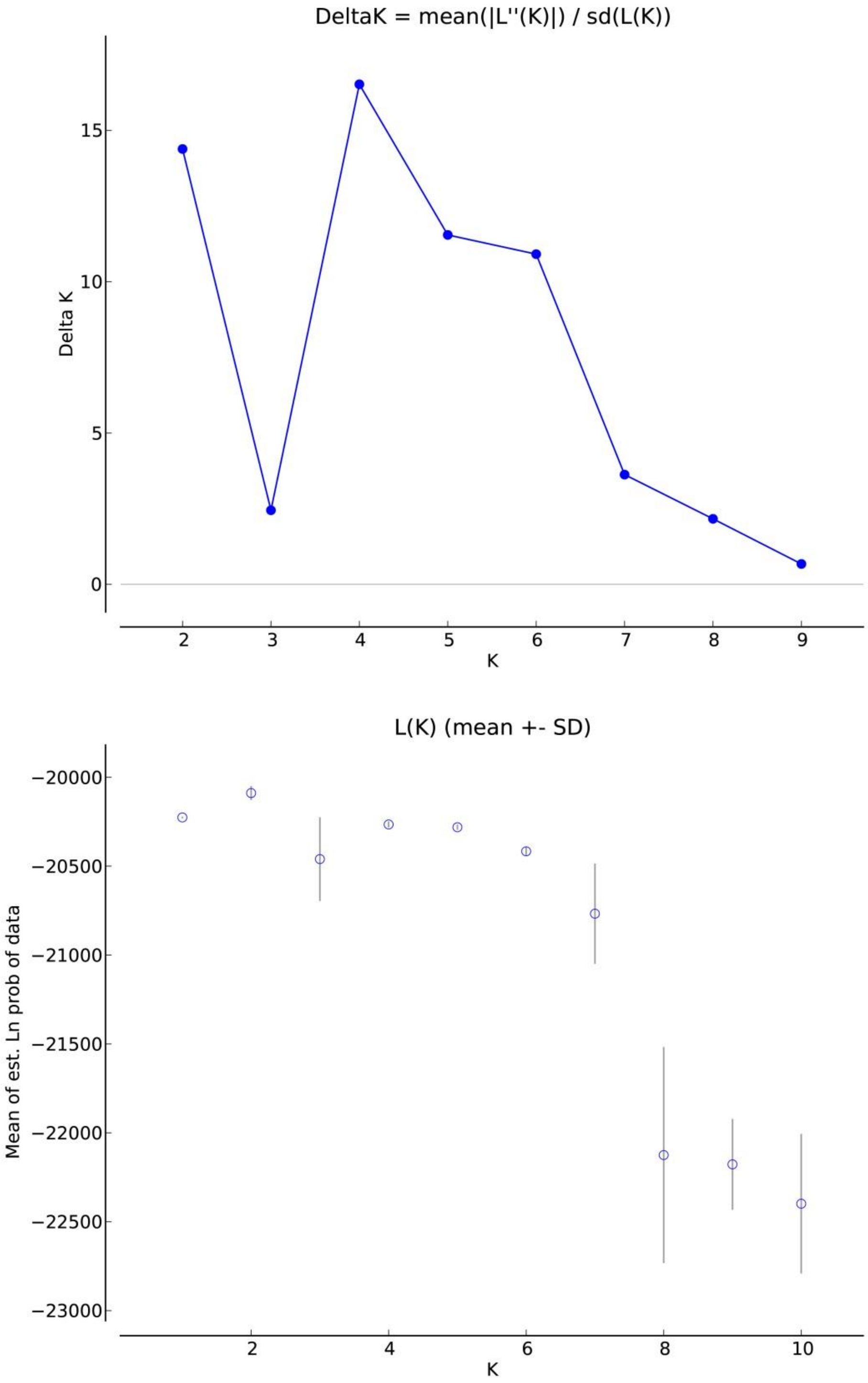


Figure S5.1. The output from STRUCTURE harvester showing Delta K (top) and mean of the likelihood of K (bottom). K=4 had the highest value of Delta K (16.5) but the value for K=2 was also high (14.4). The Delta K values were low in all cases, indicating low

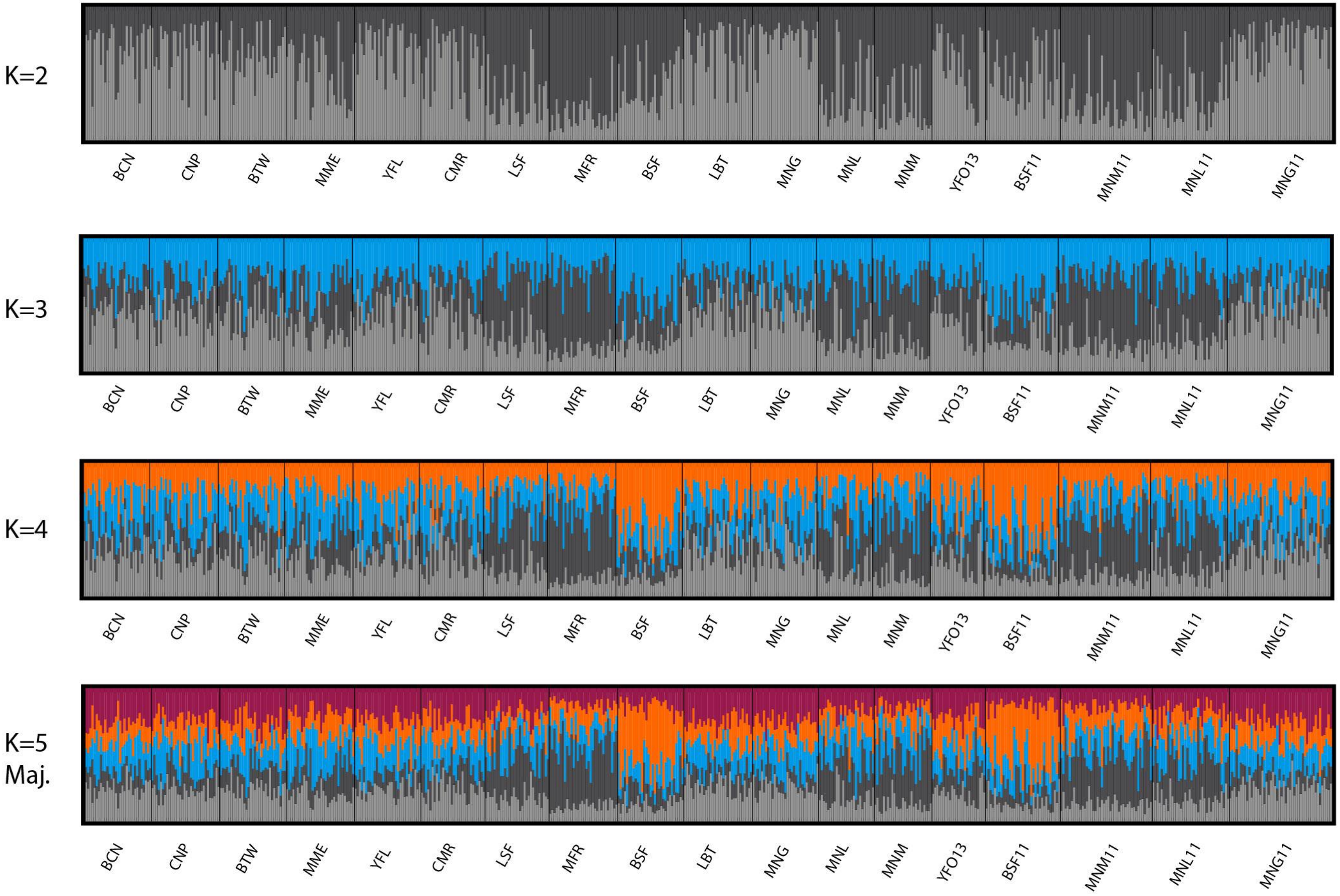


Figure S5.2. The results of the STRUCTURE analyses for K=2 through K=5. K=5 is a major cluster (11/20). SRUCTURE harvester indicated that K=4 was the most likely value of K (Figure S5.1).

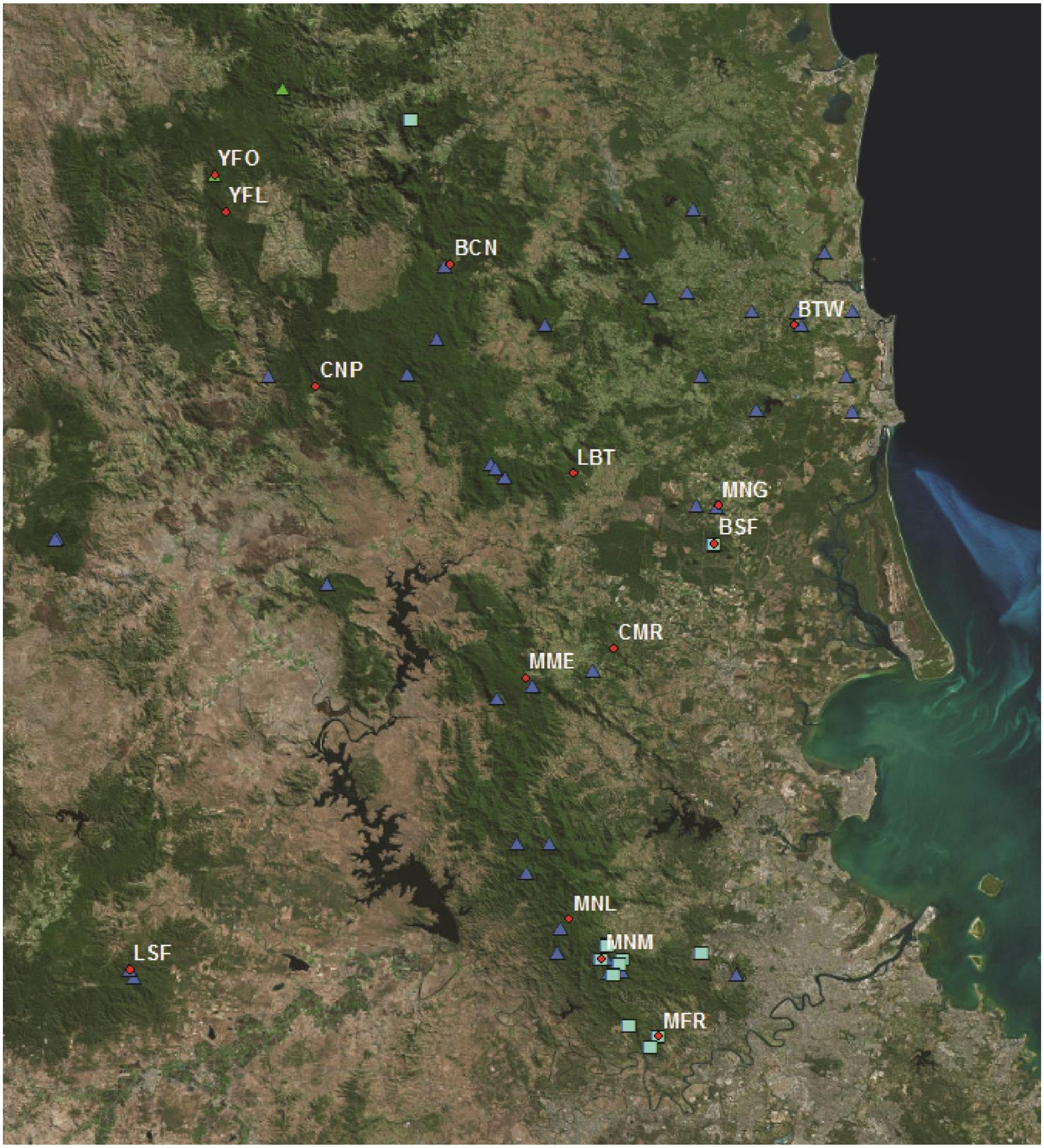


Figure S5.3. The cycad populations sampled and represented in Figure 5.1 are mapped onto this map with satellite imagery (ArcMap 10.2 basemap) to visualize the connectivity of the populations with remaining vegetation (following clearing of the landscape). Red dots are sites where individuals of *Cycadothrips chadwicki* sp. 4 were sampled from cycads for use in this study (Table 5.1). The following represent host records for the three *Macrozamia* species sampled from (HERBRECS database); teal squares – *M. macleayi*, blue triangles – *M. lucida*, and green triangles – *M. longispina*. Other populations are known for which host records do not exist, but these represent the majority of sites.

Chapter Six – Discussion: The spatial genetic structure of phytophagous insect species and the evolutionary process

6.1 Introduction

Phytophagous insects are among the most diverse organisms on earth, in part because of their particular associations with plants (Futuyma & Agrawal 2009; Jaenike 1990; Labandeira & Sepkoski 1993; Lewinsohn *et al.* 2005). Such insects exhibit considerable differences in their host specificity, even among closely-related species, but most are considered to be host specialists (Forister *et al.* 2015; Jaenike 1990; Jermy 1984). Why these insects are so speciose and how different patterns of host use evolve have been central questions in research on insect-plant interactions for a considerable time. A number of hypotheses have been proposed, and various approaches taken, to explain how these associations might evolve and thus how the patterns observed have been generated (Forister *et al.* 2012; Futuyma & Moreno 1988). Most hypotheses and approaches focus, perhaps justifiably, on the nature of the interaction between insects and their host plants (Kergoat *et al.* 2017). These include trade-offs between specialist and generalist lifestyles (Bernays 2001; Futuyma & Moreno 1988; Joshi & Thompson 1995), speciation through host specialization (Drès & Mallet 2002; Futuyma & Moreno 1988; Jermy 1984), and coevolutionary speciation (Althoff *et al.* 2014; Ehrlich & Raven 1964; Janzen 1980; Thompson 2005). Although research on the interaction between insects and their plants is clearly justified, this focus has often come at the expense of investigating those external processes that simultaneously affect both insects and their host plants.

One process that may affect insects with different patterns of host association differentially is the likelihood of population fragmentation occurring through changes in the distribution of their host plants in response to environmental change. Over long evolutionary time scales significant environmental change (a common occurrence (Petit *et al.* 1999; Zachos *et al.* 2001)) will alter the distribution of insects and their host plants (Dynesius & Jansson 2000; Walther *et al.* 2002) and may modify the associations of insect species with their host plants. For herbivorous insects, the availability of each host plant species changes through space and time (Strong 1979). Such changes in distribution could cause an insect species to interact with novel host plants or to lose contact with others. Environmental change may also affect the extent of gene flow between populations and, where no gene flow occurs, isolated insect populations may then enter upon different evolutionary trajectories. For some such populations this may mean adapting to different subsets of host plants, and perhaps even speciation (Janz *et al.* 2006; Peterson & Denno 1998).

The spatial relationships between insects and their host plants, coupled with environmental change, thus influence the evolution of insect-host associations through their effect on host plant distribution (Janz & Nylin 2008; Slove & Janz 2011; Strong 1979). The impact of a change in the distribution of a single host plant species will differ according to the host breadth of an insect. Change in the distribution of any one host plant will, in theory, be of relatively lower consequence for the population connectivity of a generalist insect, relative to a specialist. Generalist insects may, however, experience more frequent change because they are likely to be distributed over a larger geographical area (Janz 2011). There is evidence for a relationship between host breadth and insect distribution from butterflies in the subfamily Nymphalinae, where those butterfly species with a wider host breadth tend to be distributed more extensively (Slove & Janz 2011). Host plant interactions are therefore likely to influence insect population genetic structure through the number of host species with which an insect interacts, the specificity of insect host use (dependent on the specific adaptations of each insect species), and the geographical distribution of the host plants with which each insect associates.

