School of Biomedical Science

Biology and Systematics of *Trogoderma* species with Special Reference to Morphological and Molecular Diagnostic Techniques for Identification of *Trogoderma* pest species

Mark A Castalanelli

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

February 2011

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature:

Date:

Acknowledgments

Special thanks to the Cooperative Research Centre for National Plant Biosecurity for their support: Dr Simon McKirdy, Dr Kirsty Bayliss and Dr David Eagling. Kirsty you have been a great coordinator of the education stream, I appreciate the support and organisation of the yearly events across Australia. Simon McKirdy your passion for biosecurity has provided a strong backbone for the future of biosecurity in this country; let's hope the second rebid is successful.

One of the most important groups that need to be thanked are my colleagues from Curtin University, for three years this was my home away from home: Prof David Groth, Dr Kylie Munyard, Dr Keith Gregg, Danielle Giustiniano, Sharon Siva, Mel Corbett, Carla Zammit, Dr Chee Yang Lee, Jason Legder, Ryan Harris, Tash Feeley, Jaden Elphick, Rhys Cransberg, Laurton McGurk, Ibrahim Marzouqi, Rob Walker and Rob Steuart. David thank you for taking on this project when it was just an infant idea and turning it in to one of the most successful projects to come out of the CRC for National Plant Biosecurity, providing positive feedback, pushing to publish and always wanting to have a chat in the lab. Kylie you have been one of my biggest advocates and I thank you for all your support. Your molecular prowess and your willingness to provide feedback (in red ink) on my Honours and PhD theses is highly appreciated. Keith your molecular knowledge and your grasp of the English language is second to none.

My colleagues at the Department of Agriculture and Food Western Australia who have immersed me into the world of entomology: Mike Grimm, Andras Szito, Graham McAlpine, Robert Cunningham, Rob Emery, Dr Darryl Hardie, Dr Oonagh Byrne, Dr Glyn Maynard (DAFF), Pia Scalon, Michelle Chami, John Van Schagen, Lisa Vagg, Danica Collins, Dustin Severtson, Cameron Brumley, Matt Davis, Dr John Botha, Dr Nic Monzu, Dr Shashi Sharma, and the European House Borer staff. Mike the hallways have turned quiet since you have returned to Albany and Friday's just aren't the same without sitting down for quick chat that turns into a 3hr discussion on all things life related. To our morphological expert, Andras, thank you

for sharing the wonderful world of morphological taxonomy with me. From the looks of it you have a life time of whipping out *Trogoderma* genitalia ahead of you, and you think DNA is small. Rob your quest for perfection has given me a new lookout on scientific method and u 'R' a true scientific inspiration to me. Graham you are the cleverest and most practical man that I know, thank your for all your help and encouragement.

The population genetics team from Queensland University of Technology: Dr Andrew Baker, Dr Matthew Krosch, Elli and Susan Harvey – thank you for the crash course in both population genetics and molecular phylogeny. Andrew thank you for the great opportunity to study under your guidance and for making me feel welcome on my numerous trips to Brisbane. Matt thank you for showing me a range of phylogenetic techniques and programs that helped with the analysis of *Trogoderma* data presented in thesis. Ali and Susan – thank you for showing me Brisbane as only students can.

Those that provided and collected Dermestidae samples from across Australia and the world: Dr Katarina Mikac, Dr Joel Floyd, Dr Norman Barr, Roxanne Farris, Dr Evan Braswell, Darren Peck and Luke Watson, Dr Joe Holloway, and Dr Bobbie Hitchcock. Katarina thank you for sharing the *T. variabile* trapping material, while the results were not what we expected it made for a great story. Joel and Norman thank you for taking time to show me around the US agriculture institutions and making my visit to the US a memorable experience. Roxanne, my arteries are still recovering from that breakfast burrito, and Evan your taste in beer is impeccable.

Finally, the most important and influential people in my life - my family. My wife, Natalie and our son (Maxwell - 10 Months), our furry children (Jazz and George), my parents (Chris and Cass), and sister (Kim). The Frost's: (Jill, Simon, and Nina). Natalie you are my inspiration, my rock and best friend, thank you for all the years of love and support, and for happily collecting *Trogoderma* from around the South-west of Australia when you where four and eight months pregnant. My favourite little man, you are the most wondrous thing I have ever laid eyes on.

Abstract

The genus *Trogoderma* contains some of the world's most serious invasive pests of wheat and other stored grain products. Even with the application of strict quarantine measures, these grain beetles still find their way into new countries and cause imposition of onerous restrictions on exports. Two of the most serious *Trogoderma* pest species are *Trogoderma granarium* Everts and *Trogoderma variable* Ballion. At present both species can only be reliably identified by a limited number of highly skilled diagnosticians trained in traditional morphological based keys. To differentiate between these two species, delicate dissections of the genitalia and mouthparts are required. These dissections can take over an hour, even for skilled diagnosticians. However, accurate identification of larvae or a damaged specimen is difficult. To overcome these identification problems, the use of molecular markers in differentiating *Trogoderma* species was investigated.

In Australia, *T. variabile* is regarded as a minor yet persistent pest, but is considered a concern as it could mask the presence of *T. granarium*. To track the distribution of *T. variabile* the Commonwealth Scientific and Industrial Research Organisation (CSRIO) conducted a national trapping program from 2001 to 2003. Using partial sequences of two mitochondrial genes (Cytochrome Oxidase I and Cytochrome b) and the nuclear gene (18S), these samples were used in a phylogenetic study to examine the distribution, dispersal pathways, and species boundaries of *T. variabile*. Based on the molecular results, only 47% of the samples analysed were *T. variabile,* and the remaining were a mixture of six putative species. Unfortunately, all specimens were macerated to extract the DNA and so it was not possible to re-examine the specimens morphologically to confirm their identity.

A non-destructive DNA extraction method is an important step in ensuring samples can be re-examined when molecular and morphological results are incongruent. Using EDNA HiSpEx, a rapid and non-destructive DNA extraction technique for arthropod specimens was developed. This technique was tested on four arthropod orders, using specimens that were fresh, preserved by air drying, stored in ethanol, or collected with sticky or propylene glycol traps. The extraction could be completed in

20 minutes for Coleoptera, Diptera and Hemiptera, and two minutes for the subclass Acarina, without significant distortion, discolouration, or other damage to the specimens. All samples analysed hereafter had their DNA extracted using this method.

Using partial sequences of two mitochondrial genes, Cytochrome Oxidase I and Cytochrome b, and the nuclear gene, 18S, a phylogenetic study was conducted to differentiate *T. granarium* from *T. variabile* and other closely related species that were collected from across Australia. The aim of this study was to generate a reference database that could serve as a frame of reference for the identification of quarantine intercepts, rediscover the putative species reveal in the CSIRO trapping program, and evaluate the existing morphologically-based polyphyletic positioning of *Trogoderma*. Molecular phylogenetic reconstruction revealed that the *Trogoderma* genus was paraphyletic. However, this finding was only supported by Bayesian analysis because Parsimony analysis exhibited polytomy and the putative species identified in the CSIRO trapping program were not rediscovered in this study. The sequence data was also used to estimate time to most recent common ancestor for *Trogoderma* and four closely related genera from the Megatomini and Attagenini tribes. Estimation of Dermestidae origins exceeded 175 million years, placing the origins of this family in Pangaea.

Without an extensive sampled database that has both phylogenetic and morphological support, using a DNA barcoding approach could lead to inaccurate identification of intercepted material. Furthermore, DNA barcoding is time consuming when DNA sequences need to be processed offsite, with results within days rather than hours. The aim was to develop a PCR test that had high sensitivity and specificity. A quantitative PCR test was developed by multiplexing hydrolysis probes, amplification controls, and including a melt curve analysis step. As well as accurate identification of *T. variabile* and *T. granarium*, this multiplex test provided an internal checking system thereby reducing the incidence of false positives and false negatives.

This study highlights the importance of combining both molecular methods and morphological characteristics. With out morphological data, it is difficult to ascertain

when molecular results are incongruent because there is no standard genetic measure that can be used to separate a species. Likewise, morphological taxonomy needs to be checked by molecular markers as to ensure important morphological characters are not overlooked and that morphologically similar species are not grouped together. Only once both methods are combined can a pest species be accurately diagnosed.

Abbreviations

A.v	*A. verbasci*
18S	nuclear small subunit rRNA / rDNA
AFLPs	Amplified Fragment Length Polymorphism
AQIS	Australian Quarantine and Inspection Service
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
bp	base pair(s)
COI	Cytochrome oxidase subunit 1
Cp	Crossing Point
CSIRO	Commonwealth Scientific and Industrial Research Organization
CYT b	Cytochrome b
DNA	Deoxyribonucleic acid
dsDNA	double stranded DNA
EF-1a	Elongation Factor 1 Alpha
ESS	Effective Sample Sizes
GTR	General Time Reversal
HRM	High Resolution Melt curve analysis
Indels	Insertions or deletions
ITS1	Internal Transcribed Spacer 1
kb	kilo base
MCA	Melt Curve Analysis
ML	Maximum Likelihood
mtDNA	Mitochondrial DNA
MA	Million years ago
NUMTs	Nuclear Mitochondrial DNA
PAUP	Phylogenetic Analysis Using Parsimony
PCR	Polymerase Chain Reaction
PHT	Partition Homogeneity Test
RAPD	Random Amplification of Polymorphic DNA
rDNA	Ribosomal DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
SNP	Single Nucleotide Polymorphism
s/s/Myr	substitutions per site per million years
T.g	*T. granarium*
T.v	*T. variabile*
TMRCA	Time to Most Recent Common Ancestor
tRNA	transfer RNA
ya	years ago

Table of Contents

Figures

Tables

Chapter 1

Use of Molecular Methods to study an Invasive Entomological Pest Problem

1.1 Invasive Pests

Invasive pests are organisms that enter new often distant environments, proliferate due to suitable environmental conditions and a lack of natural control agents (predators, parasites, pathogens), and typically spread extensively, often to the detriment of the existing natural ecosystem (Mack *et al.*, 2000, Cock, 2003). They not only have the potential to cause adverse effects to the natural ecosystem, but can also impose a significant cost to agricultural industries (Mack *et al.*, 2000). For example, the cost to the United States agricultural sector of crops affected by insects is estimated at $US 13.9 billion annually and the introduction of the Formosan termite is estimated to cost $US 1 billion annually (Pimentel *et al.*, 2004). While an introduction may appear insignificant it can have a dramatic economic effect upon the invaded area and the country's ability to export agricultural products.

To limit the effect of an insect invasion, measures can be directed at eradication or at least curtailing the expansion of the insects range through application of control measures and management programs (Mack *et al.*, 2000, Miller and Rossman, 1995, Cock, 2003). Control measures include restricted movement of potentially infested materials, pesticide application, blocking points of entries, and the use of biological controls (Miller and Rossman, 1995, Cock, 2003). However, management efforts can be misdirected without an understanding of the pest species' biology, ecology and evolutionary pathways, resulting in considerable wastage of time and resources (Mack *et al.*, 2000, Suarez *et al.*, 2008, Schutze *et al.*, 2006, Walter, 2003).