Here I investigate the feasibility of using the population genetics literature on phytophagous insects, in combination with the results presented in Chapters 3 and 5, to analyse the relationship between the spatial genetic structure of insect populations and their diet breadth. Given the difficulty of obtaining sensible results from an analysis such as this (Nakagawa *et al.* 2017), an initial exploratory approach seemed appropriate. The aim of the analysis presented here is therefore to produce a data set to test the suitability of the proposed approach, including the choice of genetic measures, what constitutes an appropriate insect species, study suitability, marker selection, and the categorization of insect host use. The number of studies that might be included in a more comprehensive study, given the strictures of the proposed approach, is estimated. My specific prediction is that the populations that make up relatively specialist herbivorous insect species should be more fragmented and so have increased genetic differentiation over shorter geographic distances than the populations of more generalist herbivorous insects.

6.2 Methods and materials

6.2.1 General selection criteria for each study and organism

For a study to qualify for inclusion in the analysis outlined above, the following criteria were established. The reproductive mode of the insect had to be entirely sexual and at least one life stage of both sexes must depend entirely on higher plants. An arbitrary minimum number of five different geographic populations had to be analysed in the study. The insect populations included in the study had to have been distributed on a single contiguous landmass (because of the difficulties presented by island populations) and samples had to have been relatively equally distributed geographically and temporally. If, in any study, one population did not meet these spatiotemporal criteria (or even a few) then these populations were excluded and this is noted in Table S6.1. Studies undertaken over extremely small spatial scales (less than 50 kilometres (km) between the two most distant sampled sites), were also excluded. The genus, family, and order of each insect was recorded to assess the frequency with which different insect groups were represented in the data set.

6.2.2 Molecular marker selection criteria, literature search, and statistical approach

For a study to be included, it must have contained insect microsatellite genotype data and have used these data to produce pairwise F_{ST}s. Further, pairwise geographic distance values must have been available (or been possible to calculate from other available information). Pairwise F_{ST}s and pairwise geographical distances were used to perform isolation by distance (IBD) analyses for each study. Mean pairwise F_{ST} and highest pairwise F_{ST} were also recorded and used as simpler genetic measure fors each study, and these values was also used to compare the most genetically distinct populations from each study. From the IBD analyses conducted for each study, the IBD slope was used to infer the broad relationship between genetic and geographic distance, and the IBD R^2 value used to assess the 'fit' of the relationship. The scatter of the IBD plots can be used to assess whether genetic and geographic distance were continuously distributed (Peterson & Denno 1998), and so the broad suitability of each data set could be assessed. In some cases authors were contacted to clarify certain information, or to provide it when it was not contained in a publication. In each of these cases the source of this additional information is listed in the notes section of Table S6.1.

Only a single marker type was used so that marker type did not have to be treated as a distinct factor (see Riginos *et al.* (2011). Microsatellite markers were chosen because of their consistent and continued use in population genetics studies (Guichoux *et al.* 2011) and because they were used in the case studies of *N. viridula* (all *N. viridula* from Chapter 2, and *N. viridula* E from Chapter 3) and *C. chadwicki* (all central clade *C. chadwicki* from Chapter 4, and *C. chadwicki* sp. 4 from Chapter 5). Studies with few loci, and those in which very high null allele estimates were common (>10%), were excluded. If mitochondrial sequence data were available this was recorded but this was not a requirement for a study to be included. The specific advanced search criteria for the Web of Science database was broad and excluded two commonly found insect groups that were deemed inappropriate; TS=(microsatellite*) AND TS=(insect*) NOT TS=(Culicidae OR Formicidae). The search started with the most recently published studies, beginning in February 2017, and a data set was assembled for comparison with the data from *N. viridula* (Chapter 3) and *C. chadwicki* (Chapter 5).

Two sets of analyses were performed for each of the insect species that were investigated in previous chapters of this thesis, but only one analysis for each species was included in the final dataset. The first data set included only the individuals genotyped from populations of a single putative species from each system, and these were *N. viridula* E from eastern Australia (and which has already been represented in Figure 3.2b) and *C. chadwicki* sp. 4 from southern Queensland (already presented in Figure 5.3a). It was these data sets that were included in the final data set and analysed alongside the other studies. The second data set for each species included all genotyped individuals, regardless of species status, those from Chapter 2 for *N. viridula* and Chapter 4 for *C. chadwicki*. This was done so that the IBD plots, mean pairwise F_{ST} and highest pairwise F_{ST} could be compared for data sets that were known to include populations of uncertain species status (Chapters 2 and 4), with already published data sets for which such eventualities could only be postulated.

6.2.3 Categorization of host use

A brief literature review for each insect species was conducted to determine the host plants used by each. Review articles were used in preference because the host records would already have been scrutinized for accuracy. Databases such as those from the Centre for Agriculture and Bioscience International (CABI) were treated as reviews when species information could be found in them. The categorization of host use by the insect herbivores concerned is perhaps the most critical aspect of this analysis as well as the most error prone. Any categorization of host use will inevitably include incorrect or poor host records, unrecognized species complexes of insects and plants, variation in the role and importance of hosts for insects, and judgements of host suitability made using laboratory observations rather than information from the field. Host records for generalist insects would more likely include errors (relative to specialists) because of the greater number of hosts involved. Nevertheless, as much error as possible was removed to limit the impact on the results of the analyses through the use of the categorisation scheme outlined below. In some cases, the risk of such errors was obvious, and so studies were excluded (e.g. those insect species with geographically separate and highly divergent mitochondrial clades (Kindler *et al.* 2012; Maroja *et al.* 2007)).

A categorisation scheme was developed to give a weighting for the host plant specificity of each insect species at the host plant taxonomic levels of genus, family, and order. For each insect species included in the analysis, I determined the total number of plant species it uses, as well as the genus, family, and order to which each host plant belonged. Plant cultivars were not considered separate species. A ratio of the number of species from the most frequently found plant group of each taxonomic rank, relative to the total number of host plant species, was calculated. In some cases, the host specificity at a taxonomic rank was not clear because of poor host plant records for insect species. In these cases the weighted host specificity of the next highest taxonomic level was used. The moth *Plutella xylostella* illustrates how this weighting was applied. Host records from CABI for *P. xylostella* indicate that this moth feeds on 21 host plant species in 15 genera, five families, and four orders. At each taxonomic rank the most abundant host plant species were as follows; five host species from a single genus, 17 host species from a single family, and 18 host species from a single order. The weighted host specificity values at the ranks of Genus, Family, and Order were thus estimated at 0.24, 0.81, and 0.86, respectively. The mean of these values was used to represent total host specificity, in this case 0.64. A value close to zero represents extremely low host specificity and a value close to one represents extremely high host specificity.

The terms monophagous, oligophagous, and polyphagous are used here to describe the broad host relationships of each insect species based on weighted host specificity (Table S6.2), but they are not used in the presentation of any of the ANOVAs (the numerical weighted host specificity values are used for the ANOVAs). Monophagous represents insects that are specific to one host plant Genus and oligophagous to one plant Family. No insects appeared to be specific at the taxonomic level of Order and so polyphagous represents insects that are not specific to host plants at the levels of Genus or Family. The first time the weighted host specificity of each insect passed 80% at a particular host plant taxonomic level it is considered a specialist at that level. For example, the weighted host specificity of *P. xylostella* first passed 80% at the taxonomic level of family (0.81, Table S6.2) and so this insect is considered to be oligophagous.

6.2.4 Spatial analysis of genetic structure

Mantel tests were performed using the Isolation by Distance Web Service (IBDWS) 3.23 (Jensen *et al.* 2005) to assess the correlation between genetic distance and geographical distance with 10,000 bootstrap replicates. The log of geographic distance was used, and genetic distance was transformed using the calculation F_{ST} / (1- F_{ST}) (Rousset 1997). Although F_{ST} can be calculated in multiple ways the assumption is made that any differences between these methods should have only a small impact on the IBD relationship, an approach that has been taken in other similar studies (Riginos *et al.* 2011). The independent variables taken from these analyses for use in later statistical analyses are the IBD slope and the R^2 value of the IBD relationship. The IBD slope does not account for the absolute distances involved, and so mean pairwise geographic distance was used as a measure of the overall study distance. Specific information for each study can be found in Table S6.2, and this includes weighted host specificity (at the taxonomic levels of genus, family, order, and the mean of these three weighted values), IBD slope, IBD R^2, highest pairwise F_{ST}, mean pairwise geographical distance (km), and the number of populations.

6.2.5 Statistical analyses of IBD and weighted host specificity

Analysis of variance (ANOVA) was performed in R (R 2016). The *aov* function was used and effect size measured with Eta squared which was calculated using the *lsr* R package. A total of 12 analyses was performed. Weighted host specificity was the dependent variable, and ANOVAs were carried out independently for each taxonomic rank separately (Genus, Family, and Order) as well as separately for the Mean of these weighted host specificity values. The independent variables were IBD slope, IBD R^2, and highest pairwise F_{ST}. These variables were analysed separately in ANOVA for the four host specificity values detailed above. Mean pairwise geographic distance (km) was included as another independent variable in each ANOVA, and so was its interaction with other independent variables.

6.3 Results

6.3.1 Literature search results

A final data set was assembled using data from 20 insect species, with 18 species obtained from studies found using the Web of Science database. Data for two insects came from chapters included in this thesis, Chapters 3 (*N. viridula*) and 5 (*C. chadwicki*). The low number of suitable studies relative to the number of entries searched in the Web of Science database was not surprising given that the search terms were so broad. However, the intent of the search was to capture as many suitable studies as possible and so broad search terms were used and unsuitable studies eliminated on inspection. The major reasons for excluding studies were; (i) that the study species did not meet the life history criteria outlined above, (ii) a lack of usable geographic data, (iii) no reported pairwise F_{ST}s, and (iv) that highly divergent mitochondrial clades were present and that these may represent separate species being included within a data set. The frequency with which each of these reasons was invoked was not recorded. For the host specificity levels of monophagy, oligophagy, and polyphagy, the Order to which the respective insect species in each category belonged was also recorded, as well as data relating to the geographical distances for those categories.

The number of studies that might be available for a comprehensive future analysis can be estimated using the ratio of suitable studies obtained from the search to the number of database entries that were reviewed. The total number of entries returned from the Web of Science search was 4449 (27/03/2017) and a further 50% should be found in other major databases (Mongeon & Paul-Hus 2016; Nakagawa *et al.* 2017). I scrutinized 633 of the papers that the search returned, and 18 studies fit the criteria for inclusion in the analysis. A final estimate of 187 studies is obtained when this value is multiplied by the ratio of suitable to unsuitable studies (~2.8%). Mitochondrial sequence data were available together with microsatellite data in 8 of these 16 studies, giving an estimate of 94 studies for which an analysis of such paired data would be possible. It is also likely that the ratio of database entries to suitable studies will decrease as earlier years of the database are searched, and so these estimates may be slightly high.

Table 6.1. Results from ANOVA analyses of the relationship between host specificity indices at different taxonomic levels (of host plant), and various independent variables, including IBD Slope, IBD R^2, mean pairwise F_{ST} (labelled as mean F_{ST}), and highest pairwise F_{ST} (labelled as highest F_{ST}). Mean pairwise geographical distance (labelled as geo. distance) was included alongside each other independent variable in each ANOVA, as was the interaction between them. Degrees of freedom were one in all cases, and the sum of squares, F value, and p value are reported for the ANOVAs. For residuals, mean of squares is shown in brackets. Significance levels for p values are '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.10 ' '. Significant and near significant values are shown in bold. Eta squared is used as a measure of effect size.

Independent variables	**Genus**				**Family**				**Order**				**Mean**			
	Sum sq.	F value	p value	Eta sq.	Sum sq.	F value	p value	Eta sq.	Sum sq.	F value	p value	Eta sq.	Sum sq.	F value	p value	Eta sq.
IBD slope	0.70	3.50	**0.08.**	0.18	0.44	2.08	0.17	0.12	0.48	2.23	0.15	0.13	0.59	2.85	0.11	0.00
Geo. distance	0.01	0.06	0.80	0.00	0.06	0.28	0.60	0.02	0.07	0.31	0.59	0.02	0.05	0.23	0.64	0.00
Interaction	0.02	0.12	0.73	0.01	0.03	0.13	0.72	0.01	0.02	0.11	0.74	3.39	0.00	0.00	1.00	0.00
Residuals	3.22 (0.20)			0.81	3.43 (0.21)			0.87	3.39 (0.21)			0.86	3.32 (0.21)			0.00
IBD R^2	0.03	1.07	0.32	0.06	0.00	0.02	0.91	0.00	0.00	0.00	0.96	0.00	0.00	0.12	0.73	0.01
Geo. distance	0.03	1.02	0.33	0.05	0.03	1.03	0.33	0.05	0.03	1.12	0.31	0.06	0.03	1.13	0.30	0.06
Interaction	0.02	0.62	0.44	0.03	0.07	2.39	0.14	0.12	0.06	2.40	0.14	0.12	0.04	1.56	0.23	0.08
Residuals	0.44 (0.03)			0.86	0.43 (0.03)			0.82	0.43 (0.03)			0.82	0.44 (0.03)			0.85
Mean F_{ST}	0.03	3.62	**0.08.**	0.18	0.03	3.54	**0.08.**	0.19	0.03	4.27	**0.06.**	0.22	0.03	4.22	**0.06.**	0.21
Geo. distance	0.00	0.03	0.85	0.00	0.00	0.31	0.59	0.02	0.00	0.38	0.55	0.02	0.00	0.22	0.65	0.01
Interaction	0.00	0.10	0.76	0.01	0.01	0.62	0.44	0.03	0.00	0.64	0.44	0.03	0.00	0.28	0.61	0.01
Residuals	0.13 (0.01)			0.81	0.13 (0.01)			0.78	0.12 (0.01)			0.75	0.13 (0.01)			0.77
Highest F_{ST}	0.16	3.65	**0.07.**	0.18	0.13	2.99	0.10	0.16	0.16	3.97	**0.06.**	0.20	0.16	3.89	**0.07.**	0.20
Geo. distance	0.03	0.08	0.79	0.00	0.02	0.38	0.55	0.02	0.02	0.49	0.49	0.02	0.01	0.31	0.59	0.01
Interaction	0.01	0.11	0.75	0.01	0.04	0.98	0.34	0.05	0.04	0.93	0.35	0.04	0.02	0.43	0.52	0.02
Residuals	0.70 (0.04)			0.81	0.69 (0.04)			0.79	0.65 (0.04)			0.75	0.68 (0.04)			0.78

6.3.2 ANOVA results

Among all ANOVA results (Table 6.1), the most significant values were obtained when host specificity was compared with the mean pairwise F_{ST} and highest pairwise F_{ST} as the genetic independent variables. P values ranged from 0.06 to 0.8 for mean pairwise F_{ST} and 0.06 to 0.10 for highest pairwise F_{ST}. All values were not significant, but only marginally so ($p = 0.06$ and 0.07 respectively). Effect sizes (Eta squared) were moderate for both mean pairwise F_{ST} (0.18-0.22) and highest pairwise F_{ST} (0.16-0.20) ANOVA. The IBD slope was near significant ($p = 0.08$) only at the level of genus and with a moderate eta squared of 0.18. IBD R^2 was not significant in any ANOVA. Geographic distance, and its interaction with each of the genetic predictors, was non-significant in each instance. While all results are not significant, the outcome is promising given the relatively low number of studies and the consistently low significance values for both F_{ST} measures across all levels of weighted host specificity.