Fundamental to the implementation of control measures is the correct identification of the organism. Morphological identification problems can occur when dealing with immature stages or with cryptic species in the invaded ecosystem (Schutze *et al.*, 2006). Immature stages of insects generally lack unique morphological characteristics to enable them to be distinguished from closely related species and

morphological keys, used to identify adult and larvae samples, may be geographically limited and/or effective for only a particular life stage (Hebert *et al.*, 2003). While morphologically distinct characters are often not present in larvae, there are many instances where two or more adult species are unified under the same taxonomic identity (referred to as a cryptic species), but have different life cycles and host preference (Walter, 2003). For instance, *Bactrocera tryoni* and *Bactrocera aquilonis* are two species that are morphologically similar but differ markedly in host specificity, locality, and time of daily optimum activity (Wang *et al.*, 2003). The presence of cryptic species can create problems for control methods and management strategies as each species may react quite differently. For example, Western Flower Thrips (*Frankliniella occidentalis* Pergande) were controlled using one type of pesticide, although this pest was controlled only at particular sites (Rugman-Jones *et al.*, 2010). The authors later confirmed that *F. occidentalis* includes two strains, each requiring different insecticide control methods (Rugman-Jones *et al.*, 2010). Correct identification is paramount to successful control measures and management programs

1.2 Phylogenetic Markers

DNA marker technology is increasingly being applied to solve identification problems (e.g. DeSalle, 2006, Foottit *et al.*, 2008, Gwiazdowski *et al.*, 2006, Loxdale and Lushai, 1998, Lambert, 1983, Carew *et al.*, 2009). Sequence based molecular markers, used to study invasive insect pests, can be split into two broad categories: DNA taxonomy or population analysis. DNA taxonomy is the use of a DNA sequence to identify a species (Tautz *et al.*, 2003), more specifically the use of the 5` end of the Cytochrome oxidase I gene is referred to as DNA barcoding (Hebert *et al.*, 2003). DNA taxonomy has aided in the identification of characteristic morphological features, understanding phenotypic diversity, and revealing the presence of cryptic species (DeSalle, 2006, Foottit *et al.*, 2008, Hebert *et al.*, 2003, Ros and Breeuwer, 2007, Scheffer *et al.*, 2006). In comparison, population analysis relies on measurements of diversity between orthologous genes. These measurements allow investigation of invasion routes, population structure, distribution, human-aided versus natural dispersal paths, maternal origins, and biotypes (Armstrong and Ball, 2005, Gwiazdowski *et al.*, 2006, Kurose, 2005, Lee *et al.*, 2008, Scheffer *et al.*,

2006, Nadel *et al.*, 2009, Tsutsui *et al.*, 2001, Jenkins *et al.*, 2007). However, before population analysis can be employed the species needs to be correctly identified and selecting the correct marker needs to be considered.

Careful consideration must be applied when selecting a molecular marker, as different markers vary in their suitability for addressing particular questions. Often a marker is selected because it has been used in a similar study or is based on in-house experience (Armstrong and Ball, 2005). In other instances, markers are species specific and provide valuable information about the organism, but results may not be reproducible between laboratories, across similar taxa and to a lesser extent within the same taxon (Vignal *et al.*, 2002, Armstrong and Ball, 2005, Groth and Wetherall, 2000, Behura, 2006). This can make it difficult to select the most appropriate molecular marker to discriminate between species or populations.

While most molecular markers are selected using the criteria discussed above, there are three population genetic concepts that have to be considered before discussing the markers types, their usage, and issues relating to their use. These concepts are Founder Effect, Genetic Bottlenecks and Genetic Drift. Each of these concepts influence the pest species' chance of survival and have an effect on sequence based marker variability and subsequently on how the marker can be applied to analysing an invasive pest (e.g. Kalinowski *et al.*, 2010, Puillandre *et al.*, 2008, Tsutsui *et al.*, 2003, Tsutsui *et al.*, 2000)

The Founder Effect concerns the loss of genetic variation that occurs when a very small number of individuals from a larger population give rise to a new population which remains genetically isolated. Invasive species are very likely to display such an effect, as very low numbers of individuals typically succeed in forming a viable population that exploits the new environment. The new population may be distinctively different, genetically and phenotypically, from its parent population through loss of genetic variation and the non-random sub-sampling of the genes from the original population. These effects were first described by Ernst Mayr in 1952 when studying the limited human gene pools of Iceland, Easter Islander, those native to Pitcairn Island, and the Dutch founders of the Boer population in South Africa (Provine, 2004).

Genetic Bottlenecks typically occur when population size is reduced for at least one or more generations, which causes a decrease in allele frequency. In evolutionary history, such events can occur due to environmental factors such as ice ages and volcanic effects on climate, or periods of high mortality through disease or predation (Page and Holmes, 1998). Evidence for Genetic Bottlenecks is mainly from vertebrate studies where predation has reduced populations of animals to very low numbers of reproducing pairs. Examples have been seen in elephant seals (Hoelzel *et al.*, 2002, Weber *et al.*, 2000) and the east African cheetahs (O'Brien *et al.*, 1987). In the context of an invasive species, Genetic Bottlenecks appear to exhibit similar patterns to those seen in Founder Effects as a few individuals may initiate the new population.

The third population genetic concept, Genetic Drift or Allelic Drift, changes the relative frequency at which a gene variant (allele) occurs within a population (Futuyma, 1998) and can cause loss of genetic variation, particularly in small populations. The alleles in offspring are a random sample of those in the parents, and then chance has a role in determining whether a given individual survives and reproduces. This can change the allele frequency over time and may therefore be an important process in evolution. This contrasts to natural selection, where gene variants become more or less common depending on their influence on reproductive success (Avers, 1989). The changes due to genetic drift are not driven by environmental or adaptive pressures, and may be beneficial, neutral, or detrimental to reproductive success. Rates of Genetic Drift are increased by Founder Effects and Genetic Bottlenecks because the rate of drift is inversely proportional to the population size (Hartwell *et al.*, 2004).

In this Chapter the current issues in the selection and application of DNA markers used to study invasive insects are discussed. Particular emphasis is placed on mitochondrial and nuclear DNA markers, on the premise that these markers provide the best way for calculating divergence times (Hwang and Kim, 1999).

1.3 Mitochondrial Markers

The mitochondrial genome (mtDNA), in the majority of metazoan organisms, is a circular DNA molecule typically 15-20 kb in size (Hwang and Kim, 1999, Boore, 1999). This genome carries genes for 12 to 13 proteins, two rRNA, and 22 tRNA genes, which collectively build the metabolic machinery for oxidative phosphorylation of pyruvate within mitochondria to provide energy for the cells (Figure 1.1; Boore, 1999, Hwang and Kim, 1999). During replication of the mtDNA inefficiency of the DNA polymerase editing function causes new mutations to occur within the mtDNA (Brown *et al.*, 1979). Since the mtDNA is normally maternal inherited, new mutations are shared along the maternal lineage allowing the sequence information to be used to examination the phylogenetic structure of lower level classifications (families, genera and species; Hwang and Kim, 1999, Hebert *et al.*, 2003). Furthermore, conserved regions of sequence within families makes it possible to design universal primers for polymerase chain reaction (PCR) and sequencing (Hwang and Kim, 1999, Behura, 2006). The genes that are most commonly used are, in order of usage: the barcoding region Cytochrome oxidase I (*COI*; Hebert *et al.*, 2003), Cytochrome oxidase II, *16S*, and to a lesser extent *12S* and Cytochrome b (*CYT b*; Caterino *et al.*, 2000). This Chapter will investigate *COI* which is, debatably, the main DNA taxonomy region, as well as the *16S*, *CYT b* and the control region.

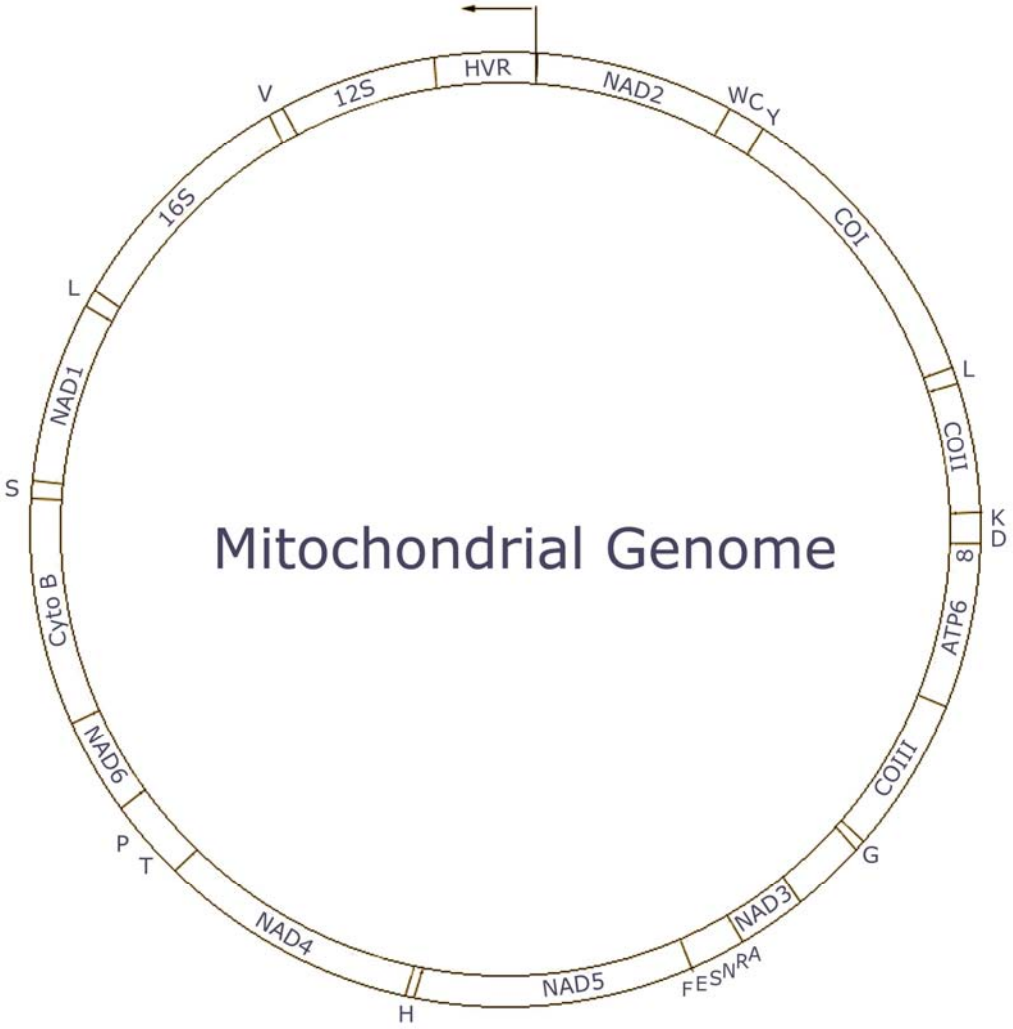

Figure 1.1. Mitochondrial gene organisation of *Drosophila yakuba* by Clary and Wolstenholme (1985). Direction of replication is represented by the arrow. Abbreviation: ATP6 & 8 (ATP synthase, subunits 6 & 8), COI, II & III (Cytochrome oxidase, subunits I to III), NAD1, 2, 3, 4, 5 & 6 (NADH dehydrogenase, subunits 1 to 6), 12S (small subunit of mitochondrial ribosomal DNA), 16S (large subunit of mitochondrial rDNA), tRNA- one letter amino acid abbreviation (parenthesis three letter amino acid abbreviation): A (ala), C (cys), D (asp), E (glu), F (phe), G (gly); H (his), I (ile); K (lys), L (leu), M (met), N (asn), P (pro), Q (gln), R (arg), S (ser), T (thr), V (val), W (trp), and Y (tyr).