6.4 Discussion

6.4.1 Host breadth, host distribution, and genetic differentiation

The results presented above suggest that the genetic structure of phytophagous insect populations is meaningfully correlated with their host breadth. Relatively specialised insects have more genetically differentiated populations than those with a broader host range (Table 6.1). This result was found when considering mean pairwise F_{ST} (average population differentiation) and highest pairwise F_{ST} (the two most genetically distinct populations from a given study). Specialist phytophagous insect species thus have lower population genetic connectivity between all populations, and as well as more highly differentiated populations overall. Counter to the predictions of this study, the significant relationship between host breadth and genetic differentiation was independent of absolute geographic distance (Table 6.1). The evidence for this comes from the near significant relationship between almost all weighted host breadth values and both pairwise F_{ST} values, compared with the lack of significance between almost all weighted host breadth values and IBD slope, and so too for that with IBD R^2 (Table 6.1). Geographic distance and its interaction with each genetic variable were also never significant correlated with host breadth (Table 6.1). The weighted index of specificity at the host genus level was the only value that had a near significant relationship with IBD slope (Table 6.1). Genetic differentiation for monophagous insects (those host specific at the level of genus) thus increases at a faster rate compared with polyphagous and oligophagous insects. Again though, this is independent of absolute geographic distance.

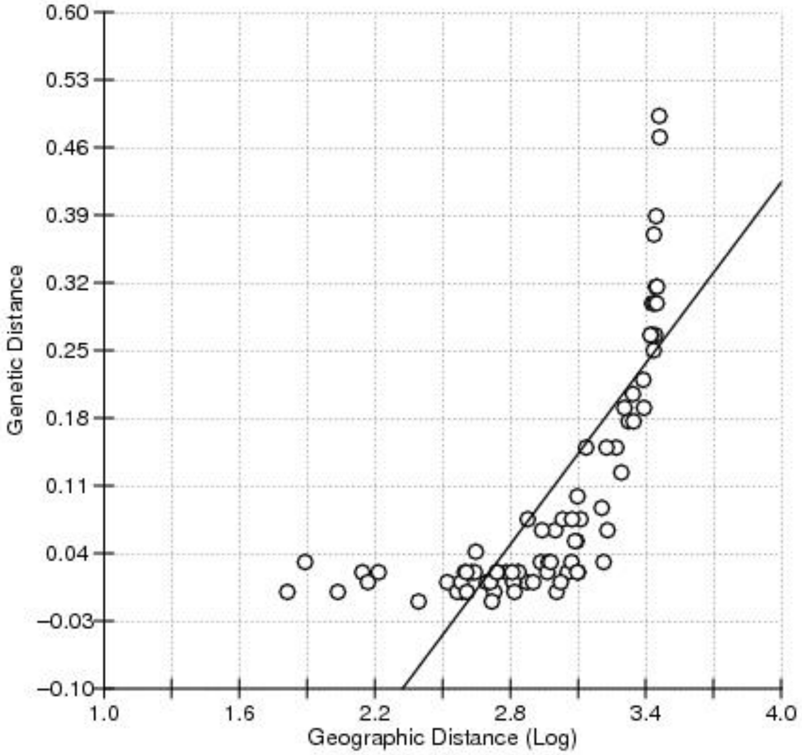
The apparent importance of outlier populations, as indicated by the near significance of highest pairwise F_{ST} values at three levels of host specificity (Table 6.1), raises questions about the relationship between these outliers and the other populations in each study. How long have these outliers been separated from other populations, and does gene flow still occur between the most genetically differentiated populations and the other populations? Further, even if there is no gene flow, it may be that these outlier populations have only recently become isolated from the main body of the species gene pool, but also possible is that unrecognized cryptic species could be included together in one data set. Indeed, *N. viridula* (Chapter 2 and 3) and *C. chadwicki* (Chapter 4 and 5) illustrate how complexities in the relationship between geographic and genetic distance could cause problems if data were used uncritically. Unless species-specific biological information can be used to make a distinction between 1) a pattern of geographic isolation, and 2) a pattern of cryptic species diversity, the outlier populations from different studies cannot be compared equally. In the approach taken here, the scatter of IBD plots was used to determine the relationship between populations in each study, and so investigating each IBD plot individually may provide further understanding.

In some cases, such as with the IBD plot for *Bombus morio* (Francisco *et al.* 2016), the linear IBD relationship does not at all match the scatter of the IBD plot (Figure S6.1). In cases such as this it is easy to justify removing the outliers, as including them confounds any attempt to meaningfully assess the IBD relationship. The same pattern is not evident for *N. viridula* and *C. chadwicki,* where the overall IBD plots appear to be relatively continuous whether all populations are included (Figures 6.1a and 6.2a) or if only data of genotyped individuals from a single putative species is included (*N. viridula* E, Figure 6.1b and *C. chadwicki* sp. 4, Figure 6.2b). A similar pattern of outliers with more subtle differentiation can be seen in the IBD plots of *Graellsia isabellae* (Marí-Mena *et al.* 2016) and *Rosalia alpina* (Drag *et al.* 2015) (Figure S6.2). Of the four species just mentioned, it is only for the group of *C. chadwicki* that are referred to as the central clade thrips (and which includes *C. chadwicki* sp. 4 and some other species, Chapter 5) that there is strong evidence for these genetically distinct populations representing distinct species (molecular data from Chapter 4, and pollinator-plant volatile data from L. I. Terry unpublished). However, this is not evident from this genetic data alone.

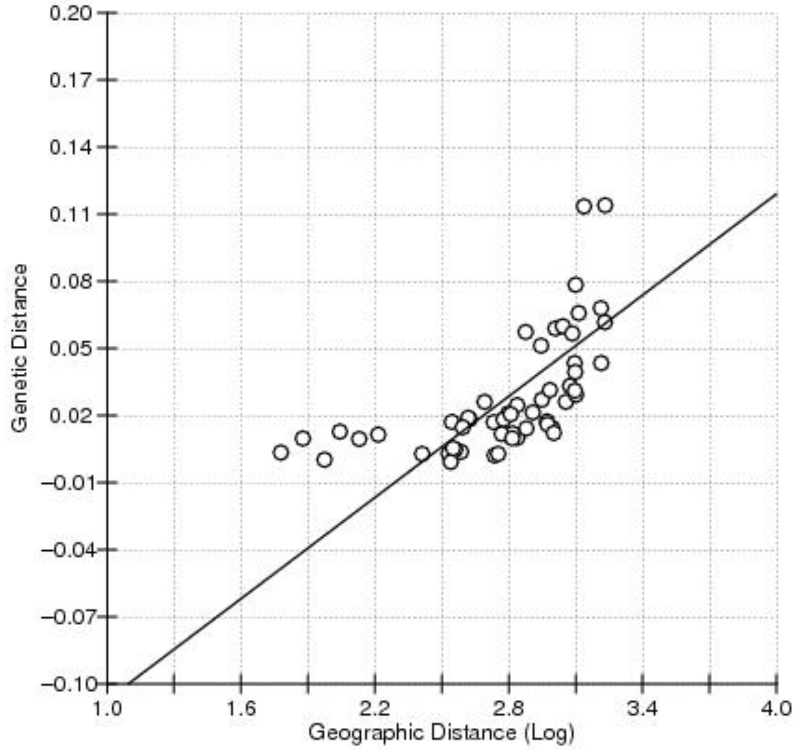
That cryptic species diversity might so easily go unnoticed in IBD plots is a crucial point, and shows to some degree, that the processes that led to highly genetically differentiated populations in the individual studies may not be the same across studies. Whether the most genetically differentiated populations represent distinct species or whether they are only geographically isolated populations of a single species cannot be determined using this approach. Further, without intimate knowledge of the biology of each system this will not become clear. These insights do provide a rationale for investigating the species status of these genetically distinct populations in more depth, but these situations impede, to some degree, the ability to make generalisations using data from such systems. Using mean pairwise F_{ST} may minimize the effects of cryptic diversity on the data, but only meaningfully so if one species is in the minority of populations sampled. Clearly, population genetics studies that cover multiple populations and localities need to consider the species status of highly divergent populations. The lack of such information not only affects interpretation within a system, but hinders the development of realistic generalisations when incorporated into larger analyses. This is however difficult, and has not even been fully resolved for *N. viridula* (Chapter 2). For an analysis such as the one proposed here to be effective, it must be able to incorporate data where species status is uncertain and this is discussed further below.

6.4.2 Refinement of the approach

The analysis presented in this chapter provides a framework from which a comprehensive analysis of the relationship between insect host breadth and population genetic differentiation can be conducted. The main issue with the approach taken here is that the species status of the populations in each study is uncertain. For *N. viridula* and *C. chadwicki* (and many other species), using mitochondrial DNA sequences alongside microsatellite genotype data has allowed for the estimation of the timing of the divergence of populations when there is presently no gene flow among them. A large number of studies have mtDNA sequence data as well as microsatellite genotype data, so a future analysis should incorporate both sets of data (Avise *et al.* 1987). The pairwise distance between mtDNA haplotypes (%) is explored for insect species where these data were available in the studies included in the analysis of Table 6.2. While this does not resolve the questions of species diversity, it does allow for estimation of the relatedness of populations and so populations with similar levels of divergence can be compared. The case with *N. viridula* highlights that even mtDNA divergence can be misleading with regard to the genetic relationship between populations, but estimates exist for the rate of occurrence of nuclear-mitochondrial discordance (Toews and Brelsford 2012) and authors are likely to have detect such occurrences in the investigation of their systems.

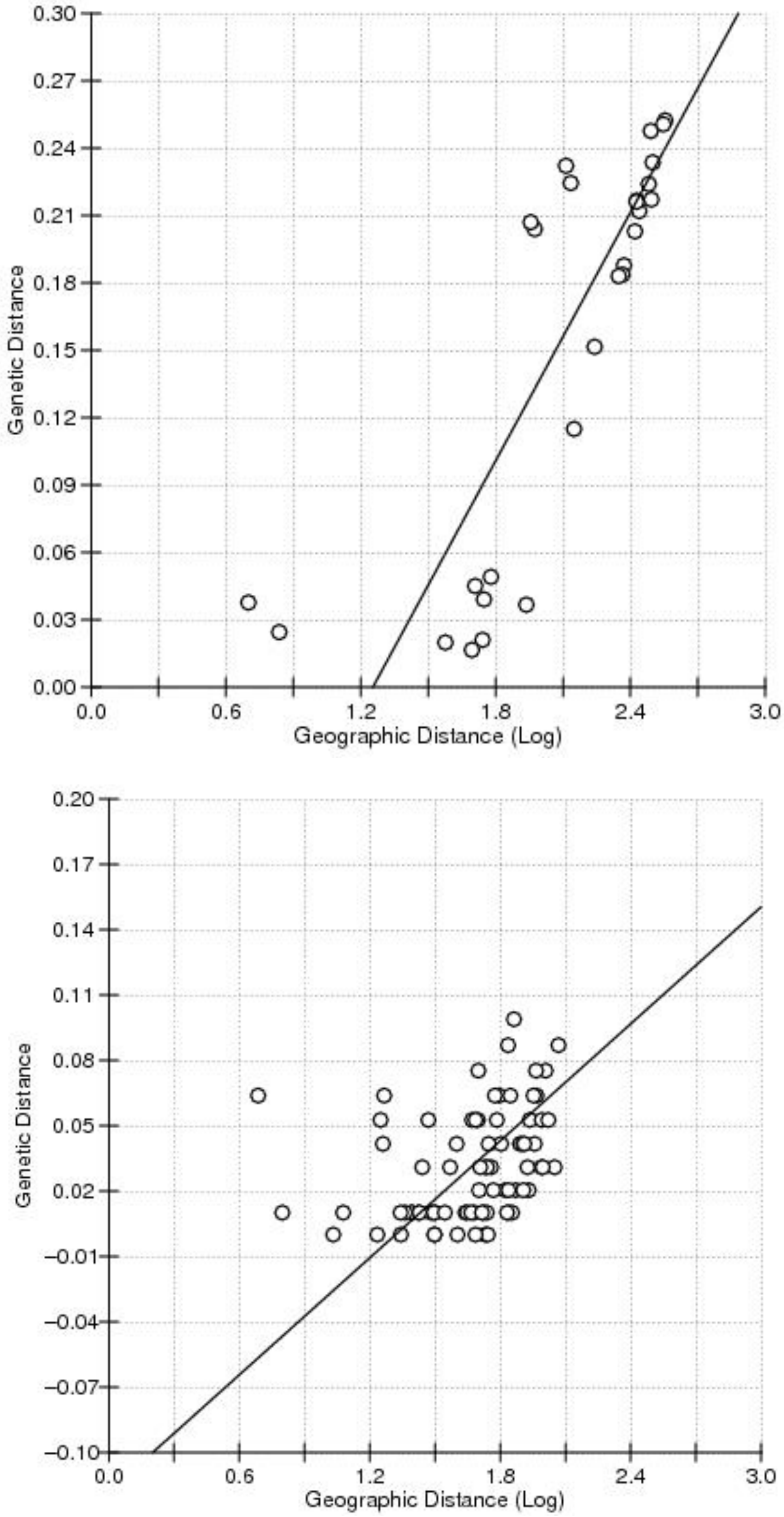


a)



b)

Figure 6.1. IBD relationships (from IBDWS) for *Nezara viridula* are shown. All Australian *N. viridula* populations are included in the top figure (a) (see Chapter 2) and the eastern Australian populations are included below (b) (this is Figure 3.2b - see Chapter 3). IBD R^2 values are 0.54 (a) and 0.39 (b).



a)

b)

Figure 6.2. IBD relationships (from IBDWS) for *Cycadothrips chadwicki sensu lato.* All central clade *C. chadwicki* species are included in the top figure (a) (those from Chapter 4) and individuals of *C. chadwicki* sp. 4 are included alone in the figure below (b) (this is Figure 5.3a - see Chapter 5). IBD R^2 values are 0.66 (a) and 0.12 (b).

The three species with highest pairwise F_{ST}s were *Graellsia isabellae* (0.70 – monophagous), *Rosalia alpina* (0.44 - polyphagous), and *Grapholita molesta* (0.68 – polyphagous). For these three species, the highest pairwise distance between mtDNA haplotypes was 1.3%, 1.2%, and 1.3% respectively. These values are low enough, compared with the typical divergence between species (Hebert *et al.* 2003), that the populations of each of these insects are most likely con-specific. *Rhynchophorus ferrugineus*, by contrast, had the highest mtDNA pairwise distance (4.8%) and has since been revealed to comprise at least two species (Rugman-Jones *et al.* 2013), though this is uncertain for the individuals genotyped in the study included in Table 6.1 (Wang *et al.* 2015). The most important aspect of such a combined approach (highest pairwise F_{ST} and mtDNA pairwise distance) is that it makes no judgement about the species status of the genetically distinct groups. Even if the populations do differ in some respects, as for example might the sub-species of *G. isabellae* (Marí-Mena *et al.* 2016), then this species information could be included in an analysis rather than be used to justify excluding populations for which not enough biological information exists.