1.3.1 The Cytochrome oxidase I Gene

Cytochrome oxidase I is the terminal enzyme in the mitochondrial respiratory chain (Michel *et al.*, 1998). The length of this gene is approximately 1500 bps and it is the most common mitochondrial gene used in molecular genetic studies. Two recent examples of barcoding studies using *COI* are the species identification of aphids (Aphididae; Foottit *et al.*, 2008) as well as cryptic and introduced moths (deWaard *et al.*, 2009). Population studies using *COI* have included eucalyptus (Schutze *et al.*, 2006, Nadel *et al.*, 2009) and Psyllid pests (Lee *et al.*, 2008), and the phylogeography of a spider mites (Tetranychidae; Ros and Breeuwer, 2007). Cytochrome oxidase I has increased in popularity due to the ease of designing universal primers (within a family), lack of insertions or deletion (indels) and its relatively high mutational rate (Hebert *et al.*, 2003, Armstrong and Ball, 2005). Indels, which may result in a shift in the reading frame, are usually detrimental to the individual and therefore rare in this gene. Lack of indels also helps overcome computational problems associated with correct sequence alignment (Hebert *et al.*, 2003). Mutational rates of approximately 1.1 to 2.3% nucleotide divergence per million years (Brower, 1994) or 0.011 to 0.023 substitutions per site per million years (s/s/Myr), makes this region more suitable for phylogeographical studies than other mitochondrial genes (Hebert *et al.*, 2003).

1.3.2 The 16S Gene

Two ribosomal subunits are encoded by the mitochondrial genome, *12S* and *16S* (Hwang *et al.*, 2000). These are two of the most highly conserved mitochondrial genes, with the substitution rates estimated at >0.001 s/s/Myr (Hwang and Kim, 1999, Simon *et al.*, 1994). *16S* has been shown to evolve slightly faster than *12S*, making it more suitable for calculating the evolutionary trajectories of species, genera and families (William and Piesman, 1994, Hwang and Kim, 1999). In comparison, *COI* is more suited to phylogenetic studies between and within species. The *16S* region has been used to examine the phylogenetic relationship between closely related species, such as the relationships among *Bactrocera* species (Smith *et al.*, 2003, Muraji and Nakahara, 2001) and *Reticulitermes* species (Marini and Mantovani, 2002). While this region is primarily used for comparing species within a

genus it has also been used as a population marker in mayflies and termites (Tojo and Matsukawa, 2003, Jenkins *et al.*, 2007).

1.3.3 The Cytochrome b gene

Cytochrome b is a component of the respiratory chain complex III involved in ATP generation (Berg *et al.*, 2001). The use of this gene has been well established in vertebrate studies for calculating evolutionary trajectories of species within the same genus and/or family (Castresana, 2001). Its use as a phylogenetic marker for invasive pests has revealed Genetic Bottleneck events for the invasive *Tecia solanivora* (Puillandre *et al.*, 2008) and origins of Argentine ants (*Linepithema humile*; Tsutsui *et al.*, 2001).

1.3.4 The Control Region

The control region (or Hyper Variable Region) is a non-coding area of mitochondrial genome that can vary in nucleotide composition and length (Hwang and Kim, 1999). The control region is often composed of >85% A+T bases with up to 96% in the honey bee (*Apis mellifera)* and fruit fly (*Drosophila melanogaster*; Zhang and Hewitt, 1997). Unlike the non-coding regions within the nuclear genome (e.g. Internal Transcribed Spacers) which show typically high mutational rates due to minimal evolutionary constraints (Eickbush and Eickbush, 2007), this is not the case for the control region. In some insect species the control region has been shown to be highly conserved, suggesting that the remaining G+C content may be functionally important (Zhang and Hewitt, 1997). Conservation of this region has made it suitable for differentiating multiple species within the *Anopheles gambiae* species complex (Caccone *et al.*, 1996). Alternatively, when under limited mutational constraint this region has the potential to mutate rapidly with substitution rates as high as 0.96 and 1.56 s/s/Myr (Lambert *et al.*, 2002, Hay *et al.*, 2008).

1.4 Nuclear Markers

Unlike the mainly maternally derived mitochondrial markers, nuclear markers provide genetic information from both maternal and paternal lineages. The most utilised nuclear gene for phylogenetic studies, particularly classification at the family level and higher, is the ribosomal array (rDNA; Jenkins *et al.*, 2007). Other nuclear markers and methods which have been used for phylogenetic reconstruction include Microsatellites, RAPDs (Random Amplification of Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism) RFLPs (Restriction Fragment Length Polymorphism) and allozymes (Vignal *et al.*, 2002). Microsatellites are excellent markers for examining variability between individuals in a population, as each microsatellite locus can display a high number of alleles in a population (Vignal *et al.*, 2002). However, species identification can be problematic when a microsatellite locus is ancient and shared between species (refer to as convergent and divergent evolution). Moreover, development of an array of microsatellite loci can be time-consuming. RAPDs are typically problematic because of their dominant nature, high numbers of artifactual bands result from PCR induced mutations, and results are rarely reproducible between laboratories (Behura, 2006; Vignal *et al.*, 2002; reviewed by Groth and Wetherall, 2000). While AFLPs are dominant in allelic nature, and time consuming (requiring restriction enzyme digest, ligation, PCR followed by scoring on specialised gels), they are highly reproducible and variable (Behura, 2006). Since DNA sequences can provide the best information for calculating divergence (Hwang and Kim, 1999) the use of rRNA arrays (*18S* and the Internal Transcribed Spacers) and Elongation Factor 1 Alpha (*EF-1a*) will be examined.

1.4.1 Ribosomal array

The rRNA transcriptional unit (or rDNA array) consists of three genes (*18S*, *5.8S* and *28S*), two internal transcribed spacers (*ITS1* and *ITS2*), two external transcribed spacers (*ETS*), an intergenic spacer (*IGS*), and a non-transcribed spacer (*NTS*; Figure 1.2; Eickbush and Eickbush, 2007, Hwang and Kim, 1999). Depending upon the species the number of rDNA arrays in a genome can vary, from a single set in the protist *Tetrahymena* to several thousand in the lizard *Amphiuma means* (Page and Holmes, 1998). In most individuals, variation among their rDNA arrays is minimal

due to the processes of concerted evolution (Eickbush and Eickbush, 2007, Weider *et al.*, 2005). The mechanisms behind concerted evolution are unequal crossing over, gene conversion, and gene amplification, which are collectively known as molecular drive (reviewed in Eickbush and Eickbush, 2007). Despite these conservation mechanisms the rDNA array contains regions that mutate at different rates (Weider *et al.*, 2005). The ribosomal genes are highly constrained, as they are responsible for protein production, whereas the non-coding internal transcribed spacers are less constrained and therefore evolve faster (Weider *et al.*, 2005). This makes the ribosomal array suitable for both species and higher order studies.

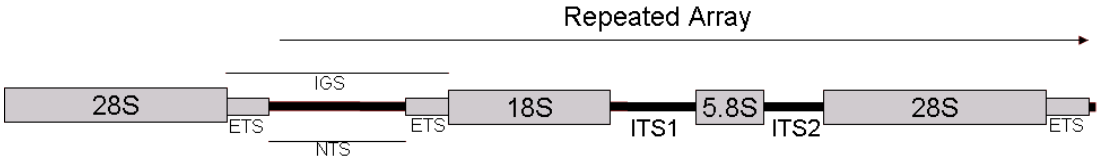

Figure 1.2. Schematic diagram of the ribosomal repeated array based. Each array consists of three genes (18S, 5.8S and 28S), two internal transcribed spacers (ITS1 and ITS2), two external transcribed spacers (ETS), an intergenic spacer (IGS), and a non-transcribed spacer (NTS). Arrow indicates the direction of transcription.

1.4.2 The 18S gene

The three ribosomal genes (*18S*, *5.8S* and *28S*) form part of the machinery responsible for protein production (Weider *et al.*, 2005). The *18S* comprises several variable domains, surrounded by highly conserved regions (Hwang and Kim, 1999). These slowly evolving regions are often used for higher level classifications (above the level of order) whereas the variable domains have been used for lower level classifications (Vogler *et al.*, 1997, Gillespie *et al.*, 2008). For example, Volger and Colleagues (1997) used the variable domain V4 to examine the phylogeny of tiger beetles (Carabidae: Cicindellinae).

1.4.3 The Internal Transcribed Spacer regions

Two Internal Transcribed Spacers are situated between the three ribosomal subunits all of which are expressed as part of a polycistronic mRNA. During biogenesis of the rDNA array the Internal Transcribed Spacers are removed. Consequently, the spacer regions evolve much faster than other nuclear rRNA regions, with new mutations or expansions occurring every few generations (Weider *et al.*, 2005). Since these mutations don't alter the structure of the nuclear subunits there are no detrimental effects on the organism (Hwang and Kim, 1999). The high mutation rate makes ITS1 and ITS2 some of the most commonly used regions at examining population structure and geneflow (e.g. (Tang *et al.*, 1996, Presa *et al.*, 2002, Lorenz-Lemkwe *et al.*, 2005).

1.4.4 Elongation Factor 1 Alpha

During translation, *EF-1a* is involved in the GTP-dependent binding of charged tRNAs to the acceptor site of the ribosome (Sanchis *et al.*, 2001). While the amino acid sequence of *EF-1a* is highly conserved, variation within introns has made this single copy gene suitable for both species and within species studies (Sanchis *et al.*, 2001). Studies involving the use of this gene include differentiating bee species (Apidae; Danforth and Ji, 1998), parasitoid wasps from the genus *Pausia* (Sanchis *et al.*, 2001), and fruit fly (*Bactrocera depressa*) populations (Mun *et al.*, 2003).

1.5 Marker Usage

To investigate the pattern of use of the *16S, COI, CYT b, Control Region, 18S, ITS* and *EF-1a* loci for phylogeny studies, a search between the years 1991 and 2009 was conducted using the journal databases ScienceDirect, Springerlink and Wiley, within the life sciences journals. The articles were required to have the keywords "phylogeny and/or phylogenetic" in any field. The results were then separated into individual genes by selecting those studies that had the gene name in the title, abstract or keyword fields (studies that used more than one gene were included and each gene counted). The datasets were further scrutinised by including the keywords "invasive insects" (in any field).

The data were plotted to show the number of studies over time and the gene use (as a percentage) for all phylogeny studies as well as for studies restricted to invasive insects (Figure 1.3). While the number of phylogeny studies increased between 1991 and 2009 (from 16 to 1443), the change in gene usage was minimal. The only noteworthy change was the gradual decrease in the use of *16S*, which decreased from 56% in 1991 to 33% in 2009. This can be attributed to changes of emphasis on the type of phylogenetic study, from family to species studies. The use of *CYT b* as a phylogenetic marker has been relatively stable since 1992, but its use as an invasive pest marker has decreased since it was first reported in 2002. The percentage use of *COI* has varied little since it was first proposed by Hebert *et al.* (2004) as the barcoding region for species identification (ranging between 22% and 29%). This was surprising considering the numerous papers that endorse the use of this gene and the number of *COI* sequences stored in Genbank (http://www.ncbi.nlm.nih.gov/; 89,046 as of 19/03/2010). For invasive pest species, an increase in the use of all other genes was observed, relative to the use of *COI*. This may be attributed to the problems associated with the use of only one molecular maker and will be discussed.

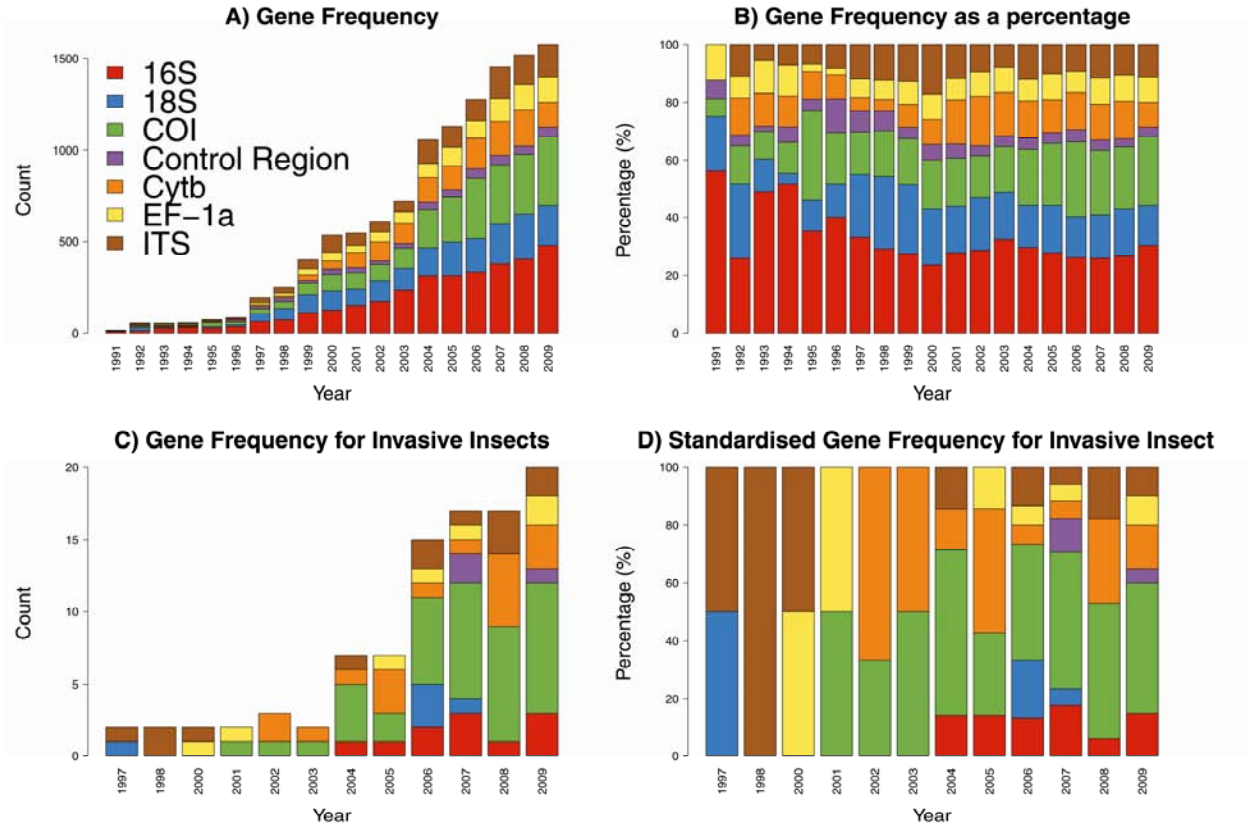

Figure 1.3. Frequency of use of the *16S*, *COI*, *Control Region*, *18S*, ITS and *EF-1a* loci for phylogeny studies between the years 1991 and 2009. A) Gene usage frequency for all types of phylogenetic studies. B) Gene usage frequency as a percentage data for all types of phylogenetic studies. C) Gene usage frequency for invasive pests. D) Gene usage frequency for invasive pests as a percentage.

1.6 Considering Each Molecular Marker

The studies that were described in the preceding text all measured the divergence between orthologous genes in order to either identify a species or analyse population structure. These types of analyses depend upon correct identification of the species being studied and thus, by default, the species boundary. The species boundary defines how a group of similar organisms are split into species and depending on the discipline (morphology, ecology or molecular phylogeny) this definition can change significantly. Traditional taxonomists group organisms based on similar or diagnostically significant morphological characteristics (Balakrishnan, 2005). Biological species are defined as reproductively isolated organisms and not phenotypic variants, although proving reproductive isolation is often difficult (Balakrishnan, 2005, Page and Holmes, 1998). A molecular species is defined by individuals that cluster together, either by lineage or through a character based analysis (Balakrishnan, 2005, Sites and Marshall, 2003). While DeSalle (2007) strongly supports the use of character based analyses over lineage, it is rarely used because of the problems in determining when a character becomes fixed within a species (Balakrishnan, 2005). In comparison, lineage analysis uses a database of DNA sequence to construct phylogenetic trees and genetic divergence values (Balakrishnan, 2005). Species are then delineated based upon the clustering or splitting of individuals. However, using lineage based methods as the sole means to determine species boundaries is difficult and can lead to incorrect inferences if results are misinterpreted or data is incomplete.

Using a standard percentage of pairwise divergences, for example, 2% or ten times the intra-specific variation for *COI*, has been suggested as a means to determine a species boundary (Hebert *et al.*, 2003). However, using these values as a cut-off for a species boundary may be problematic. One study using *COI* to identify introduced and cryptic moth species in Canada calculated a mean pairwise divergence between congeneric species of 9.50% (range of 1.54 to 15.32%) and within species of 0.25% (range of 0 to 3.59%; deWaard *et al.*, 2009). This resulted in a 2% overlap between intra and inter-specific diversity. A similar study in aphid species by Foottit and Colleagues (2008) calculated the average divergence between congeneric species at 7.25% (range of 0.46 to 13.1%) and within species at 0.2% (ranging from 0 to

1.27%). In these cases, using a standard percentage divergence or ten times intra-specific variation (which was in these cases ~2%) would have resulted in an incorrect assignment of individuals to a species. These two cases are not isolated (e.g. Jenkins *et al.*, 2007; Nadel *et al.*, 2010; Ros and Breeuwer, 2007).

Cognato (2006) reviewed the use of standard percentage pairwise divergence to delineate species. The main gene investigated was *COI*, which ranged from 0.04 to 26% and 0.15 to 25.7%, intra and inter-specific divergence respectively. Cognato (2006) concluded that in 39% of the cases the reliance on a standard percentage of 2% to differentiate species would provide an incorrect conclusion. Foottit and Colleagues (2008) observed 27 cases where different aphid species had divergences less than 2%. Particularly, two aphid species within the *Illinoia* genus varied by only 0.61%, but feed on different vegetation (conifers versus Spiraea; Foottit *et al.*, 2008). Yet, in a species of *Ips* defined on strong morphological evidence variation of 0 to 10% was observed (Cognato and Sun, 2007). These examples should raise considerable concern about identifying a species based solely upon sequence divergence, especially when divergence is <10%.

The mitochondrial *16S*, which is one of the most conserved mitochondrial genes, should show very little divergence within a species. However, this region seems to be more incongruent with even greater intra and inter specific divergence overlap. Intra and inter-specific divergence for the *16S* ranged from 0.1 to 2.4% and 0.3 to 11% respectively (Smith *et al.*, 2003, Cognato, 2006, Muraji and Nakahara, 2001, Marini and Mantovani, 2002, Jenkins *et al.*, 2007, Tojo and Matsukawa, 2003). While this divergence range was lower than the *COI*, *16S* is no better at resolving species relationships when divergence levels are low (Funk *et al.*, 1995, Lee *et al.*, 2008). One author suggested that *16S* is more suited to resolving taxa when combined with *COI*, if *COI* divergence is intermediate (approx 5-15%; Funk, 1999).

Successful species delineation and population analysis using *CYT b*, the control region and/or *EF-1a* depends upon the organism studied. The use of *CYT b* as a population marker has been successfully used to examine Genetic Bottleneck events in *Tecia solanivora* (Puillandre *et al.*, 2008), however, it was unable to differentiate between several introduced populations of the Argentine ant (Tsutsui *et al.*, 2001).

The *CYT b* data in the Argentine ant study indicated that each introduced population was derived from one main population and a more accurate population structure was only achieved after additional markers were included. Similarly, inconsistent results were also noted in studies attempting to delineate species. While the *Anopheles gambiae* species complex was separated into 5 species based upon variability within the mitochondrial control region (Caccone *et al.*, 1996), no length variation and minimal sequence divergence was observed within the mitochondrial control region in a study of six species of the butterfly (genus *Jalmenus*). This was despite phylogenetic studies showing very little sequence conservation across Lepidopteran genera (Taylor *et al.*, 1993). Several authors have reported conflicting results using *EF-1a* as well. One study on a parasitoid wasp from the genus *Pausia* showed variation within and between species supporting the discovery of three new species which was contradictory to existing morphological evidence (Sanchis *et al.*, 2001). In the case of bees, one study showed that the numbers of introns in *EF-1a* varied between species from one up to five (Danforth and Ji, 1998). While a single gene may provide suitable data to delineate or examine the structure of one group of species, in another it may not be as useful.

Not surprisingly, the utility of the multi-copy rDNA array as a phylogenetic marker is highly debatable. For the ITS regions, it is not the intra and inter-specific overlap that is the problem, rather the high levels of intra-individual variation (Vogler and DeSalle, 1994, Cruz *et al.*, 2006, Harris and Crandall, 2000). For example, Cruz *et al.* (2006) observed a mean intra-individual variation of about 5% (ranging from 0 to 13%) when studying a population of bees. In contrast, minimal intra and inter-specific variation (0.2 to 1.5% and 5.7 to 7.3% respectively) and no intra and inter-specific overlap was observed when investigating the *Anopheles* species complex (Fritz *et al.*, 1994). It seems as long as a high level of homogeneity is maintained within a species, the spacer region is a possible marker suitable for phylogenetic studies (Bower *et al.*, 2009). In contrast, the *18S* is highly conserved within a genome. While species boundaries have been identified using the *18S*, this region been shown to be incongruent with other molecular regions when trying to deduce phylogenetic trajectories (Vogler *et al.*, 1997). However, the use of rDNA runs counter to current practices for phylogenetic reconstruction (Alvarez and Wendel, 2003, Page and Holmes, 1998). With the number of arrays varying between

individuals, as well as species, it would be difficult to ascertain whether the sequences were orthologous or paralogous (Alvarez and Wendel, 2003, Page and Holmes, 1998). This may lead to inaccurate inferences being made, especially when intra-individual variation is high.