6.5 Conclusions

In a previous attempt to address similar questions, Peterson and Denno (1998) found no significant relationship between the host breadth of insects and their population genetic structure. The advantages of the approach I took, and perhaps the reason for the different result, are illustrated with the following examples. i) The inclusion of weighted measures of host specificity reduces the error in assessing host use. The two species *Thaumetopoea pityocampa* and *Plutella xylostella* would have had their host specificity over-generalized with an approach that considered only the taxonomic classification of their two most distantly related host plants (Table S6.2). ii) Analysing multiple genetic measures that characterise different aspects of the genetic relationship between populations has shown how different the results they produce can be. That the analyses that use IBD slope and IBD R^2, compared with mean pairwise F_{ST} and highest pairwise F_{ST}, produce different results suggests that the choice of genetic measures requires good justification with respect to the aims of such analyses.

In summary, the genetic connectivity of herbivorous insect populations is related to their host specificity, and this has important consequences for understanding the evolution of these systems. The interaction between the host specificity of insects, changes in host plant and insect distributions, and the genetic connectivity of insect populations may explain some observations relating to the evolution of insect-plant interactions. If populations of highly specialised insects are more easily fragmented, and so separated from one another more frequently, then they are also likely to have higher rates of speciation. This may help explain why specialist insect species are more abundant than generalists. That specialist insects rely on so few host plant species may also mean that their host use is of more consequence to their associated species diversity, as appears to be the case with *Cycadothrips* (compared with *N. viridula*). In future analyses, the inclusion of aspects of insect feeding behaviour, and other features that relate to host specificity, may also help to explain why some insect groups are so much more speciose than others.

Table 6.2. For studies that had mitochondrial sequence data together with microsatellite data the following information is shown - species name and host breadth (in square brackets as either M – monophagous, O – oligophagous, and P – polyphagous), highest pairwise F_{ST}, highest pairwise distance between mtDNA haplotypes (%), and gene region sequenced. Highest mtDNA pairwise distance was obtained manually using sequences from Genbank that were associated with each study. Sequences were aligned using MAFFT and then the highest distance between any two haplotypes was calculated.

Species	**Highest pairwise F_{ST}**	**mtDNA seq. distance (%)**	**Gene region**
Thaumetopoea pityocampa [O]	0.23	1.9	COII
Graellsia isabellae [M]	0.70	1.3	COI
Bombus morio (all) [P]	0.20	3.1	COI
Bombus morio (outliers removed) [P]	0.04	2.0	COI
Rosalia alpina [P]	0.44	1.2	COI
Grapholita molesta [O]	0.68	1.3	COI
Prodoxus decipiens [M]	0.28	0.9[a]	COI
Nezara viridula (all Australia) [P]	0.33	3.0	COI
Nezara viridula (east Australia) [P]	0.12	0.0[b]	COI
Cycadothrips chadwicki (central clade) [M]	0.09	2.0	COI
Cycadothrips chadwicki sp. 4 [M]	0.20	1.1	COI
Trioza barrettae [M]	0.23	1.5	COI
Rhynchophorus ferrugineus [O]	0.31	4.8	COI

[a] Obtained from sequences of Althoff *et al.* (2002).

[b] Excluding mtDNA haplotypes that have resulted from mitochondrial introgression (Chapter 2).

6.6 Supporting information

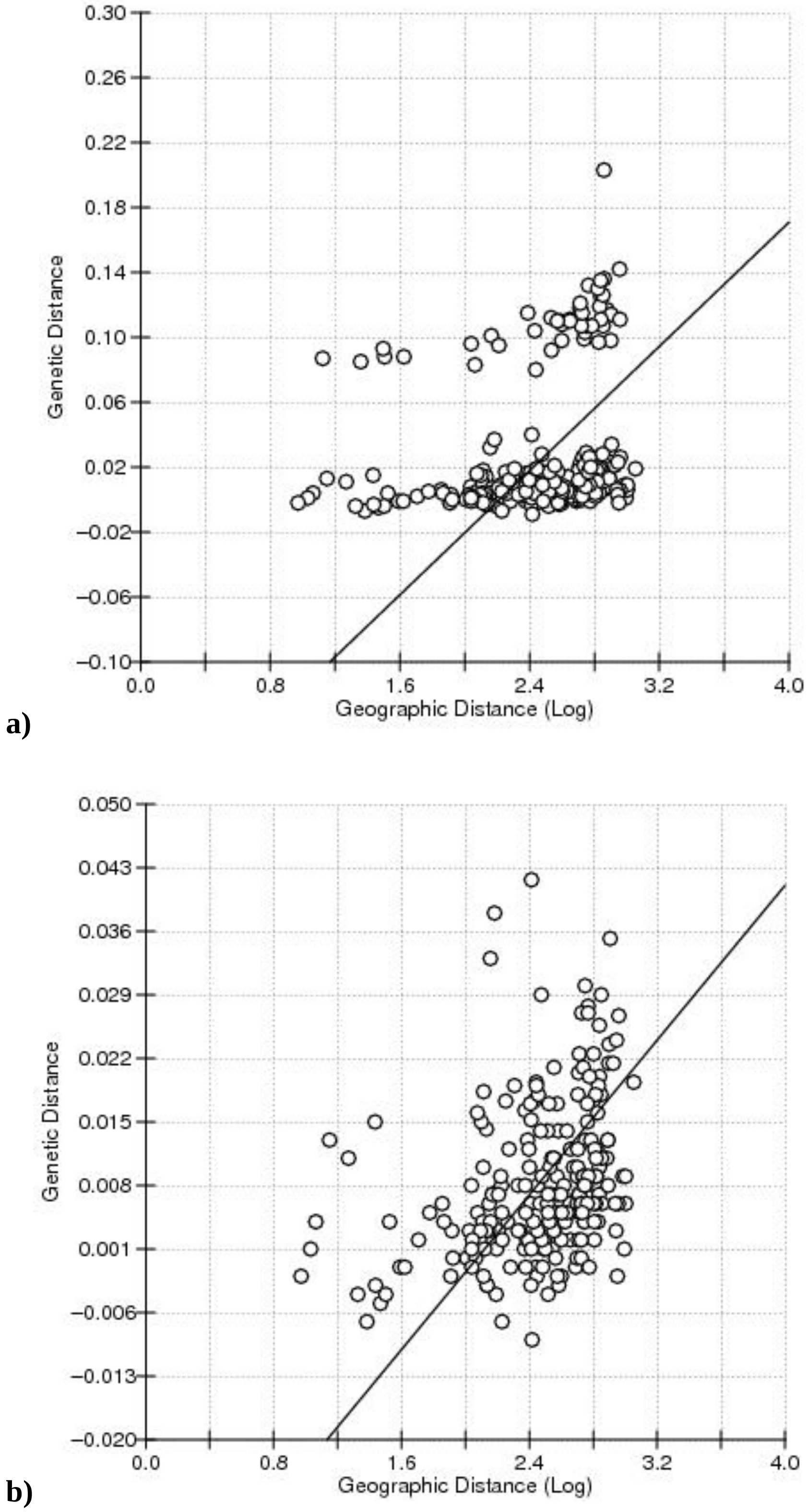
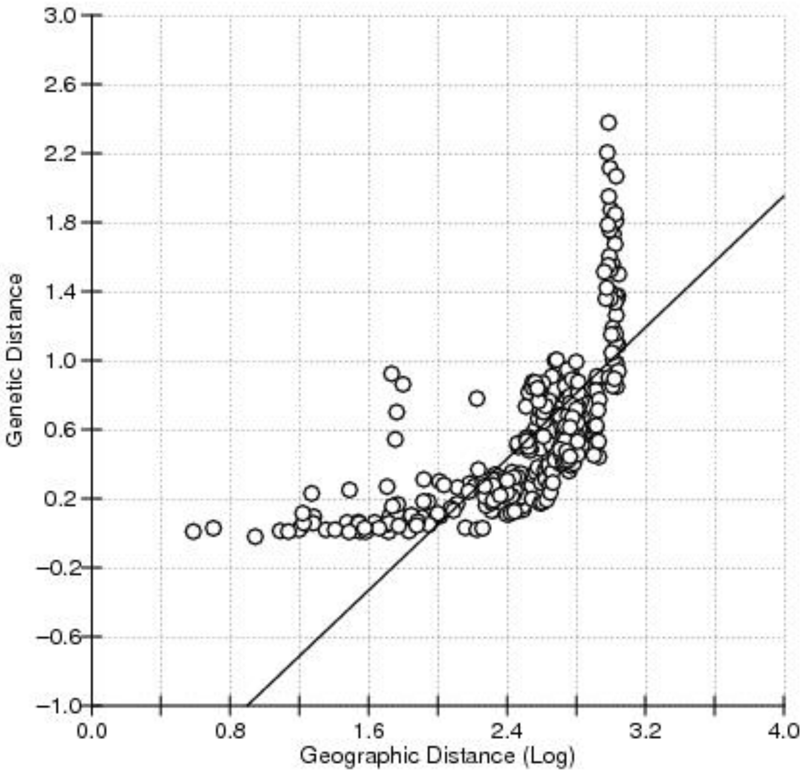
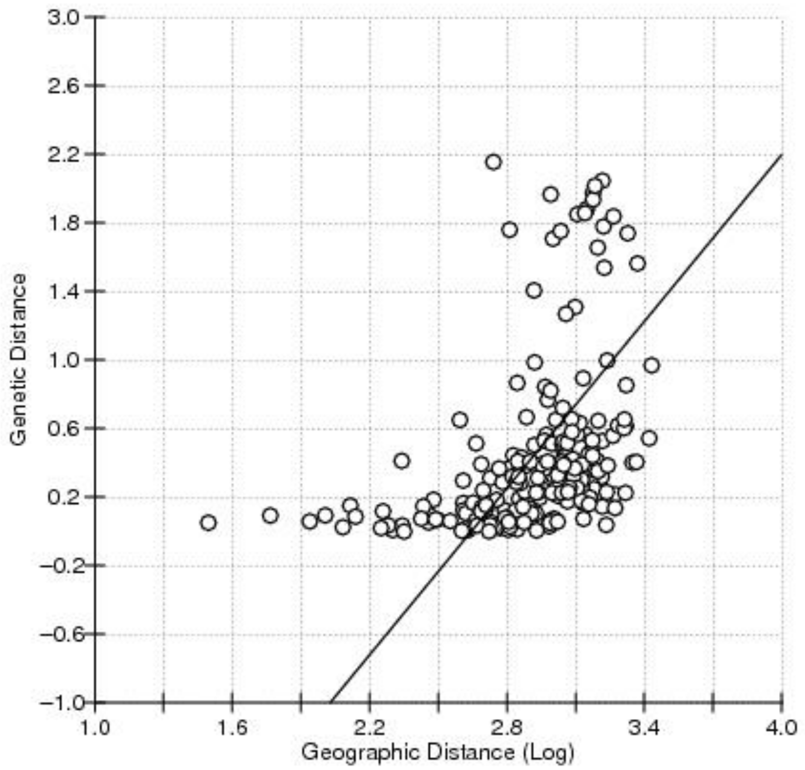


Figure S6.1. IBD relationship (from IBDWS) for *Bombus morio* (Francisco *et al.* 2016) prior to the populations IVIT and TSAM being removed in the top figure (a) and after their removal below (b).



a)



b)

Figure S6.2. IBD relationships (from IBDWS) for *Graellsia isabellae* (Marí-Mena *et al.* 2016) in the top figure (a), and for *Rosalia alpina* (Drag *et al.* 2015) below (b).

Table S6.1. Details of the studies included in the analysis presented here that were found using the Web of Science database. Study species, study authors, Digital Object Identifier (DOI), and notes relevant to how the data from each study was incorporated into the data set presented here.

Species	**Authors**	**DOI**	**Study Notes**
Thaumetopoea pityocampa	Kerdelhué *et al.* (2016)	10.1002/ece3.2194	Only the populations labelled as ‘macro’ in the study were used.
Dendroctonus valens	Taerum *et al.* (2016)	10.1111/bij.12781	Invasive Chinese populations were removed.
Graellsia isabellae	Marí-Mena *et al.* (2016)	10.1186/s12862-016-0708-y	None.
Bombus morio	Francisco *et al.* (2016)	10.1007/s10841-016-9872-z	Two highly divergent populations, IVIT and TSAM, were removed.
Plutella xylostella	Ke *et al.* (2015)	10.1002/ece3.1850	None.
Rosalia alpina	Drag *et al.* (2015)	10.1111/bij.12624	None.
Lobesia botrana	Reineke *et al.* (2015)	10.1017/S0007485315000267	Author provided geographic coordinates and verified error in one F_{ST} value.
Grapholita molesta	Wei *et al.* (2015)	10.1111/mec.13300	None.
Aleurocanthus spiniferus	Wang *et al.* (2015)	10.1016/j.gene.2015.01.050	None.
Prodoxus decipiens	Darwell *et al.* (2014)	10.1111/jeb.12529	Island populations excluded.
Matsucoccus feytaudi	Kerdelhué *et al.* (2014)	10.1038/hdy.2014.39	Gargas (temporal outlier) and Corsica (island) populations removed.
Bactrocera dorsalis	Aketarawong *et al.* (2014)	10.1186/1471-2156-15-70	F_{ST}s from the same site were averaged. Only *B. papayae* populations used.
Episyrphus balteatus	Raymond *et al.* (2013)	10.1111/mec.12483	Poor host records. Weighted host specificity based on host plant order only.
Sphaerophoria scripta	Raymond *et al.* (2013)	10.1111/mec.12483	Poor host records. Weighted host specificity based on host plant order only.
Trioza barrettae	Moir *et al.* (2016)	10.1016/j.biocon.2016.04.012	None.
Ramburiella hispanica	Ortego *et al.* (2015)	10.1111/eva.12273	None.
Rhynchophorus ferrugineus	Wang *et al.* (2015)	10.1111/eea.12282	None.
Anastrepha fraterculus	Manni *et al.* (2015)	10.3897/zookeys.540.6713	None.

Table S6.2. Columns G, F, O, and M, are weighted host specificity values for Genus, Family, Order, and the Mean of these three weighted host specificity values respectively. The first time the weighted host specificity of each insect passes 80% at a taxonomic level it is considered a specialist at that level. Host use is designated into monophagous (host specific at the level of Genus), oligophagous (host specific at the level of Family), or polyphagous (not host specific at the levels Genus or Family) according to weighted host specificity. Mean pairwise FST and Highest pairwise FST are shown without transformations.