As the published data shows the use of molecular markers as the sole means to identify a species can often be difficult, this is especially true where inter-specific divergence percentages are below 2%. The difficulty in identifying a species is further increased when only mitochondrial genes are used, with problems arising from paralogy, resulting from the transfer of mtDNA into the nuclear genome (NUMTs), male-biased gene flow, and *Wolbachia* infections (Moritz and Cicero, 2004, Pamilo and Nei, 1988). Any of these factors could result in the clumping or splitting of a species or individuals within a population and underlines the importance of including morphological, biological, and ecological data to help define the species boundaries.

1.7 Analysis Issues

The foregoing discussion provides evidence that there are problems associated with the use of the most commonly used phylogenetic markers; in particular, how intra and inter-specific divergence values may affect the ability to identify species boundaries. However, selection and correct interpretation of the appropriate marker(s) is not the only reason that incorrect inferences may be drawn. There are six additional factors that are critical in ensuring correct and effective species identification and population analysis.

The use of single gene trees for the discovery or identification of a species has been shown to have high error rates and lead to incorrectly inferred identifications (DeSalle, 2006). This is especially true if the gene is not congruent with the true species tree. While NUMTs, male-biased gene flow, and inherited *Wolbachia* infections influence mitochondrial markers, lineage sorting and polytomy can affect all markers (Pamilo and Nei, 1988, McCracken and Sorenson, 2005). The process of lineage sorting is whereby three or more species radiate rapidly such that an ancient polymorphism is retained in all descendent lineages (McCracken and Sorenson,

2005). Genetic drift will lead to the monophyly of alleles at each locus in each species and by chance each species may then share a set of alleles at different loci (McCracken and Sorenson, 2005). If lineage sorting has occurred then separate analysis of each locus may generate different phylogenetic trees. Alternatively, a marker may lack phylogenetic signal and cause multifurication at a number of nodes (McCracken and Sorenson, 2005). Multifurication can occur when there is insufficient data (referred to as "soft" polytomy) or when real biological events such as rapid radiation occurs ("hard" or true polytomy; Maddison, 1989, Humphries and Winker, 2010). If only one gene has been used to examine the phylogenetic history of a group of species it can be extremely difficult to detect lineage sorting or overcome polytomy. Furthermore, an empirical study on the number of genes required to resolve a tree correctly has shown that as more taxa are added there is a reduction in tree confidence (Rokas and Carroll, 2005). Single gene analysis should not be relied upon as a means of species discovery, but rather a technique for determining divergence levels and help identify individuals that should be further scrutinised using traditional morphological taxonomy (Cognato and Sun, 2007).

Many scientists have taken up the proposal that "DNA barcodes" should to be lodged in global databases e.g. (Wheeler, 2008), such as Genbank, for use in global comparisons. Reviews of DNA sequences lodged in Genbank have revealed that some sequences have been attributed to the wrong species (Ros and Breeuwer, 2007, Wheeler, 2008), which has resulted in inaccurate assumptions in subsequent samples analyses. Misidentification can occur as a result of the presence of cryptic species, a lack of reliable morphological characteristics, taxonomic inexperience within the taxon, or museum specimens that have been poorly curated (Wheeler, 2008). Comparison of sequence data from specimens under investigation with sequences retrieved from Genbank should be an exploratory step, as any matches or mismatches may be in error, depending on the reliability of the source.

Small datasets of DNA sequences based on a few reference samples can lead to incorrect inferences and assumptions. For example, Ros and Breeuwer (2007) showed that a previous study on spider mites (Acari: Tetranychidae) had concluded there were two clades, one from the Mediterranean region and the other of mixed origin. When Ros and Breeuwer (2007) expanded the study by including a larger

sample size and subsequently additional sequences from the same gene, the two original clades were split into multiple clades. In this example, the initial phylogeographical patterns dissolved, indicating that in this DNA study sample size is critical to understanding the population structure (Ros and Breeuwer, 2007). The effect of sample size on the ability to accurately delineating species was investigated by Meyer and Paulay (2005) using an extensively sampled group of marine gastropods (>2,000 individuals in 263 taxa, representing >93% of the total known species). The authors found that even with a thoroughly sampled phylogeny, approximately 4% of samples were identified incorrectly using lineage based methods (Meyer and Paulay, 2005). When the size of the reference group was reduced, incorrect delineation increased. This was a result of unknown samples falling outside the existing species intra-specific divergence, and therefore, being grouped with another species or forming a novel clade (Meyer and Paulay, 2005). Moreover, the addition of more closely related samples showed an increase in overlap, resulting in less accurate identification of a sample. These documented cases indicate that accuracy is proportional to sample size and a range of closely related taxa, or samples from different geographical regions are required in order to accurately delineate a species.

Without a strong taxonomic understanding of a group of organisms, misidentification is inevitable. Imperfect taxonomy can create problems by increasing the range of the intra-specific variation, resulting in polyphyly and an increase in inter-specific overlap (Funk and Omland, 2003). This can result from plesiomorphic characters that are considered phenotypic variation and can result in the lumping of multiple species (Funk and Omland, 2003). Furthermore, morphological characteristics can be retained throughout history (parallelism) as well as arising independently (convergent evolution; Wiens *et al.*, 2003). This can give the impression that morphologically similar species are more closely related than they are at a genetic level. Under these circumstances molecular phylogenetics can be an extremely powerful tool in providing additional support to the biological or morphological disciplines by clumping or splitting individuals and supporting apomorphic or plesiomorphic characteristics.

In the case of invasive species, the genetic diversity of a population can be affected by the number of incursions that created the new population and how the species disperse. A single incursion would exhibit a Founding Effect with the new population containing a subset of the original population's alleles. A population resulting from multiple incursions would contain alleles from each original population. However, a population arising from a single event that contains multiple individuals may have a high genetic diversity that would be difficult to differentiate from the latter. It has been shown that a species intra-specific variation can also be affected by natural dispersal mechanisms. In a study on cowries, the authors showed that larvae lacking the planktonic larvae stage have a restricted dispersal mechanism and consequently, smaller intra-specific divergence ranges (mean 0.29%), whereas cowries that do disperse had an average intra-specific divergence of 0.7% (Meyer and Paulay, 2005). Knowing the local history of the invasive pest and sampling all known international population is an important step in understanding what has happened and may occur within a new environment.

The data and concepts discussed in this chapter highlight the need for careful consideration of what markers to use, including an understanding of their limitations. Many datasets and papers reviewed here have relied on a single marker gene or used a small population sample. The population samples may be restricted in two ways; first, a small number of individuals may be taken from a population; and second, a narrow biogeographical range may have been sampled. In the first case, there is no surety that the results reflect the variability within a population and in the second there is no guarantee that results are stable over a wide biogeographical range where a species may be present. Thus, the results may not reliably distinguish between intra and inter-specific variants of a species, a trait well known from classical alpha-taxonomy based on morphology.

1.8 The *Trogoderma* Problem

The Family Dermestidae is a group of mostly xerophilic, necrophagic beetles that scavenge for dried, proteinaceous materials such as dried fish, meat and hides (Kiselyova and McHugh, 2006, Zhantiev, 2009, Peacock, 1993). This type of material is consumed mainly during the larval stages, whereas the majority of adults forms feed on plant material, generally nectar and pollen, to help with egg maturation (Kiselyova and McHugh, 2006, Zhantiev, 2009). Exceptions do occur; for example, adults within the *Dermestis* genus do not feed on plant material and some species of *Trogoderma* and *Anthrenus* can complete their entire lifecycle on plant material making them pests of stored products (Mroczkowski, 1968, Hinton, 1945).

The *Trogoderma* genus contains some of the world's most serious pests of stored products. In particular, the Khapra beetle (*Trogoderma granarium* Everts) is one of the most serious pests of stored products and is ranked in the world's top 100 invasive species (Lowe *et al.*, 2000). This pest is highly invasive, with reports of infestations occurring in 52 countries (Banks, 1977). While *T. granarium* feed on most dried plant or animal matter they prefer cereals and cereal products, particularly wheat, barley, oats, rye, maize, rice, flour, malt and noodles (Hinton, 1945). It is a destructive feeder with reported weight loss of the stored product up to 30% and in extreme cases 70% (Szito, 2007b). Furthermore, it is difficult to eradicate because it has the ability to diapause for up to eight years (Burges, 1962). An introduction of *T. granarium* into Australia would result in an estimated loss of approximately $AU 1.83 billion annually due to the loss of Pest Freedom Status (Cook, 2003).

While Australia is currently free of *T. granarium*, the closely related exotic *Trogoderma variabile* Ballion has succeeded in invading and establishing in this country. This species was first discovered in Griffith, Australia, in 1977 and has since spread throughout the remainder of Australia (Hartley and Greening, 1983). In Australia, this species is regarded as a minor and persistent pest (Rees *et al.*, 2003). However, it is of considerable concern because it could mask the presence of the more damaging Khapra beetle, due to the morphological similarity between these two species (Figure 1.4; Rees *et al.*, 2003). Furthermore, the majority of *Trogoderma* species inhabit Australia (52 species, not taking into account the proposed synonymies: A. Szito, personal communications). Differences between *Trogoderma*

species include: setae colour, elytra pattern, number of segments that form the antennal club, and genitalia shape. These morphological features are, in most cases, diagnostically unique and are alone sufficient to identify adults of the species. However, adult specimens collected by quarantine or biosecurity agencies maybe missing some of these characteristics because of feeding by necrophagic larvae, including *Trogoderma*, or by physical abrasion. The movement of stored products can act like abrasive paper removing antenna, legs, elytra and setae from the adults. The loss of key morphological features can make identification uncertain. In the case of larvae, there are fewer unique morphological features, and physically damaged specimens can make identification very difficult.