Species	**G**	**F**	**O**	**M**	**Host use**	**IBD Slope**	**IBD R^2**	**Mean F_{ST}**	**Highest F_{ST}**	**Mean dist. (km)**	**No. pops.**
Thaumetopoea pityocampa	0.71	0.93	0.93	0.86	Oligophagous	0.19	0.22	0.12	0.23	285	16
Dendroctonus valens	1.00	1.00	1.00	1.00	Monophagous	0.27	0.55	0.13	0.33	2413	14
Graellsia isabellae	1.00	1.00	1.00	1.00	Monophagous	0.95	0.42	0.31	0.70	447	26
Bombus morio	0.14	0.24	0.24	0.21	Polyphagous	0.02	0.11	0.01	0.04	390	22
Plutella xylostella	0.24	0.81	0.86	0.63	Oligophagous	0.10	0.06	0.02	0.09	239	9
Rosalia alpina	0.25	0.25	0.50	0.33	Polyphagous	0.33	0.41	0.13	0.44	585	32
Lobesia botrana	0.17	0.29	0.44	0.30	Polyphagous	0.08	0.17	0.08	0.20	1473	14
Grapholita molesta	0.57	1.00	1.00	0.86	Oligophagous	1.62	0.16	0.24	0.68	992	21
Aleurocanthus spiniferus	0.67	1.00	1.00	0.89	Oligophagous	-0.11	0.01	0.05	0.09	902	10
Prodoxus decipiens	1.00	1.00	1.00	1.00	Monophagous	0.18	0.05	0.11	0.28	1064	18
Matsucoccus feytaudi	1.00	1.00	1.00	1.00	Monophagous	1.06	0.23	0.30	0.63	852	12
Bactrocera dorsalis	0.04	0.09	0.17	0.10	Polyphagous	-0.12	0.04	0.10	0.17	531	12
Episyrphus balteatus	0.33	0.33	0.33	0.33	Polyphagous	0.04	0.04	0.01	0.03	147	13
Sphaerophoria scripta	0.33	0.33	0.33	0.33	Polyphagous	-0.02	0.00	0.00	0.03	154	12
Nezara viridula	0.07	0.22	0.22	0.17	Polyphagous	0.08	0.39	0.03	0.13	574	21
Cycadothrips chadwicki sp. 4	1.00	1.00	1.00	1.00	Monophagous	0.09	0.12	0.03	0.09	55	18
Trioza barrettae	1.00	1.00	1.00	1.00	Monophagous	0.33	0.28	0.08	0.23	48	5
Ramburiella hispanica	0.50	1.00	1.00	0.83	Oligophagous	0.03	0.02	0.02	0.05	28	18
Rhynchophorus ferrugineus	0.22	1.00	1.00	0.74	Oligophagous	-0.31	0.00	0.18	0.31	488	12
Anastrepha fraterculus	0.09	0.18	0.23	0.17	Polyphagous	0.21	0.30	0.09	0.18	1244	6

Table S6.3. The species from each study are classified into monophagous, oligophagous, or polyphagous and comparison made between them. Mean geographic distance, and its standard deviation, are shown for the included studies. Mean distance is calculated a second time after excluding studies that exist outside the standard deviation of geographic distance. The Order to which insects from each host breadth category belong is shown, with the number of species from each Order shown in brackets.

Host breadth	**Mean dist. (km)**	**Std. dev. (km)**	**Mean dist. (km) w/o outliers**	**Insect Orders (count)**
Monophagous	824	874	506	Coleoptera (1), Hemiptera (2), Lepidoptera (2), Thysanoptera (1)
Oligophagous	489	385	260	Coleoptera (1), Hemiptera (1), Lepidoptera (3), Orthoptera (1)
Polyphagous	1115	524	1034	Coleoptera (1), Diptera (4), Hemiptera (1), Hymenoptera (1), Lepidoptera (1)

References

Afreh-Nuamah K (1983) Insect pests associated with winged bean in Ghana. *Annals of Tropical Research* **5**, 1-5.

Aketarawong N, Isasawin S, Thanaphum S (2014) Evidence of weak genetic structure and recent gene flow between *Bactrocera dorsalis* ss and *B. papayae*, across Southern Thailand and West Malaysia, supporting a single target pest for SIT applications. *BMC Genetics* **15**, 70.

Aldrich J, Oliver J, Lusby W, Kochansky J, Lockwood J (1987) Pheromone strains of the cosmopolitan pest, *Nezara viridula* (Heteroptera: Pentatomidae). *Journal of Experimental Zoology* **244**, 171-175.

Aldrich J (1990) Dispersal of the southern green stink bug, Nezara viridula (L.)(Heteroptera: Pentatomidae), by hurricane Hugo. *Proceedings of the Entomological Society of Washington* **92**, 757-759.

Ali JG, Agrawal AA (2012) Specialist versus generalist insect herbivores and plant defense. *Trends in Plant Science* **17**, 293-302.

Ali M, Ewiess M (1977) Photoperiodic and temperature effects on rate of development and diapause in the green stink bug, *Nezara viridula* L.(Heteroptera: Pentatomidae). *Zeitschrift für angewandte Entomologie* **84**, 256-264.

Althoff DM, Pellmyr O, Bernatchez L (2002) Examining genetic structure in a bogus yucca moth: a sequential approach to phylogeography. *Evolution* **56**, 1632-1643.

Althoff DM, Segraves KA, Johnson MT (2014) Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology & Evolution* **29**, 82-89.

Althoff DM, Segraves KA, Smith CI, Leebens-Mack J, Pellmyr O (2012) Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. *Molecular Phylogenetics and Evolution* **62**, 898-906.

Anderson C, Tay WT, McGaughran A, Gordon K, Walsh TK (2016) Population structure and gene flow in the global pest, *Helicoverpa armigera*. *Molecular Ecology* **25**, 5296-5311.

Anderson RS (1993) Weevils and plants: phylogenetic versus ecological mediation of evolution of host plant associations in Curculioninae (Coleoptera: Curculionidae). *Memoirs of the Entomological Society of Canada* **125**, 197-232.

Anstett MC, Hossaert-McKey M, Kjellberg F (1997) Figs and fig pollinators: evolutionary conflicts in a coevoled mutualism. *Trends in Ecology & Evolution* **12**, 94-99.

Avise JC (1994) Molecular markers, natural history and evolution. *Chapman and Hall*, New York.

Avise JC, Arnold J, Ball RM, *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155-165.

Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology* **47**, 773-815.

Bernays E (2001) Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annual Review of Entomology* **46**, 703-727.

Blackburn TM, Pyšek P, Bacher S, *et al.* (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* **26**, 333-339.

Blows MW, McGuigan K (2015) The distribution of genetic variance across phenotypic space and the response to selection. *Molecular Ecology* **24**, 2056-2072.

Bossart J, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution* **13**, 202-206.

Brookes D, Hereward J, Terry L, Walter G (2015) Evolutionary dynamics of a cycad obligate pollination mutualism–Pattern and process in extant *Macrozamia* cycads and their specialist thrips pollinators. *Molecular Phylogenetics and Evolution* **93**, 83-93.

Brookes DR, Hereward JP (2013) Characterisation of 18 microsatellite markers for the thrips *Cycadothrips chadwicki* (Thysanoptera: Aeolothripidae), an obligate mutualist pollinator of 11 *Macrozamia* cycad species. *Conservation Genetics Resources* **5**, 1167-1170.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004-1006, 1008-1010.

Buckman RS, Mound LA, Whiting MF (2013) Phylogeny of thrips (Insecta: Thysanoptera) based on five molecular loci. *Systematic Entomology* **38**, 123-133.

Byrne M, Steane DA, Joseph L, *et al.* (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography* **38**, 1635-1656.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.

Cho S, Mitchell A, Regier JC, *et al.* (1995) A highly conserved nuclear gene for low-level phylogenetics: elongation factor-1 alpha recovers morphology-based tree for heliothine moths. *Molecular Biology and Evolution* **12**, 650-656.

Clarke AR (1992) Current distribution and pest status of *Nezara viridula* (L.)(Hemiptera: Pentatomidae) in Australia. *Australian Journal of Entomology* **31**, 289-297.

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

Čokl A, McBrien HL, Millar JG (2001) Comparison of substrate-borne vibrational signals of two stink bug species, *Acrosternum hilare* and *Nezara viridula* (Heteroptera: Pentatomidae). *Annals of the Entomological Society of America* **94**, 471-479.

Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24**, 1999-2017.

Cook JM, Segar ST (2010) Speciation in fig wasps. *Ecological Entomology* **35**, 54-66.

Crisp MD, Cook LG (2011) Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. *New Phytologist* **192**, 997-1009.

Cristescu ME (2015) Genetic reconstructions of invasion history. *Molecular Ecology* **24**, 2212-2225.

Crosiers C, Malaisse FP (1995) *Ex situ* pollination and multiplication of *Encephalartos laurentianus* De Wild.(Zamiaceae, Cycadales). *Biodiversity and Conservation* **4**, 767-775.

Cruaud A, Rønsted N, Chantarasuwan B, *et al.* (2012) An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. *Systematic Biology* **61**, 1029-1047.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772-772.

Darwell C, Fox K, Althoff D (2014) The roles of geography and founder effects in promoting host-associated differentiation in the generalist bogus yucca moth *Prodoxus decipiens*. *Journal of Evolutionary Biology* **27**, 2706-2718.

Devictor V, Julliard R, Couvet D, Jiguet F (2008) Birds are tracking climate warming, but not fast enough. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2743-2748.

Dhami MK, Dsouza M, Waite DW, Anderson D, Li D (2016) Real-time PCR assay for the identification of the brown marmorated stink bug (*Halyomorpha halys*). *Frontiers in molecular biosciences* **3**.

Dinsdale A, Cook L, Riginos C, Buckley Y, De Barro P (2010) Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America* **103**, 196-208.

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology* **24**, 2095-2111.

Donaldson JS (2003) *Cycads - status survey and conservation action plan.* IUCN/SSC Cycad Specialist Group, IUCN, Gland, Switzerland and Cambridge, UK.

Downie DA, Donaldson JS, Oberprieler RG (2008) Molecular systematics and evolution in an African cycad-weevil interaction: Amorphocerini (Coleoptera: Curculionidae: Molytinae) weevils on *Encephalartos*. *Molecular Phylogenetics and Evolution* **47**, 102-116.

Drag L, Hauck D, Bérces S, *et al.* (2015) Genetic differentiation of populations of the threatened saproxylic beetle Rosalia longicorn, *Rosalia alpina* (Coleoptera: Cerambycidae) in Central and South-east Europe. *Biological Journal of the Linnean Society* **116**, 911-925.

Drès M, Mallet J (2002) Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **357**, 471-492.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences* **97**, 9115-9120.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Edwards DL, Melville J (2010) Phylogeographic analysis detects congruent biogeographic patterns between a woodland agamid and Australian wet tropics taxa despite disparate evolutionary trajectories. *Journal of Biogeography* **37**, 1543-1556.

Egas M, Dieckmann U, Sabelis MW (2004) Evolution restricts the coexistence of specialists and generalists: the role of trade-off structure. *American Naturalist* **163**, 518-531.

Ehrlich PR, Raven PH (1964) Butterflies and plants, a study in coevolution. *Evolution* **18**, 586-608.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435-445.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.

Ferrari A, Schwertner CF, Grazia J (2010) Review, cladistic analysis and biogeography of *Nezara* Amyot & Serville (Hemiptera: Pentatomidae). *Zootaxa* **2424**, 1-41.

Fitzpatrick B, Fordyce J, Gavrilets S (2009) Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology* **22**, 2342-2347.

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

Forister M, Dyer L, Singer M, Stireman J, Lill J (2012) Revisiting the evolution of ecological specialization, with emphasis on insect–plant interactions. *Ecology* **93**, 981-991.

Forister ML, Novotny V, Panorska AK, *et al.* (2015) The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 442-447.

Forster PI, Machin PJ, Mound L, Wilson GW (1994) Insects associated with reproductive structures of cycads in Queensland and northeast New South Wales, Australia. *Biotropica* **26**, 217-222.

Francisco FO, Santiago LR, Mizusawa YM, Oldroyd BP, Arias MC (2016) Genetic structure of island and mainland populations of a Neotropical bumble bee species. *Journal of Insect Conservation* **20**, 383-394.

Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences* **106**, 18054-18061.

Futuyma DJ, Mayer GC (1980) Non-allopatric speciation in animals. *Systematic Biology* **29**, 254-271.

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. *Annual Review of Ecology and Systematics* **19**, 207-233.

Gaggiotti O, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**, 1513-1520.

Gariepy T, Haye T, Zhang J (2014) A molecular diagnostic tool for the preliminary assessment of host–parasitoid associations in biological control programmes for a new invasive pest. *Molecular Ecology* **23**, 3912-3924.

Garnas JR, Auger-Rozenberg M-A, Roques A, *et al.* (2016) Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biological Invasions* **18**, 935-952.

Georghiou GP, Lagunes-Tejeda A (1991) *The occurrence of resistance to pesticides in arthropods.* FAO United Nations, Rome.

Griffith M, Calornje M, Stevenson D, *et al.* (2012) Time, place, and relationships: cycad phenology in a phylogenetic and biogeographic context. *Memoirs of the New York Botanical Garden* **106**, 59-81.

Grimaldi D, Shmakov A, Fraser N (2004) Mesozoic thrips and early evolution of the order Thysanoptera (Insecta). *Journal of Paleontology* **78**, 941-952.

Groot A, Classen A, Inglis O, *et al.* (2011) Genetic differentiation across North America in the generalist moth *Heliothis virescens* and the specialist *H. subflexa*. *Molecular Ecology* **20**, 2676-2692.

Guichoux E, Lagache L, Wagner S, *et al.* (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**, 591-611.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704.

Guillot G, Estoup A, Mortier F, Cosson JF (2005) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.

Gwiazdowski RA, Foottit RG, Maw HEL, Hebert PD (2015) The Hemiptera (Insecta) of Canada: constructing a reference library of DNA barcodes. *PLOS ONE* **10**, e0125635.

Haine ER, Martin J, Cook JM (2006) Deep mtDNA divergences indicate cryptic species in a fig-pollinating wasp. *BMC Evolutionary Biology* **6**, 83.

Hall J, Walter G (2014) Relative seed and fruit toxicity of the Australian cycads *Macrozamia miquelii* and *Cycas ophiolitica*: further evidence for a megafaunal seed dispersal syndrome in cycads, and its possible antiquity. *Journal of Chemical Ecology* **40**, 860-868.

Hall JA, Walter GH, Bergstrom DM, Machin P (2004) Pollination ecology of the Australian cycad *Lepidozamia peroffskyana* (Zamiaceae). *Australian Journal of Botany* **52**, 333-343.

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

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812-14817.

Hebert PD, Ratnasingham S, de Waard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, S96-S99.

Hembry DH, Althoff DM (2016) Diversification and coevolution in brood pollination mutualisms: Windows into the role of biotic interactions in generating biological diversity. *American Journal of Botany* **103**, 1783-1792.

Hembry DH, Kawakita A, Gurr NE, *et al.* (2013) Non-congruent colonizations and diversification in a coevolving pollination mutualism on oceanic islands. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130361.

Hembry DH, Yoder JB, Goodman KR (2014) Coevolution and the diversification of life. *American Naturalist* **184**, 425-438.

Herbert JJ, Toews MD (2012) Seasonal abundance and population structure of *Chinavia hilaris* and *Nezara viridula* (Hemiptera: Pentatomidae) in Georgia farmscapes containing corn, cotton, peanut, and soybean. *Annals of the Entomological Society of America* **105**, 582-591.

Hereward JP, Walter GH, DeBarro PJ, Lowe AJ, Riginos C (2013) Gene flow in the green mirid, *Creontiades dilutus* (Hemiptera: Miridae), across arid and agricultural environments with different host plant species. *Ecology and Evolution* **3**, 807-821.

Herre EA (1989) Coevolution of reproductive characteristics in 12 species of New World figs and their pollinator wasps. *Experientia* **45**, 637-647.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

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

Ingham JA, Forster PI, Crisp MD, Cook LG (2013) Ancient relicts or recent dispersal: how long have cycads been in central Australia? *Diversity and Distributions* **19**, 307-316.

Ivbijaro M, Bolaji O (1990) Effects of cypermethrin+ dimethoate and extracts of *Piper guineense* and *Azadirachta indica* on the pests and yield of cowpea, *Vigna unguiculata*. *The Journal of Agricultural Science* **115**, 227-231.

Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* **21**, 243-273.

Janz N (2011) Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. *Annual Review of Ecology, Evolution, and Systematics* **42**, 71-89.

Janz N, Nylin S (2008) The oscillation hypothesis of host-plant range and speciation. In: *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects.*, pp. 203-215. University of California Press.

Janz N, Nylin S, Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology* **6**, 5.

Janzen DH (1980) When is it coevolution. *Evolution* **34**, 611-612.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* **6**, 1.

Jeraj M, Walter G (1998) Vibrational communication in *Nezara viridula*: response of Slovenian and Australian bugs to one another. *Behavioural Processes* **44**, 51-58.

Jermy T (1984) Evolution of insect/host plant relationships. *American Naturalist* **124**, 609-630.

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

Jones DL, Forster PI, Sharma IK (2001) Revision of the *Macrozamia miquelii* (F. Muell.) A. DC. (Zamiaceae section *Macrozamia*) group. *Austrobaileya* **6**, 67-94.

Jones WA (1988) World review of the parasitoids of the southern green stink bug, *Nezara viridula* (L.)(Heteroptera: Pentatomidae). *Annals of the Entomological Society of America* **81**, 262-273.

Jones WA, Sullivan M (1982) Role of host plants in population dynamics of stink bug pests of soybean in South Carolina. *Environmental Entomology* **11**, 867-875.

Joseph L, Omland KE (2009) Phylogeography: its development and impact in Australo-Papuan ornithology with special reference to paraphyly in Australian birds. *Emu* **109**, 1-23.

Joshi A, Thompson JN (1995) Trade-offs and the evolution of host specialization. *Evolutionary Ecology* **9**, 82-92.

Jousselin E, van Noort S, Berry V, *et al.* (2008) One fig to bind them all: host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. *Evolution* **62**, 1777-1797.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian protein metabolism* (ed. Munro HN), pp. 21-132. Academic Press, Inc., New York.

Jung S, Duwal RK, Lee S (2011) COI barcoding of true bugs (Insecta, Heteroptera). *Molecular Ecology Resources* **11**, 266-270.

Kariya, H (1961) Effect of Temperature on the development and mortality of the southern green stink bug, *Nezara viridula*, and the oriental stink bug, *N. antennata*. *Japanese Journal of Applied Entomology and Zoology* **5**, 191-196.

Kavar T, Pavlovčič P, Sušnik S, Meglič V, Virant-Doberlet M (2006) Genetic differentiation of geographically separated populations of the southern green stink bug *Nezara viridula* (Hemiptera: Pentatomidae). *Bulletin of Entomological Research* **96**, 117-128.

Kawakita A (2010) Evolution of obligate pollination mutualism in the tribe Phyllantheae (Phyllanthaceae). *Plant Species Biology* **25**, 3-19.

Kawakita A, Kato M (2006) Assessment of the diversity and species specificity of the mutualistic association between *Epicephala* moths and *Glochidion* trees. *Molecular Ecology* **15**, 3567-3581.

Kawakita A, Kato M (2009) Repeated independent evolution of obligate pollination mutualism in the Phyllantheae-*Epicephala* association. *Proceedings of the Royal Society B: Biological Sciences* **276**, 417-426.

Kaye L, Walter G, Raghu S (2016) Patchy distribution and varied habitats of *Macrozamia lucida* cycads explained by constancy in a key environmental variable. *Australian Journal of Botany* **64**, 285-294.

Ke F, You S, He W, *et al.* (2015) Genetic differentiation of the regional *Plutella xylostella* populations across the Taiwan Strait based on identification of microsatellite markers. *Ecology and Evolution* **5**, 5880-5891.

Kelley ST, Farrell BD (1998) Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). *Evolution* **52**, 1731-1743.

Kerdelhué C, Boivin T, Burban C (2014) Contrasted invasion processes imprint the genetic structure of an invasive scale insect across southern Europe. *Heredity* **113**, 390-400.

Kerdelhué C, Burban C, Battisti A, Chakali G, Simonato M (2016) Genetic differentiation of the pine processionary moth at the southern edge of its range: contrasting patterns between mitochondrial and nuclear markers. *Ecology and Evolution* **6**, 4274-4288.

Kergoat G, Meseguer A, Jousselin E (2017) Evolution of plant–insect interactions: insights from macroevolutionary approaches in plants and herbivorous insects. *Advances in Botanical Research* **81**, 25-53.

Kershaw A (1997) A bioclimatic analysis of early to middle Miocene brown coal floras, Latrobe Valley, south-eastern Australia. *Australian Journal of Botany* **45**, 373-387.

Kershaw A, Martin H, Mason JM (1994) The Neogene: a period of transition. In: *History of the Australian vegetation: Cretaceous to Recent* (ed. Hill R), pp. 299-327. Cambridge University Press.

Khaemba B, Mutinga MJ (1982) Insect pests of sunflower (*Helianthus annuus* L.) in Kenya. *Insect Science and its Application* **3**, 281-286.

Kindler E, Arlettaz R, Heckel G (2012) Deep phylogeographic divergence and cytonuclear discordance in the grasshopper *Oedaleus decorus*. *Molecular Phylogenetics and Evolution* **65**, 695-704.

Kiritani K (1970) Studies on the adult polymorphism in the southern green stink bug, *Nezara viridula* (Hemiptera: Pentatomidae). *Researches on Population Ecology* **12**, 19-34.

Kiritani K (2011) Impacts of global warming on *Nezara viridula* and its native congeneric species. *Journal of Asia-Pacific Entomology* **14**, 221-226.

Kiritani K, Hokyo N, Yukawa J (1963) Co-existence of the two related stink bugs *Nezara viridula* and *N. Antennata* under natural conditions. *Researches on Population Ecology* **5**, 11-22.

Kishimoto-Yamada K, Kamiya K, Meleng P, *et al*. (2013) Wide host ranges of herbivorous beetles? Insights from DNA bar coding. *PLOS ONE* **8**, e74426.

Kiritani K, Sasaba T (1969) The differences in bio- and ecological characteristics between neighbouring populations in the southern green stink bug, *Nezara viridula* L. *Japanese Journal of Ecology* **19**, 177-184.

Kon M, Oe A, Numata H, Hidaka T (1988) Comparison of the mating behaviour between two sympatric species, *Nezara antennata* and *N. viridula* (Heteroptera: Pentatomidae), with special reference to sound emission. *Journal of Ethology* **6**, 91-98.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Labandeira C, Sepkoski J (1993) Insect diversity in the fossil record. *Science* **261**, 310-315.

Labandeira CC, Kvaček J, Mostovski MB (2007) Pollination drops, pollen, and insect pollination of Mesozoic gymnosperms. *Taxon* **56**, 663-695.

Lai Y, Sun F (2003) The relationship between microsatellite slippage mutation rate and the number of repeat units. *Molecular Biology and Evolution* **20**, 2123-2131.

Lee D-H (2015) Current status of research progress on the biology and management of *Halyomorpha halys* (Hemiptera: Pentatomidae) as an invasive species. *Applied Entomology and Zoology* **50**, 277-290.

Leigh JW, Bryant D (2015) popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110-1116.

Lewinsohn TM, Novotny V, Basset Y (2005) Insects on plants: diversity of herbivore assemblages revisited. *Annual Review of Ecology Evolution and Systematics* **36**, 597-620.

Li M, Liu Q, Xi L, *et al.* (2014) Testing the potential of proposed DNA barcoding markers in *Nezara virudula* and *Nezara antennata* when geographic variation and closely related species were considered. *Journal of Insect Science* **14**, 1-11.

Li M, Xi L, Fan ZH, *et al.* (2010) Phylogeographic relationships of the southern green stink bug *Nezara viridula* (Hemiptera: Pentatomidae). *Insect Science* **17**, 448-458.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Machado CA, Robbins N, Gilbert MT, Herre EA (2005) Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences of the United States of America* **102 Suppl 1**, 6558-6565.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963.

Manni M, Lima KM, Guglielmino CR, *et al.* (2015) Relevant genetic differentiation among Brazilian populations of *Anastrepha fraterculus* (Diptera, Tephritidae). *ZooKeys* **540**, 157-173.

Marí-Mena N, Lopez-Vaamonde C, Naveira H, Auger-Rozenberg M-A, Vila M (2016) Phylogeography of the Spanish moon moth *Graellsia isabellae* (Lepidoptera, Saturniidae). *BMC Evolutionary Biology* **16**, 139.

Maroja LS, Bogdanowicz SM, Wallin KF, Raffa KF, Harrison RG (2007) Phylogeography of spruce beetles (*Dendroctonus rufipennis* Kirby)(Curculionidae: Scolytinae) in North America. *Molecular Ecology* **16**, 2560-2573.

Marvaldi A, Oberprieler R, Lyal C, Bradbury T, Anderson R (2006) Phylogeny of the Oxycoryninae *sensu lato* (Coleoptera: Belidae) and evolution of host-plant associations. *Invertebrate Systematics* **20**, 447-476.

Matsubayashi KW, Ohshima I, Nosil P (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata* **134**, 1-27.

Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. *Harvard University Press*.

Mayr E (1970) Populations, species, and evolution: an abridgment of animal species and evolution. *Harvard University Press*.

McGowran B, Holdgate G, Li Q, Gallagher S (2004) Cenozoic stratigraphic succession in southeastern Australia. *Australian Journal of Earth Sciences* **51**, 459-496.

McLain DK (1980) Female choice and the adaptive significance of prolonged copulation in *Nezara viridula* (Hemiptera: Pentatomidae). *Psyche* **87**, 325-336.

McPeek MA (1996) Trade-offs, food web structure, and the coexistence of habitat specialists and generalists. *American Naturalist* **148**, S124-S138.

Meglécz E, Costedoat C, Dubut V, *et al.* (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.

Miklas N, Renou M, Malosse I, Malosse C (2000) Repeatability of pheromone blend composition in individual males of the southern green stink bug, *Nezara viridula*. *Journal of Chemical Ecology* **26**, 2473-2485.

Moir ML, Coates DJ, Kensington WJ, Barrett S, Taylor GS (2016) Concordance in evolutionary history of threatened plant and insect populations warrant unified conservation management approaches. *Biological Conservation* **198**, 135-144.

Molbo D, Krieger M, Herre E, Keller L (2002) Species-diagnostic microsatellite loci for the fig wasp genus *Pegoscapus*. *Molecular Ecology Notes* **2**, 440-442.

Molbo D, Machado CA, Sevenster JG, Keller L, Herre EA (2003) Cryptic species of fig-pollinating wasps: implications for the evolution of the fig-wasp mutualism, sex allocation, and precision of adaptation. *Proc Natl Acad Sci U S A* **100**, 5867-5872.

Mongeon P, Paul-Hus A (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* **106**, 213-228.

Mound L (1991) The first thrips species (Insecta, Thysanoptera) from cycad male cones, and its family level significance. *Journal of Natural History* **25**, 647-652.

Mound LA, Marullo R (1998) Biology and identification of Aeolothripidae (Thysanoptera) in Australia. *Invertebrate Systematics* **12**, 929-950.

Mound LA, Terry I (2001) Thrips pollination of the central Australian cycad, *Macrozamia macdonnellii* (Cycadales). *International Journal of Plant Sciences* **162**, 147-154.

Moussalli A, Hugall AF, Moritz C (2005) A mitochondrial phylogeny of the rainforest skink genus *Saproscincus*, Wells and Wellington (1984). *Molecular Phylogenetics and Evolution* **34**, 190-202.

Muraji M, Kawasaki K, Shimizu T (2001) Nucleotide sequence variation and use of mitochondrial DNA for phylogenetic analyses in anthocorid bugs (Hemiptera: Anthocoridae). *Japan Agricultural Research Quarterly* **35**, 85-90.

Musolin DL (2007) Insects in a warmer world: ecological, physiological and life-history responses of true bugs (Heteroptera) to climate change. *Global Change Biology* **13**, 1565-1585.

Musolin DL, Tougou D, Fujisaki K (2011) Photoperiodic response in the subtropical and warm-temperate zone populations of the southern green stink bug *Nezara viridula*: why does it not fit the common latitudinal trend? *Physiological Entomology* **36**, 379-384.

Nagalingum NS, Marshall CR, Quental TB, *et al.* (2011) Recent synchronous radiation of a living fossil. *Science* **334**, 796-799.

Nakagawa S, Noble DW, Senior AM, Lagisz M (2017) Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC biology* **15**, 18.

Nee S, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences* **344**, 305-311.

Nielsen AL, Holmstrom K, Hamilton GC, Cambridge J, Ingerson-Mahar J (2013) Use of black light traps to monitor the abundance, spread, and flight behavior of *Halyomorpha halys* (Hemiptera: Pentatomidae). *Journal of Economic Entomology* **106**, 1495-1502.

Norstog K, Nicholls TJ (1997) *The Biology of the Cycads* Cornell University Press, Ithaca, NY.

Norstog KJ, Stevenson DW, Niklas KJ (1986) The role of beetles in the pollination of *Zamia furfuracea* L. fil.(Zamiaceae). *Biotropica* **18**, 300-306.

Nuismer SL, Gomulkiewicz R, Ridenhour BJ (2010) When is correlation coevolution? *American Naturalist* **175**, 525-537.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa* **1668**, 491-520.

Oleksyn J, Karolewski P, Giertych M, *et al.* (1998) Primary and secondary host plants differ in leaf-level photosynthetic response to herbivory: evidence from *Alnus* and *Betula* grazed by the alder beetle, *Agelastica alni*. *New Phytologist* **140**, 239-249.

Olson DM, Ruberson JR, Andow DA (2012) Effects on stink bugs of field edges adjacent to woodland. *Agriculture Ecosystems & Environment* **156**, 94-98.

Olson DM, Ruberson JR, Andow DA (2016) Relative longevity of adult *Nezara viridula* in field cages of cotton, peanut, and soybean. *Entomologia Experimentalis et Applicata* **159**, 30-36.

Olson DM, Ruberson JR, Zeilinger AR, Andow DA (2011) Colonization preference of *Euschistus servus* and *Nezara viridula* in transgenic cotton varieties, peanut, and soybean. *Entomologia Experimentalis et Applicata* **139**, 161-169.

Ornduff R (1990) Geographic variation in reproductive behavior and size structure of the Australian cycad *Macrozamia communis* (Zamiaceae). *American Journal of Botany* **77**, 92-99.