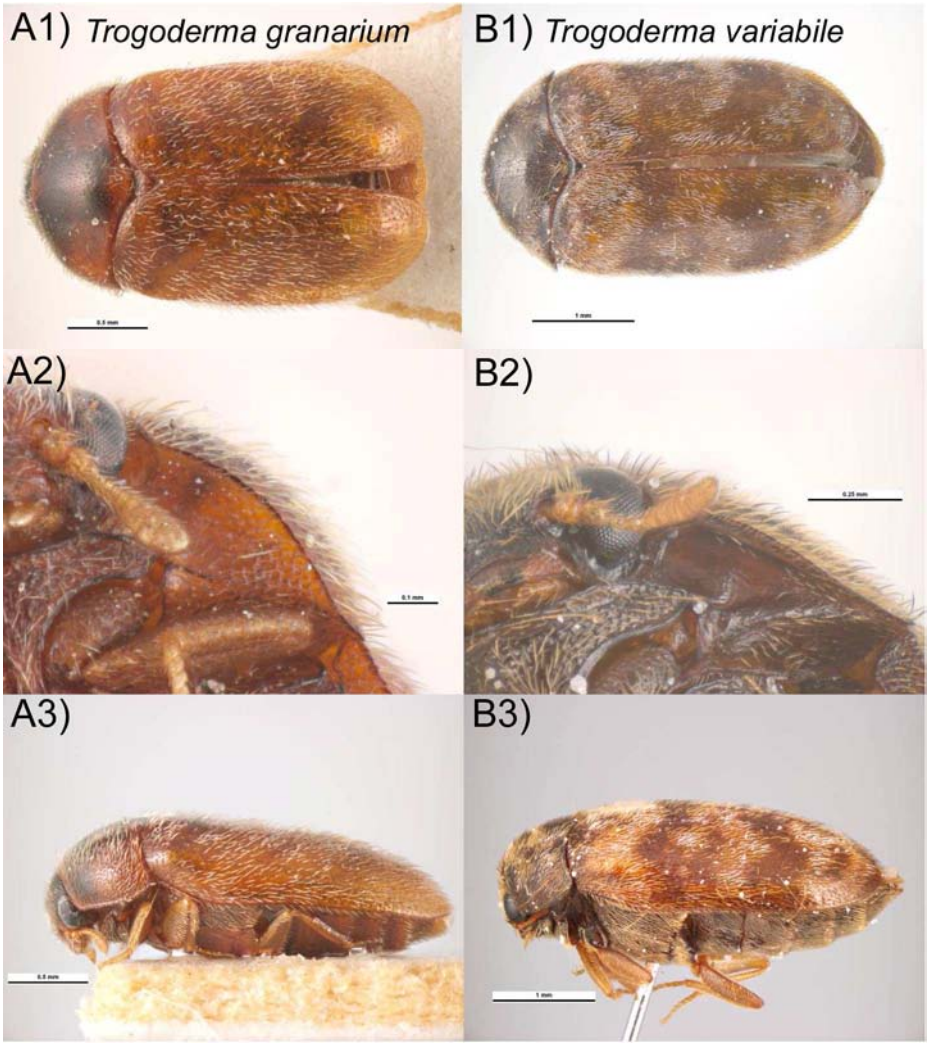

Figure 1.4. Photographs of A) *T. granarium* and B) *T. variabile*. 1) Dorsal photograph highlighting the difference in setae colour and pattern. 2) Antenna and antenna cavity. 3) Lateral view. Photographs courtesy of the Pest and Disease Image Library (PaDIL; http://www.padil.gov.au/).

The limitations of DNA markers, as discussed in foregoing sections, need to be taken into account when selecting a marker, or a combination of markers. At present no molecular study has examined the genetic relationships between *Trogoderma* species and closely related genera, and the current phylogenetic structure is based on morphological data (Table 1.1), which is often problematic. While the immediate interest is in reliably separating *T. variabile* and *T. granarium* at the species level, it is necessary to establish the phylogenetic relationship of these two species and the other related species and genera. This will ensure that any species intercepted by quarantine or biosecurity agencies can be identified using molecular markers. Clearly, using a range of markers to investigate these relationships should ensure that accurate genetic relationships between *Trogoderma* species and closely related genera are calculated

Table 1.1 Species used in this study and their current hierarchical nomenclature.

Order	Series	Family	Tribe	*Genus*	*species*	Australia/ Naturalised	Exotic
Coleoptera							
	Bostrichiformia						
		Dermestidae					
			Attagenini	*Attagenus*			E
			Megatomini	*Anthrenus*	*verbasci*	A (nat)	E
				Anthrenocerus	*australia*	A	
				Orphinus		A	
				Trogoderma	*angustum*		E
				Trogoderma	*glabrum*		E
				Trogoderma	*granarium*		E
				Trogoderma	*megatomoidies*		E
				Trogoderma	*ornatum*		E
				Trogoderma	*variabile*	A (nat)	E
				Trogoderma	*versicolor*		E
				Trogoderma		A	
	Cucujiformia						
		Tenebrionidae					
				Tribolium	*castaneum*	A (nat)	E

1.9 Hypothesis and Approach

Considering the foregoing discussion I hypothesise that molecular markers can differentiate *T. granarium* from *T. variabile* and other closely related species and genera.

DNA extraction methods often destroy the external physical characteristics essential for morphological identification, especially where whole specimens are macerated. Conversely, preparation of arthropods for morphological identification often damages or destroys DNA within the specimen. A non-destructive DNA extraction method was required for this project so that specimens can be re-examined after they have been genotyped. This is especially important if the genetic data is incongruent with the morphology, and where genetic data indicates a putative new species that then requires morphological characterisation to erect a new species name.

Acquiring samples, of invasive insects from overseas countries, can be problematic with countries either denying their presence or if they are classified as a common pest there is no interest in collecting them for a second party. Museum specimens seems to offer the easiest way in acquiring samples, although curators are often reluctant to allow DNA extractions for fear of damage to the specimens, and poor management and vouchering of specimens may led to mislabelled specimens (Wheeler, 2008). Furthermore, it is difficult and expensive to carry out a structured trapping in overseas countries, for an insect group such as *Trogoderma* that has serious economic and quarantine implications for that host country. For this reason the bulk of specimens use for this study were obtained in Australia and correctly identified by an experienced taxonomist/specialist to ensure correct diagnosis and a starting point for molecular phylogeny reconstruction.

To develop an understanding of the phylogenetic the relationship between *T. variabile*, *T. granarium*, and the closely related species and genera, three markers have been selected. These markers are associated with different rates of base conservation, ranging from at least one polymorphism per 500,000 years; one per 5-10,000 years; and one within 10-50 generations. These rates of conservation are postulated to correspond to genus, species and intra-species differences within a

taxon. Three genes, one nuclear and two mitochondrial, were selected for the phylogenetic studies. The first mitochondrial gene of choice was *COI* because it is frequently used to study the phylogeny of closely related species (Foottit *et al.*, 2008, deWaard *et al.*, 2009, Hebert *et al.*, 2003). Furthermore, the availability of universal primers by Simon *et al.* (1994) and Lunt *et al.* (1996) provides a quick first pass analysis and generation of a large dataset relatively quickly from trapping programs or museum collections. Cytochrome b was also included because it evolves more slowly than *COI* (Castresana, 2001) to inform intermediate tree nodes. Finally, the entire rDNA *18S*, which is a slowly evolving gene normally used for higher classification studies (Hwang *et al.*, 2000). The *18S* included the variable domains that have been shown to help resolve genera (Hwang *et al.*, 2000, Gillespie *et al.*, 2008). The aim to resolve structure across the phylogenetic tree by including highly conserved and variable domains within the nuclear gene, as well as faster evolving mitochondrial genes.

Finally, the phylogenetic data was used to develop a novel diagnostic test. In a biosecurity and quarantine context, a diagnostic test should ideally have the following capabilities: rapid, cheap, high throughput capacity, and foremost accuracy. High throughput is a critical requirement as *Trogoderma* traps can contain >100 specimens. To meet these requirements a multiplex quantitative PCR test was developed that included species specific primers and probes, an internal amplification control, and a high resolution melt curve analysis step. These components were evaluated to determine if the incidence of false positives and negatives can be reduced.

1.9.1 Aims and Objectives

The following 4 aims were used to address the main hypothesis:

1. Develop a non-destructive DNA extraction method that is rapid, cost effective, with high throughput capacity, and without damaging important morphological characters. To meet the preceding criteria we evaluated the efficacy of EDNA Hi SpEx by testing samples with that were (i) different in age and (ii) stored in a range of mediums.

2. A phylogenetic understanding of the Australian population of the invasive stored product pest *T. variabile* in Australia to determine: (i) if the Australian population is the result of multiple incursions or a single incursion, (ii) determine whether their wide spread distribution is the result of human aided or natural dispersal, and (iii) detect the presence of cryptic species. This information may help model possible dispersal routes for *T. granarium.*

3. Evaluate the phylogenetic grouping of *Trogoderma* (in species both exotic and endemic to Australia) using a molecular phylogenetic approach. This includes the elucidation of the molecular evolutionary relationships between four closely related genera (*Anthrenocerus* Arrow, *Anthrenus* Müller, *Attagenus* Latreille, *Orphinus* Motschulsky and *Trogoderma* Dejean) to understand their current biogeographically distributions. This in turn may provide an understanding as to why *T. variabile* and *T. granarium*, originally from the Palaeartic and Indian regions, are pests of stored products while the Australian natives are not. If the phylogenetic trees show that each of the genera are monophyletic and the pest species and Australian *Trogoderma* form distinct lineages, then using a DNA taxonomy approach to identifying dermestids that are intercepted by quarantine agencies may be feasible.

4. Develop a novel diagnostic test that will increase efficacy and have minimal incidences of false negatives and positives. To meet these criteria we tested the diagnostic markers on the following: (i) the specimens that were used to examine the population of *T. variabile* in Australia, and (ii) specimens that were intercepted by Australian Quarantine and Inspection Service (AQIS) and US Customs and Border Protection.

Chapter 2

General Materials and Methods

2.1 DNA Extraction

DNA was extracted non-destructively using EDNA HiSpEx (Fisher Biotec, Perth). A master mix of EDNA HiSpEx was made by mixing Solution 1A and Solution 1B together at a ratio of 39:11. All dermestid specimens were immersed in 50 µl of EDNA mix and then heated at 99 ºC for 20 min. After incubation, a 0.25 volume of EDNA Solution 2 was added to each tube. Solution 2 and Solutions 1A and 1B were mixed by gentle pipetting and an aliquot was transferred to a sterile microcentrifuge tube and stored at -20 ºC.

2.2 DNA amplification by Polymerase Chain Reaction

2.2.1 Polymerase Chain Reaction Mixture

One of the main components of the experimental work involved the amplification of genomic regions for phylogenetic analysis. While each of the primers varied between chapters the Polymerase Chain Reaction (PCR) reaction master mix was universal, unless otherwise stated. PCR volume was 25 µl, including 2.5 µl of EDNA extracted DNA (1:10 dilution DNA:water) which was the equivalent of 20 to 25 ng/µl. The reaction mix comprised 1 × polymerase buffer (Roche), 1.8 mM $MgCl_2$, 0.2 µM of each primer, 200 µM of each dNTP, and 0.5 U of Faststart Hi Fidelity Taq polymerase (Roche).