Ortego J, Aguirre MP, Noguerales V, Cordero PJ (2015) Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications* **8**, 621-632.

Östergård H, Ehrlén J (2005) Among population variation in specialist and generalist seed predation– the importance of host plant distribution, alternative hosts and environmental variation. *Oikos* **111**, 39-46.

Panizzi AR (1997) Wild hosts of pentatomids: ecological significance and role in their pest status on crops. *Annual Review of Entomology* **42**, 99-122.

Panizzi AR (2000) Suboptimal nutrition and feeding behavior of hemipterans on less preferred plant food sources. *Anais da Sociedade Entomologica do Brasil* **29**, 1-12.

Panizzi AR, Meneguim AM (1989) Performance of nymphal and adult *Nezara viridula* on selected alternate host plants. *Entomologia Experimentalis et Applicata* **50**, 215-223.

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution* **27**, 1659-1672.

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

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

Pease CG, Zalom FG (2010) Influence of non-crop plants on stink bug (Hemiptera: Pentatomidae) and natural enemy abundance in tomatoes. *Journal of Applied Entomology* **134**, 626-636.

Pellmyr O (2003) Yuccas, yucca moths, and coevolution: a review. *Annals of the Missouri Botanical Garden* **90**, 35-55.

Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist* **152**, 428-446.

Petit J-R, Jouzel J, Raynaud D, *et al.* (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429-436.

Piry S, Luikart G, Cornuet J-M (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90**, 502-503.

Porras-Hurtado L, Ruiz Y, Santos C, et al. (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics* **4**.

Pott C, Krings M, Kerp H, Friis EM (2010) Reconstruction of a bennettitalean flower from the Carnian (Upper Triassic) of Lunz, Lower Austria. *Review of Palaeobotany & Palynology* **159**, 94-111.

Poutouli W (1995) Some oophagus parasitoids observed on Heteroptera in Togo. *Journal of African Zoology* **109**, 259-267.

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

Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* **16**, 608-627.

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* **4**, 4399-4428.

Qi B, Walter G, O'Toole A (1999) Structural incompatible? - An investigation on microstructures of genitalia of *Nezara viridula* (Heteroptera: Pentatomidae) from Australia and Slovenia. *Journal of Inner Mongolia Normal University (Natural Science Edition)* **29**, 48-54.

R Core Team (2016) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Rafter MA, Hereward JP, Walter GH (2013) Species limits, quarantine risk and the intrigue of a polyphagous invasive pest with highly restricted host relationships in its area of invasion. *Evolutionary Applications* **6**, 1195-1207.

Rajapakse CNK, Walter GH (2007) Polyphagy and primary host plants: oviposition preference versus larval performance in the lepidopteran pest *Helicoverpa armigera*. *Arthropod-Plant Interactions* **1**, 17-26.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Raymond L, Plantegenest M, Vialatte A (2013) Migration and dispersal may drive to high genetic variation and significant genetic mixing: the case of two agriculturally important, continental hoverflies (*Episyrphus balteatus* and *Sphaerophoria scripta*). *Molecular Ecology* **22**, 5329-5339.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Reineke A, Assaf H, Kulanek D, *et al.* (2015) A novel set of microsatellite markers for the European Grapevine Moth *Lobesia botrana* isolated using next-generation sequencing and their utility for genetic characterization of populations from Europe and the Middle East. *Bulletin of Entomological Research* **105**, 408-416.

Rew L, Medd R, Van de Ven R, *et al.* (2005) Weed species richness, density and relative abundance on farms in the subtropical grain region of Australia. *Animal Production Science* **45**, 711-723.

Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography* **34**, 566-575.

Rodrigues AR, Torres JB, Siqueira HA, Teixeira VW (2009) *Podisus nigrispinus* (Dallas)(Hemiptera: Pentatomidae) requires long matings for successful reproduction. *Neotropical Entomology* **38**, 746-753.

Roemer R, Terry I, Chockley C, Jacobsen J (2005) Experimental evaluation and thermo-physical analysis of thermogenesis in male and female cycad cones. *Oecologia* **144**, 88-97.

Roemer RB, Terry LI, Walter GH (2008) Unstable, self-limiting thermochemical temperature oscillations in *Macrozamia* cycads. *Plant Cell and Environment* **31**, 769-782.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rugman-Jones PF, Hoddle CD, Hoddle MS, Stouthamer R (2013) The lesser of two weevils: molecular-genetics of pest palm weevil populations confirm *Rhynchophorus vulneratus* (Panzer 1798) as a valid species distinct from *R. ferrugineus* (Olivier 1790), and reveal the global extent of both. *PLOS ONE* **8**, e78379.

Rugman-Jones PF, Hoddle MS, Stouthamer R (2010) Nuclear-mitochondrial barcoding exposes the global pest western flower thrips (Thysanoptera: Thripidae) as two sympatric cryptic species in its native California. *Journal of Economic Entomology* **103**, 877-886.

Ryan MA, Cokl A, Walter GH (1996) Differences in vibratory sound communication between a Slovenian and an Australian population of *Nezara viridula* (L.)(Heteroptera: Pentatomidae). *Behavioural Processes* **36**, 183-193.

Ryan MA, Moore CJ, Walter GH (1995) Individual variation in pheromone composition in *Nezara viridula* (Heteroptera: Pentatomidae): how valid is the basis for designating “pheromone strains”? *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **111**, 189-193.

Salas-Leiva DE, Meerow AW, Calonje M, *et al.* (2013) Phylogeny of the cycads based on multiple single-copy nuclear genes: congruence of concatenated parsimony, likelihood and species tree inference methods. *Annals of Botany* **112**, 1263-1278.

Schneider D, Wink M, Sporer F, Lounibos P (2002) Cycads: their evolution, toxins, herbivores and insect pollinators. *Naturwissenschaften* **89**, 281-294.

Schodde R, Mason I (1999) *The directory of Australian birds: Passerines* CSIRO Publishing, Melbourne.

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

Silva CCA, Laumann RA, Ferreira JBC, *et al.* (2012) Reproductive biology, mating behavior, and vibratory communication of the brown-winged stink bug, *Edessa meditabunda* (Fabr.) (Heteroptera: Pentatomidae). *Psyche* **2012**, 9.

Singh Z (1972) *Bionomics of the Southern Green Stink Bug Nezara viridula (Linn.) (Hemiptera: Pentatomidae) in Central India*, The University of Illinois.

Sithole S, de Milliano W, Kaula G, *et al.* (1986) The insect pest situation in sorghum at research stations in SADCC countries during the 1985/86 cropping season, 375-381.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Slobodkin LB, Sanders HL (1968) On the contribution of environmental predictability to species diversity **22**, 82-95.

Slove J, Janz N (2011) The relationship between diet breadth and geographic range size in the butterfly subfamily Nymphalinae - a study of global scale. *PLOS ONE* **6,** e16057.

Smith CI, Godsoe WK, Tank S, Yoder JB, Pellmyr O (2008a) Distinguishing coevolution from covicariance in an obligate pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. *Evolution* **62**, 2676-2687.

Smith CI, Pellmyr O, Althoff DM, *et al.* (2008b) Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. *Proceedings of the Royal Society B: Biological Sciences* **275**, 249-258.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PD (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proceedings of the National Academy of Sciences* **104**, 4967-4972.

Snow EL, Walter GH (2007) Large seeds, extinct vectors and contemporary ecology: testing dispersal in a locally distributed cycad, *Macrozamia lucida* (Cycadales). *Australian Journal of Botany* **55**, 592-600.

Sosa-Gómez DR, da Silva JJ, Costa F, *et al.* (2005) Population structure of the Brazilian southern green stink bug, *Nezara viridula*. *Journal of Insect Science* **5**, 1-10.

Stevenson D, Norstog K, Fawcett P (1998) Pollination biology of cycads. In: *Reproductive Biology* (eds. Owens SJ, Rudall PJ), pp. 277-294. Royal Botanical Gardens, Kew, United Kingdom.

Strong DR (1979) Biogeographic dynamics of insect-host plant communities. *Annual Review of Entomology* **24**, 89-119.

Suinyuy TN, Donaldson JS, Johnson SD (2012) Geographical variation in cone volatile composition among populations of the African cycad *Encephalartos villosus*. *Biological Journal of the Linnean Society* **106**, 514-527.

Suinyuy TN, Donaldson JS, Johnson SD (2013) Variation in the chemical composition of cone volatiles within the African cycad genus *Encephalartos*. *Phytochemistry* **85**, 82-91.

Taerum SJ, Konečný A, Beer ZW, Cibrián-Tovar D, Wingfield MJ (2016) Population genetics and symbiont assemblages support opposing invasion scenarios for the red turpentine beetle (*Dendroctonus valens*). *Biological Journal of the Linnean Society* **118**, 486-502.

Terry I (2001) Thrips and weevils as dual, specialist pollinators of the Australian cycad *Macrozamia communis* (Zamiaceae). *International Journal of Plant Sciences* **162**, 1293-1305.

Terry I, Forster PI, Moore CJ, Roemer RB, Machin PJ (2008) Demographics, pollination syndrome and conservation status of *Macrozamia platyrhachis* (Zamiaceae), a geographically restricted Queensland cycad. *Australian Journal of Botany* **56**, 321-332.

Terry I, Moore CJ, Walter GH, *et al.* (2004) Association of cone thermogenesis and volatiles with pollinator specificity in *Macrozamia* cycads. *Plant Systematics and Evolution* **243**, 233-247.

Terry I, Walter GH, Moore C, Roemer R, Hull C (2007) Odor-mediated push-pull pollination in cycads. *Science* **318**, 70-70.

Terry IL, Roemer RB, Booth DT, Moore CJ, Walter GH (2016) Thermogenic respiratory processes drive the exponential increase of volatile organic compound emissions in *Macrozamia* cycad cones. *Plant, Cell & Environment*.

Terry LI, Roemer RB, Walter GH, Booth D (2014) Thrips' responses to thermogenic associated signals in a cycad pollination system: the interplay of temperature, light, humidity and cone volatiles. *Functional Ecology* **28**, 857-867.

Terry LI, Walter GH, Donaldson JS, *et al.* (2005) Pollination of Australian *Macrozamia* cycads (Zamiaceae): effectiveness and behavior of specialist vectors in a dependent mutualism. *American Journal of Botany* **92**, 931-940.

Thompson JN (2005) *The geographic mosaic of coevolution.* University of Chicago Press, Chicago, IL.

Thompson JN, Cunningham BM (2002) Geographic structure and dynamics of coevolutionary selection. *Nature* **417**, 735-738.

Tillman PG (2014) Physical barriers for suppression of movement of adult stink bugs into cotton. *Journal of Pest Science* **87**, 419-427.

Tillman PG, Cottrell TE, Mizell RF, Kramer E (2014) Effect of field edges on dispersal and distribution of colonizing stink bugs across farmscapes of the Southeast USA. *Bulletin of Entomological Research* **104**, 56-64.

Tillman PG, Greenstone MH, Hu JS (2015) Predation of stink bugs (Hemiptera: Pentatomidae) by a complex of predators in cotton and adjoining soybean habitats in Georgia, USA. *Florida Entomologist* **98**, 1114-1126.

Tillman PG, Northfield TD, Mizell RF, Riddle TC (2009) Spatiotemporal patterns and dispersal of stink bugs (Heteroptera: Pentatomidae) in peanut-cotton farmscapes. *Environmental Entomology* **38**, 1038-1052.

Todd JW (1989) Ecology and behavior of *Nezara viridula*. *Annual Review of Entomology* **34**, 273-292.

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907-3930.

Toon A, Daglish G, Ridley A, *et al.* (2016) Random mating between two widely divergent mitochondrial lineages of *Cryptolestes ferrugineus* (Coleoptera: Laemophloeidae): A test of species limits in a phosphine-resistant stored product pest. *Journal of Economic Entomology* **109**, 2221-2228.

Tougou D, Musolin DL, Fujisaki K (2009) Some like it hot! Rapid climate change promotes changes in distribution ranges of *Nezara viridula* and *Nezara antennata* in Japan. *Entomologia Experimentalis et Applicata* **130**, 249-258.

Trapero C, Wilson IW, Stiller WN, Wilson LJ (2016) Enhancing integrated pest management in GM cotton systems using host plant resistance. *Frontiers in Plant Science* **7**, 500.

Tripp EA, Manos PS (2008) Is floral specialization an evolutionary dead-end? pollination system transitions in *Ruellia* (Acanthaceae). *Evolution* **62**, 1712-1737.

Velasco LR, Walter GH (1992) Availability of different host plant species and changing abundance of the polyphagous bug *Nezara viridula* (Hemiptera: Pentatomidae). *Environmental Entomology* **21**, 751-759.

Velasco LR, Walter GH (1993) Potential of host-switching in *Nezara viridula* (Hemiptera: Pentatomidae) to enhance survival and reproduction. *Environmental Entomology* **22**, 326-333.

Velasco LR, Walter GH, Harris VE (1995) Voltinism and host plant use by *Nezara viridula* (L.)(Hemiptera: Pentatomidae) in Southeastern Queensland. *Australian Journal of Entomology* **34**, 193-203.

Venugopal PD, Coffey PL, Dively GP, Lamp WO (2014) Adjacent habitat influence on stink bug (Hemiptera: Pentatomidae) densities and the associated damage at field corn and soybean edges. *PLOS ONE* **9**, e109917.

Virant-Doberlet M, Čokl A, Stritih N (2000) Vibratory songs of hybrids from Brazilian and Slovenian populations of the green stink bug *Nezara viridula*. *Pflügers Archiv* **439**, r196-r198.

Vivan LM, Panizzi AR (2005) Nymphal and adult performance of genetically determined types of *Nezara viridula* (L.)(Heteroptera: Pentatomidae), under different temperature and photoperiodic conditions. *Neotropical Entomology* **34**, 911-915.

Walker S, Taylor I, Milne G, *et al.* (2005) A survey of management and economic impact of weeds in dryland cotton cropping systems of subtropical Australia. *Animal Production Science* **45**, 79-91.

Wallenius T, Peakall R, Wanjura W, *et al.* (2012) Volatile emissions, thermogenesis, and dehiscence patterns of *Macrozamia communis* (Zamiaceae): implications for cycad pollination research., 395-418.

Wallingford AK, Kuhar TP, Pfeiffer DG, *et al.* (2013) Host plant preference of harlequin bug (Hemiptera: Pentatomidae), and evaluation of a trap cropping strategy for its control in collard. *Journal of Economic Entomology* **106**, 283-288.

Wallner AM, Hamilton GC, Nielsen AL, *et al.* (2014) Landscape factors facilitating the invasive dynamics and distribution of the brown marmorated stink bug, *Halyomorpha halys* (Hemiptera: Pentatomidae), after arrival in the United States. *PLOS ONE* **9**, e95691.

Walter GH (2003) *Insect Pest Management and Ecological Research.* Cambridge University Press, United Kingdom.

Walther G-R, Post E, Convey P, *et al.* (2002) Ecological responses to recent climate change. *Nature* **416**, 389-395.

Walther G-R, Roques A, Hulme PE, *et al.* (2009) Alien species in a warmer world: risks and opportunities. *Trends in Ecology & Evolution* **24**, 686-693.

Wang GH, Zhang X, Hou YM, Tang BZ (2015) Analysis of the population genetic structure of *Rhynchophorus ferrugineus* in Fujian, China, revealed by microsatellite loci and mitochondrial COI sequences. *Entomologia Experimentalis et Applicata* **155**, 28-38.

Waterhouse DF, Norris KR (1987) *Biological Control: Pacific Prospects.* Inkata Press Melbourne, Melbourne.

Wei SJ, Cao LJ, Gong YJ, *et al.* (2015) Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. *Molecular Ecology* **24**, 4094-4111.

Werle E, Schneider C, Renner M, Völker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research* **22**, 4354-4355.

Wiebes J (1979) Co-evolution of figs and their insect pollinators. *Annual Review of Ecology and Systematics* **10**, 1-12.

Williamson-Natesan EG (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics* **6**, 551-562.

Wilson DS, Yoshimura J (1994) On the coexistence of specialists and generalists. *American Naturalist* **144**, 692-707.

Wilson GW (2002) Insect pollination in the cycad genus *Bowenia* Hook, ex Hook. f. (Stangeriaceae). *Biotropica* **34**, 438-441.

Wilson L (2010) IPM in Bollgard II®: coping with changes in pests and climate. Final Report to the Cotton Catchment Communities CRC, Narrabri, NSW.

Wilson L, Downes S, Khan M, *et al.* (2013) IPM in the transgenic era: a review of the challenges from emerging pests in Australian cotton systems. *Crop and Pasture Science* **64**, 737-749.

Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **30**, 1720-1728.

Yoder JB, Nuismer SL (2010) When does coevolution promote diversification? *American Naturalist* **176**, 802-817.

Yukawa J, Kiritani K (1965) Polymorphism in the southern green stink bug. *Pacific Insects* **7**, 639-642.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686-693.

Žunič A, Virant-Doberlet M, Čokl A (2011) Species recognition during substrate-borne communication in *Nezara viridula* (L.)(Pentatomidae: Heteroptera). *Journal of Insect Behavior* **24**, 468-487.

Appendices

Appendix 1.1 Conference abstract for the paper presented at the Australian Cotton Research Conference at Toowoomba, Queensland, in 2015.

Conference abstract for the citation:

> Brookes, D.R., Hereward, J.P., Wilson, L.J. and Walter, G.H., 2015. Gene flow and host use in the Green Vegetable Bug, *Nezara viridula.* Paper presented at the *Australian Cotton Research Conference,* Toowoomba, Queensland.

Gene flow and host use in the Green Vegetable Bug, *Nezara viridula*

With the widespread use of *Bt* cotton, plant-sucking insects have become more significant pests in cotton crops. One of these pests is the Green Vegetable Bug (Hemiptera: *Nezara viridula*; GVB), a globally significant pest that probably arrived in Australia soon after European settlement. In cotton, these bugs feed on developing bolls and destroy them, but only occasionally reach numbers that require control. Similar patterns are found for many of the crop hosts GVB feed on. The fluctuations in GVB populations and their sporadic appearance in an area (or on a particular crop) make predicting when and why GVB will be present in damaging numbers in a particular location or crop challenging.

Our research aims to understand more deeply the relationship between GVB and its hosts in Australia, and to establish the relationships between Australian GVB and international populations of the pest. We are investigating the spatial and temporal population dynamics of GVB using both phylogenetic and population genetics approaches. We will thus investigate gene flow between hosts and geographic areas within Australia and then compare these results with those from overseas. This information will be used to contextualize the global research on GVB to Australian growing regions, thereby aiding the development of appropriate management strategies for GVB.

Appendix 1.2 Conference abstract for the paper presented at the World Cotton Research Conference at Goiânia - Goiás, Brazil, in 2016.

Conference abstract for the citation:

> Brookes, D.R., Hereward, J.P., Wilson, L.J. and Walter, G.H., 2016. Gene flow and host use, relative to cotton, in *Nezara viridula* (the Green Vegetable Bug or Southern Green Stink Bug). Paper presented at the *World Cotton Research Conference 6*, Goiânia - Goiás, Brazil.

Gene flow and host use, relative to cotton, in *Nezara viridula* (the Green Vegetable Bug or Southern Green Stink Bug)

With the widespread use of *Bt* cotton, plant-sucking insects have become more significant pests in cotton crops. In Australia, this includes the Green Vegetable Bug or Southern Green Stink Bug (Hemiptera: *Nezara viridula*), an insect pest of global significance. These bugs feed on the developing cotton bolls, potentially reducing yield and staining the lint, but they only occasionally reach numbers that require control. Outside of cotton these insects feed on plants from over 40 families, with the specific host composition varying regionally and temporally, so it is difficult to determine why these insects invade cotton and when they will do so.

Our sampling was designed to understand what role, if any, the presence of distinct genetic lineages within *N. viridula* plays in the association of this species with different hosts and in its distribution across continental Australia. We used both phylogenetics and population genetics approaches to assessing gene flow across populations of *N. viridula* from different host plants, as well as across populations from different geographical areas within Australia.

Mitochondrial DNA analysis reveals that *N. viridula* arrived in Australia through successive invasions, indicating a complex biogeographical history that resulted in both the Asian and European mitochondrial lineages being present. These lineages are partitioned geographically in Australia, but show no further genetic structure across different host plants within any particular region. Analysis of microsatellite markers indicates, further, that gene flow has occurred across these mitochondrial lineages, in at least some locations. Further analysis on Australian samples is underway.

Our current understanding of the genetic diversity within global *N. viridula* populations relies on previous research that focused on these different mitochondrial lineages. The discordance between the microsatellite and mitochondrial data in our results indicates how the global *N. viridula* populations could be better understood by using microsatellite markers to assess their population genetic structure. This will allow for a more accurate interpretation of the genetic makeup of *N. viridula* populations, which can then be compared with recorded patterns of host use, distribution, and regional population dynamics of this species, and thus inform why and when *N. viridula* invades cotton.

Appendix 2.1 Salt DNA extraction protocol for shotgun sequencing

The flight muscle of *Nezara viridula* was lysed using 3 µL of Proteinase K (20mg/ml) and 600 µL of lysis buffer (50mM tris-HCl pH 8.0, 20mM EDTA pH 8.0, and 2% SDS). The sample was incubated overnight at 55°C on a shaker, then at 37°C for 1 hour after 3 µL of Rnase (10mg/µL) had been added, and then placed on ice for 5 minutes. Each sample then had 200µL of NaCl (5M) added to it and was mixed through inversion before being centrifuged at high speed for 2 min. The supernatant was discarded and 1mL of cold isopropyl (100%) was added. The sample was repeatedly inverted (50x), centrifuged at high speed for 5 min, and washed twice using 1mL of EtOH (75%) (but left to rest for 15 min each time EtOH was added). The EtOH was removed a final time and the sample left to air dry overnight, when 50µL of TE buffer was added to dissolve the DNA pellet.

Appendix 2.2 *Nezara viridula* microsatellite development

The flight muscle of a single *N. viridula* individual was dissected and DNA extracted using a salt method (see above). The DNA was shotgun sequenced at the Australian Genome Research Facility (AGRF) using the Illumina MiSeq platform. Paired end reads (250 bp) were merged using FLASh 1.2.7 (Magoč & Salzberg 2011) for a total of 1,015,756 reads. The software QDD2 (Meglécz *et al.* 2010) was used to analyse these reads so as to select appropriate microsatellite loci, with up to 3°C difference allowed in the melting temperature of primer pairs and excluding short reads of 80 bp and lower. Only microsatellite loci with more than five repeats were considered, and so primers for 10,450 microsatellite loci were developed. The characteristics of these microsatellites loci can be found in Table 2.2.1.

Table 2.2.1. Characteristics of the 10,450 microsatellite loci with a repeat length of five or more identified by QDD2. The values A – G represent the design (quality) of the microsatellite loci (see QDD2 manual).

Motif Size	**A**	**B**	**C**	**D-G**
Di-	1663	1175	168	6067
Tri-	354	148	18	738
Tetra-	31	13	1	58
Penta-	4	0	0	9
Hexa-	2	0	0	1
Total	2054	1336	187	6873

We tested primer pairs for 134 microsatellite loci, selecting loci with a design of ‘A’ or ‘B’ and using the primer pairs with the lowest penalty among the options available for each locus. The motif size composition of these loci was di- (32), tri- (86), tetra- (15), and hexa- (1) nucleotides. A fluorescent M13 tail was added to the forward primer (Schuelke 2000) and a ‘pig tail’ to the reverse primer (Brownstein *et al.* 1996). Primers were initially tested on four individuals representative of the ‘Asian’ lineage collected from Kununurra in the Northern Territory, and four individuals representative of the ‘European’ lineage collected from Griffith in New South Wales (see methods for collection details). PCR protocols are presented in the following section. We used a micro electrophoresis system (MCE-202 MultiNA, Shimadzu Corporation, Kyoto Japan) to assess initial primer amplification. We genotyped 46 individuals, 23 from each lineage, at the 63 loci deemed suitable for further testing. Seven of these loci were tested with a second primer pair because they had high null allele estimates in only one lineage.

Twelve microsatellite loci were chosen to genotype all individuals (Table A2.2.2), with the choice based on allelic richness and null allele estimates (within lineages). Details of the loci can be found in Table 2.2 of Chapter 2. Null allele estimates were calculated using FreeNA (Chapuis & Estoup 2007) and those loci with high estimates (>10%) for both lineages were excluded from consideration. Ten loci that had high null allele estimates in only one lineage, so a second pair of primers was tested, as outlined above, to replace them.

Table 2.2.2. Characteristics of the 12 microsatellite loci developed for genotyping samples of the Asian and European lineages of *Nezara viridula*. Summary statistics for these loci, including null allele estimates, can be found in Table 2.2 of Chapter 2.

Locus	**Primer sequence (5'-3')**	**Motif**	**Approx. product size (bp)**	**Dye**
NEZA01	F- TTCCTCTGCCTAGGTGTGCT R- TTCAGGTTCGCCAAGTCTTT	AAT	161	6-FAM
NEZA02	F- GAAACCGGATCATCTCAAGC R- ATTATTTACGGAGTTGGCCG	AAT	150	VIC
NEZA03	F- GTGGATCGGCTTGACAATCT R- TTAGTCCACCTCGAACCCAG	AGC	140	NED
NEZA04	F- TACTCATCATGGCTGCGTGT R- ACCATCTAAGCATGGAAGCG	AAG	188	PET
NEZA05	F- CTTGTCGTAGCCAGGGAATC R- TGGCCATGAAGAATAACAACA	AAAT	107	6-FAM
NEZA06	F- GTCGATCATTTCAGGCCAAC R- TTAGCTTCCGTAAAGCGAGTC	AAG	261	VIC
NEZA07	F- GGCTGTACAATGCTGGTCTTC R- ATAATTCATAAGTGTAATGTGCTTCAG	AAAT	212	NED
NEZA08	F- GGAACAAACCAAACAGGTGAA R- GGTATTCTGGAGCAAGTGGC	AGG	182	PET
NEZA09	F- CATAAGGAGGTCGCACGAAT R- GCAAATCTCAACCTATTGATTTATGT	ACAT	119	6-FAM
NEZA10	F- AGGGAAATGTTGTGGACCTG R- GACCGAGATCACGCTCAATC	AAT	127	VIC
NEZA11	F- AAAGGACCTCTTGGGCATCT R- TTGACGTGGGAATTGTCAGA	AAG	187	NED
NEZA12	F- AAATCGATGCCAGAAACCTG R- TTGCGAATCTGGAACGAAAT	ATC	243	PET

Appendix 2.3 PCR protocols, EF1α and Tubα1 primer development

Microsatellite genotyping PCR protocol

PCR was carried out using 12 µL reactions with 1x myTAQ buffer, 0.1 µM of forward primer, 0.2 µM of reverse primer, 0.3 units of myTAQ HS DNA polymerase, 0.2 µM of M13 label (Schuelke 2000) (either 6-FAM, VIC, NED, or PET) and 3.0µl of DNA template. The primers used were those outlined above in Table 2.2.2. PCR cycling conditions were 10 min at 95°C followed by 25 cycles of 95°C for 25, 30 s at 57°C and 45 s at 72°C. A second series of 10 cycles followed, of 25 s at 95°C, 30 s at 54°C, 45 s at 72°C, and ending with an extension of 10 min at 72°C. The PCR product was cleaned using 2µl per sample of Exo-Ap (adapted from Werle *et al.* (1994)) with PCR conditions of 15min at 37°C followed by 15min at 80°C. Exo-Ap was made up of 5µl Exonuclease I (20U/µl), 10µl buffer and 85µl H_2O combined with 20µl Antarctic Phosphatase (5U/µl), 10µl buffer and 70µl H_2O. The products were separated by capillary electrophoresis on an ABI 3730XL (Macrogen, Korea).

Gene sequencing primers

The genes cytochrome C oxidase subunit I (COI), elongation factor 1 alpha (EF1α) and tubulin alpha 1 (Tubα1) were sequenced with their trimmed sequences being 608 bp, 417 bp, and 458 bp in length respectively. Primers for COI were the previously published barcoding primers HCO and LCO (Folmer *et al.* 1994). Previously developed primers for EF1α (primers Shirley and Prowler (Cho *et al.* 1995) and Tubα1 (TH_TubA forward and reverse primers (Buckman *et al.* 2013)) were used to amplify these gene regions but were largely unsuccessful. The few sequences that amplified successfully were sequenced, and those that aligned with other EF1α and Tubα1 sequences from Genbank were used to develop primers specific to *N. viridula* using Primer3.

Gene sequencing PCR protocol

PCR reactions and protocols were the same for all three genes except that the annealing steps differed. PCR was carried out using 20 µL reactions with 1x myTAQ buffer, 0.4 µM of forward primer, 0.4 µM of reverse primer, 1 unit of myTAQ HS DNA polymerase, and 2.0µL of DNA template. The primers used were those outlined above in Table 2.3.1. PCR cycling conditions were 2 min at 95°C followed by 8 cycles of 95°C for 25, 1 min at 1st annealing temperature and 1 min 15 s at 72°C. A second series of 32 cycles followed, of 30 s at 95°C, 45 s at 2nd annealing temperature, 1 min 15 s at 72°C, and ending with an extension of 10 min at 72°C. The PCR product was cleaned using Exo-Ap (adapted from Werle *et al.* (1994)) as outlined above in the microsatellite PCR protocol. The products were then sequenced using Sanger sequencing (Macrogen, Korea).

Table 2.3.1. Primers, annealing temperatures and references for the PRC protocols of the COI, EF1α, and Tubα1 gene regions.

Gene region	**Primers**	**Annealing temp. (°C)**	**Reference**
COI	F-GGTCAACAAATCATAAAGATATTGG R- TAAACTTCAGGGTGACCAAAAAATCA	51 – (40 cycles)	(Folmer *et al.* 1994)
EF1α	F- ACGCACTTCTTGCCTTCACT R- TGACAACCATACCTGGTTTCAAT	61 (8 cycles) 59 (32 cycles)	This study
Tubα1	F- TTCCATTTTGACCACTCACACT R- CCAGTTGGACACCAATCAACA	51 (8 cycles) 47 (32 cycles)	This study