2.2.2 Primers and PCR parameters for the Phylogenetic Studies

Two mitochondrial genes, *COI* and *CYT b*, and the nuclear gene *18S* were amplified for the phylogenetic studies (Table 2.1). Due to the size of the *18S* fragment (~1850bp) it was amplified in four sections. The four sections were amplified using primers that were designed by Primer 3 (http://frodo.wi.mit.edu/primer3/) to target conserved regions of previously published beetle *18S* sequences (GENBANK Accession Numbers AY748111, AY748105, AY748103, EF213875, EF213892, EF362981). Each fragment overlapped by 50 to 100bp. Universal primers were used to amplify *COI* and *CYT b*. The universal *CYT b* primers failed to amplify on *T. granarium* DNA. Consequently new primers were designed specifically for this

species using conserved regions within *Cyt b* sequences that were generated from the samples analysed in chapter 3 and using the program Primer3 (Table 2.1). Annealing temperatures and extension times varied as shown in Table 2.1. The thermocycler conditions were: 95 ºC for 10 min; 40 cycles of: 95 ºC for 30 s, annealing for 30 s, and 72 ºC for extension; with a single final extension period of 72 ºC for 5 min.

2.3 Electrophoresis

Quality and quantity of the amplified PCR products was determined by agarose gel. The concentration of agarose gel was 1.5% w/v and made up with 1 × TAE buffer. Electrophoresis of the agarose gel was carried out for 1 hr at 60 Volts, then stained with ethidium bromide and examined under UV light.

2.4 DNA Purification

PCR products were purified by the addition of 10 U of Exonuclease I (NEB) and 2.5 U of Antarctic Phosphatase (NEB) and incubation for 30 min at 37 °C followed by inactivation by heating the reaction to 80 °C for 20 min. Purified DNA was stored at -20 ºC until required.

2.5 Sequencing

Sequencing of the amplified genes was carried out by Macrogen Inc (Korea) using an Applied Biosystems ABI 3730 48-capillary DNA analyser using Big Dye Terminator Technology according to the manufacturer's protocols (Applied Biosystems). If required the PCR products were purified by Macrogen Inc (Korea).

2.6 Analysis of Sequence Data

Sequences were edited using CodonCode Aligner 3.0.3 (CodonCode Corporation) and aligned using the built-in version of CLUSTALW (Thompson *et al.*, 1994). If required the alignment were optimised by eye. Parsimony and Maximum Likelihood (ML) trees and Partition Homogeneity Test (PHT) were performed using PAUP 4.0 (Swofford, 2003). Bayesian trees was generated using Mr Bayes 3.1.2 (Ronquist and Huelsenbeck, 2003). MEGA 4 (Tamura *et al.*, 2007) was used to calculate net divergences between individual sequences, clades, and calculated transition / transversion ratios, and TCS v1.21 (Clement *et al.*, 2000) for network analysis. The criteria for each of these programs varied between chapters and are described within.

Table 2.1. PCR primers used to amplify the mitochondrial and nuclear genes.

Primer No.	Genome	Target Gene	Primer Name	Primer Sequence 5`-3`	Aneal Tm (ºC)	Ext Time (s)	Exp Size (bp)	Reference
1	Nuclear	*18S*	18SF1	TGCTTGTCTCAAAGATTAAG	45.1	30	500	This Study
			18SR1	GTATCGTTATTTTTCGTCAC				
2	Nuclear	*18S*	18SF2	TACCACATCCAAGGAAGG	45.1	30	500	This Study
			18SR2	CCTCTAACGTCGCAATAC				
3	Nuclear	*18S*	18SF3	TCTATTTTGTTGGTTTTCG	45.1	30	600	This Study
			18SR3	AGAACATCTAAGGGCATC				
4	Nuclear	*18S*	18SF4	CTTCTAGCCGCACGAGAT	45.1	30	400	This Study
			18SR4	CCTTGTTACGACTTTTACTTCC				
5	mtDNA	*COI*	C1-J-2183	CAACATTTATTTTGATTTTTTGG	43.7	30	800	(Simon *et al.*, 1994)
			TL-2-3014	TCCAATGCACTAATCTGCCATATTA				
6	mtDNA	*CYT b*	CB-J-10933	TATGTACTACCATGAGGACAAATATC	43.7	60	500	(Simon *et al.*, 1994)
			CB-N-11367	ATTACACCTCCTAATTTATTAGGAAT				
7	mtDNA	*CYT b T. granarium*	Tg-J-Cytb	GCTAATGGAGCCTCATTCTTC	43.7	60	500	This Study
			Tg-N-Cytb	TGAATTGGCGTAACTAATGGGT				

Chapter 3

Multiple Incursions and Putative Species Revealed using a Mitochondrial and Nuclear Phylogenetic approach to analysing the *Trogoderma variabile* Trapping Program in Australia

The Warehouse beetle, Trogoderma variabile *(Coleoptera: Dermestidae), is an internationally significant invasive pest of packed goods and stored grain. When it was first documented in Australia at Griffith, New South Wales, in 1977 an eradication campaign was initiated. After several years and considerable effort the eradication campaign was abandoned. To monitor the presence and spread of* T. variabile*, surveys were carried out by government agencies, in 1992 and 2002. When survey data were compared, it was concluded that the distribution of morphologically identified* T. variabile *had doubled in most Australian states. Here, samples from the 2002 survey were used to conduct a phylogenetic study using partial sequences of mitochondrial genes Cytochrome Oxidase I and Cytochrome B, and the nuclear gene 18S, to examine the distribution and dispersal of* T. variabile *and detect the presence of misidentified species. Based on the molecular results only 47% of the samples analysed were* T. variabile *and the remainder were a mixture of six putative species. In addition,* T. variabile *was found in only 78% of the trapping sites. The importance of correct diagnosis in relation to the eradication campaign is discussed. This chapter has been peer reviewed and published in the journal of Bulletin of Entomology and is attached in appendix II.*

3.1 Introduction

When *T. variabile* was first documented in Australia at Griffith, New South Wales, in 1977 and then again in 1979 on the other side of the continent at Morawa, Western Australia, Australian State and Federal agricultural departments attempted to control and eradicate this pest (Wright, 1994, Hartley and Greening, 1983, Rees *et al.*, 2003). After several years the eradication campaign was terminated because surveys showed that the beetle had spread to surrounding areas (Hartley and Greening, 1983). Although the eradication program was cancelled, surveys have been routinely carried out to monitor the distribution of the pest. One such extensive survey, conducted by The Commonwealth Scientific and Industrial Research Organization (CSIRO) from 2001 to 2003, discovered that in Victoria and South Australia the distribution of *T. variabile* had doubled since 1990, suggesting that little progress had been made in limiting its spread (Rees *et al.*, 2003).

In order to understand the apparently uncontrollable dispersal of *T. variabile* throughout Australia there is a need to better understand the mechanisms underlying their invasion biology. Without a detailed understanding of these mechanisms, efforts to control the dispersal of an organism can be rendered useless (Tsutsui *et al.*, 2003, Tsutsui *et al.*, 2000, Tsutsui *et al.*, 2001, Schutze *et al.*, 2006). Studies on other invasive insect species have shown that molecular markers can provide valuable information about population structure, gene flow and dispersal pathways (Mikac and FitzSimmons, 2010, Mikac and Clarke, 2006). In turn, this can provide an insight into the dispersal of species and help determine the presence of cryptic species (Loxdale and Lushai, 1998, Mikac and FitzSimmons, 2010, Mikac and Clarke, 2006).

In this study three genes were selected, two mitochondrial and one nuclear, as tools to investigate the population structure, dispersal pathways of *T. variabile*, and determine the presence of any cryptic species. The mitochondrial gene *COI* is frequently used to understand mechanisms underlying invasion biology (Nadel *et al.*, 2009, Jenkins *et al.*, 2007). Therefore, *COI* was selected to resolve patterns of geographical distribution and determine if the dispersal of *T. variabile* throughout Australia has been the result of multiple incursions and/or subsequent human-aided dispersal. Dispersal due to human activities is likely because the distance between grain storage facilities within Australia ranges from 9 to 3760km, which far exceeds the estimated average dispersal distance for individual *T. variabile* of only 75m (Campbell and Mullen, 2004). We also wanted to test the possibility that samples identified morphologically as *T. variabile* could in fact be *Trogoderma* species endemic to Australia, as there are currently 60 described endemic species (Booth *et al.*, 1990). While *COI* is well-suited for examining intra and inter-specific variation at both the species and genus level, *CYT b* was also included in this study because it is a gene that evolves more slowly than *COI*. A partial *18S* fragment was also included, to help resolve deeper nodes and check for congruence between the mitochondrial and nuclear genomes.

3.2 Materials and Methods

3.2.1 Collection

One hundred and forty one specimens were collected from 27 grain storage sites throughout Australia, between 2001 and 2003 (Table 3.1), using baited sticky flight traps, as described by Rees *et al.* (2003). These specimens were identified using morphological techniques that were not recorded and unknown to this project. Ten additional specimens were collected from nine Western Australian sites in 2007 as part of an ongoing trapping program (Table 3.1). These ten samples were the only samples that were verified as *T. variabile* using traditional morphological keys (Szito, 2007a), and were subsequently defined as representative *T. variabile* reference samples for this study. Single specimens of *T. granarium* and *Anthrenus verbasci* Linnaeus, both closely related non native pest species, were taken from the Department of Agriculture and Food Western Australian insect collection and used as outgroups. The *T. granarium* sample was intercepted by quarantine officers and was not part of a local population.

Table 3.1. Details of the number of samples analysed, putative species (based on DNA Taxonomy using either *COI* or *Cytb*), and the location of the 35 grain storage facilities.

		Putative Species									
Location	n	*T. variabile*	B	C	D	E	F	G	State	Latitude	Longitude
Ariah Park	6	2		1	3				NSW	-34.332	147.218
Balaklava	6	6							SA	-34.147	138.416
Booroowa	2			2					NSW	-34.482	148.751
Bowman	1	1							NSW	-33.15	151.292
Caracabal	5	2			3				NSW	-33.795	147.65
Carnanmah *	3	3							WA	-29.69	115.884
Colleambally	4				3		1		NSW	-34.799	145.785
Coolaman	8	7			1				NSW	-34.848	147.185
Cowra	7			2	3	2			NSW	-33.819	148.658
Crystal Brook	4	4							SA	-33.353	138.205
Dalwallinu *	1	1							WA	-30.249	116.669
Dimboola	8	1		5		2			Vic	-36.454	142.028
Garrah	2	2							SA	-35.483	140.3
Geelong	3			3					Vic	-38.149	144.357
Geraldton *	1	1							WA	-28.78	114.613
Griffith	7	3					4		NSW	-34.282	146.035
Katanning *	1	1							WA	-33.415	117.333
Little desert NP	8			8					Vic	-36.544	141.84
Lockhart	6				2		4		NSW	-35.215	146.768
Manangatang	8	7		1					Vic	-35.051	142.882
Narrabri West	3	3							NSW	-30.333	149.767
Nhill	6			6					Vic	-36.333	141.651
Pithara *	1	1							WA	-30.389	116.667
Port Adelaide	10	1	5	4					SA	-34.846	138.503
Quirindi	1	1							NSW	-31.5	150.683
Rainbow	4	1		1	1		1		Vic	-35.9	141.993
Swan Hill	8	8							Vic	-35.339	143.555
Tamworth	6	6							NSW	-31.091	150.932
The Rock	4	2			1		1		NSW	-31.865	149.685
Three Springs*	1	1							WA	-29.536	115.761
Varley	5							5	WA	-32.706	119.641
Wagin *	1	1							WA	-33.187	117.208
West Wyalong	7	2			3	2			NSW	-33.932	147.201
Willow Tree	2	2							NSW	-31.649	150.735
Wyalkatchem *	1	1							WA	-31.179	117.382
Total	151	71	5	33	20	6	11	5			

Notes. n = total number of samples analysed. * Samples from the 2007 trapping program

3.2.2 DNA extraction

The number of specimens used in this study ranged from 1 to 8 per site, dependent on the number of specimens collected. DNA was isolated from the whole individual by crushing them in 5% Chelex beads (Biorad) following the methods of (Walsh *et al.*, 1991.

EDNA HiSpEx (Fisher Biotec, Perth) was used to extract DNA from the ten *T. variabile* samples collected in Western Australia and the single specimens of *T. granarium* and *A. verbasci* following the method described in Chapter 2.1.

3.2.3 Amplification of *COI*

A semi-nested Polymerase Chain Reaction (PCR) was used to amplify a section of *COI* using the primers UEA 5 (5` AGTTTTAGCAGGAGCAATTACTAT 3`) and UEA 10 (5` TCCAATGCACTAATCTGCCATATT A 3`) for the first round of PCR (PCR1) followed by amplification of first round PCR product using the primer combination UEA 7: (5` TACAGTTGGAATAGACGTTGATAC 3`) and UEA 10 in a second PCR amplification (PCR2;(Lunt *et al.*, 1996). PCR were carried out in 12.5 µl volumes using 25-50 ng DNA, which was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc; PCR1) or 1 µl of PCR product (PCR2), 0.3 µM each primer (PCR1: UAE 5 and UAE 10; PCR2: UAE 7 and UAE 10), 0.2 mM each dNTP, 1 × PCR buffer (New England Biolabs), 3 mM $MgCl_2$, and 1.5 U of Taq DNA polymerase (New England Biolabs). PCR1 cycling conditions consisted of an initial denaturation step of 94 ºC for 3 min, followed by 30 cycles of 94 ºC for 40 s, 54 ºC for 1 min 40 s, and 72 ºC for 60 s, with a final extension temperature of 72 ºC for 8 min. PCR2 cycling conditions consisted of an initial denaturation step of 94 ºC for 3 min followed by 40 cycles of 94 ºC for 40 secs, 52 ºC for 1 min 40 s, and 72 ºC for 60 s, with a final extension temperature of 72 ºC for 8 mins. Ten percent of the final volume of the PCR2 products were electrophoresed on 2% w/v agarose gel containing 1 × TBE buffer and visualized with SYBR safe (Molecular Probes Inc.). Unpurified PCR products were sent to Macrogen (Korea) for cleanup and sequencing (Chapter 2.5).

3.2.4 Amplification of *CYT b* and *18S*

Amplification of the *CYT b* and a partial *18S* fragment (primers *18S*F2 and *18S*R2; Chapter 2 Table 2.1) were performed using the Reaction volumes, PCR reaction mixture, PCR cycling conditions, DNA purification, electrophoresis, and sequencing methods described in Chapter 2. The only alteration was to the PCR reaction mixture whereby the concentration of $MgCl_2$ was increased to 3 mM.

3.2.5 Data Analysis

Sequences (Accession numbers HM243239 - HM243470) were edited using CodonCode Aligner 3.0.3 and aligned using the built-in version of CLUSTALW. PAUP 4.0 was used to perform a PHT to determine the level of congruence between the three genes. The three genes were concatenated and PAUP 4.0 used to generate Parsimony and ML gene trees. In addition, PAUP 4.0 was used to generate a *COI* parsimony gene tree which was compared to a Bayesian *COI* tree generated by Mr Bayes 3.1.2. MEGA 4 was used to calculate net divergences between clades, TCS v1.21 for network analysis and an R statistics (Team, 2007) function that I wrote to plot the distribution of each clade as well as the *T. variabile* haplotypes. The PHT of the three genes was performed using the parameters hsearch, randomseed=0 and nrep=1000. The Parsimony and ML analysis of the three concatenated genes was calculated under the following conditions: hsearch, addseq=random, nrep=1000, swap=TBR, MaxTrees=1000, with the additional parameters added for the bootstrapping analysis nchuck=5 and chuckscore=1. The ML tree construction used likelihood settings that were selected by Modeltest 3.7 (Posada and Crandall, 1998) from the best-fit model (GTR+I+G) and were Lset Base=(0.2884 0.1771 0.2319); Nst=6; Rmat=(0.1618 10.5497 2.0353 0.3522 1.6956); Rates=gamma; Shape=1.8432; and Pinvar=0.6063. Bayesian analysis was performed on the *COI* sequence data using the evolution model 4by4, gen=10,000,000, sample freq=100, sump burnin=2,500, and sumt burnin=2,500, the posterior probability values were compared to the Parsimony bootstrapping results. Parsimony analysis of the *COI* data used the same parameters as concatenated gene analysis. *Anthrenus verbasci* was the outgroup for the generation of all trees.

3.3 Results

3.3.1 Sequence Data

The *COI* gene fragment (620 bp, of which 264 were parsimony-informative) exhibited 46 haplotypes in 87 individuals, with an average sequence divergence of 16.12% (S.E 0.87%). In comparison, *CYT b* (434 bp of which 171 were parsimony-informative) produced fewer haplotypes (24 in 112 individuals), but had a similar average sequence divergence of 14.73% (S.E. 0.95%). The number of sequences obtained varied between the mitochondrial genes as the amplification of each gene was unsuccessful in some individuals, therefore each gene was analysed separately. For *COI* and *CYT b* the polymorphisms resulted in 79 (38%) and 33 (22.9%) non-synonymous changes in the amino acid sequence, respectively. The *18S* nuclear fragment length was 518bp and, as expected, was evolving more slowly than *COI* and *CYT b*, with only 7 polymorphic nucleotides in 19 individuals. The average sequence divergence for this gene was only 3.9% (S.E. 0.17%). No insertions or deletions were observed in any gene fragment.

3.3.2 *COI* tree

Since the *COI* gene was the most phylogenetically informative fragment (46 haplotypes and 264 polymorphic nucleotides) it was used to generate both Parsimony and Bayesian trees (Figure 3.1). The *COI* tree revealed nine distinct clades (denoted using a unique letter or reference name) that were separated by deep branch lengths and supported by high net divergence between clades (from 10.72% to 26.11%; Table 3.2). All clades were highly supported by Bayesian posterior probability values greater than 0.71. Parsimony bootstrapping values were slightly less supportive for all clades (6 to 8% less), except the divergence of clade c from d and e, and clade f from g which were more than 22% different (53% and 77%, respectively). The *T. variabile* reference specimens that were identified using classical morphological techniques formed a distinct well supported clade with only 47% of the samples. The pairwise divergence between *T. variabile* and the other clades ranged from 18.88 to 25.68% (ca. mean 22.88%, S.E. 2.11%), the lowest of these values was a comparison with the *T. granarium* reference sample.

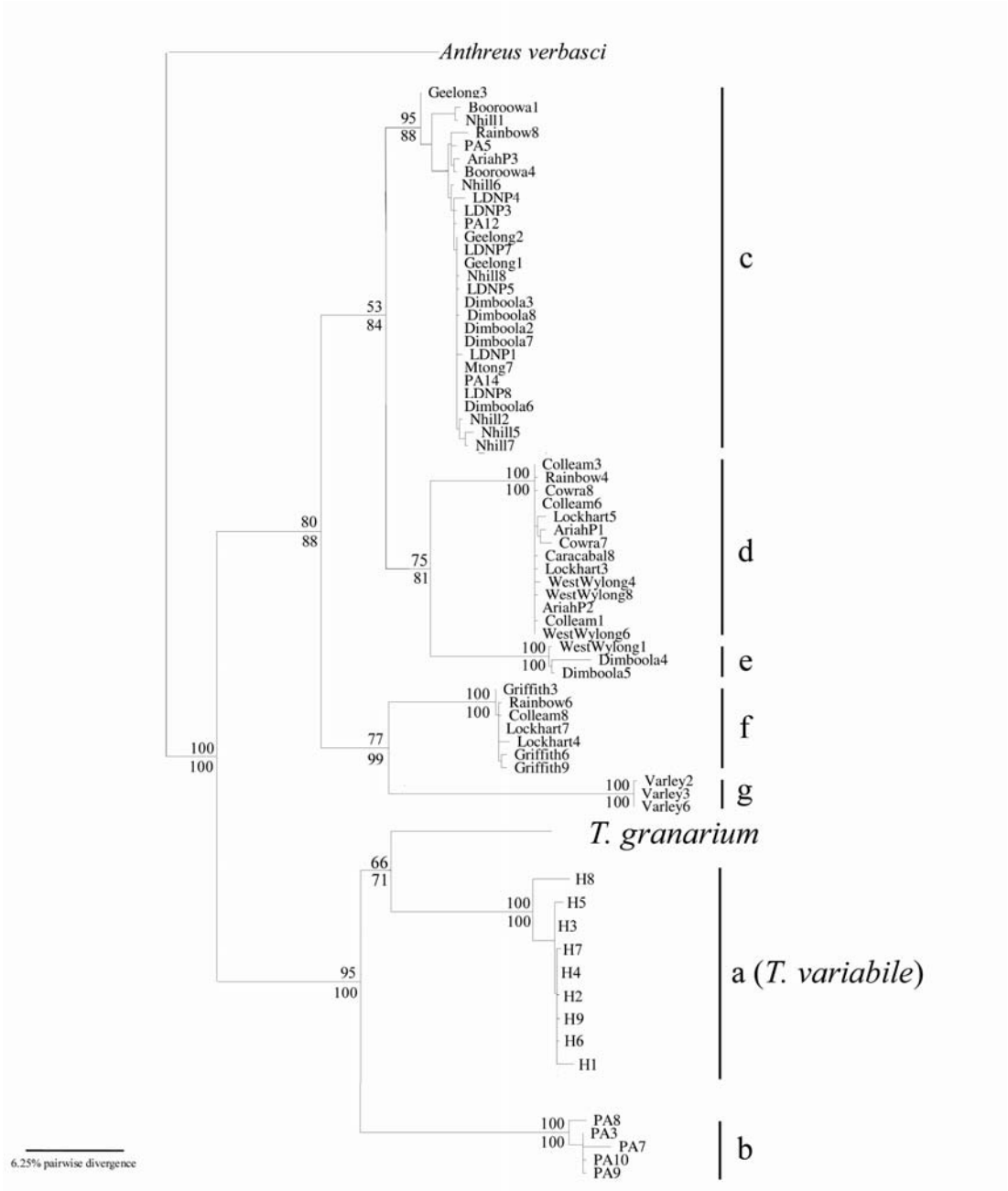

Figure 3.1. Cytochrome oxidase I Parsimony and Bayesian tree. The value above the branch represents the parsimony bootstrapping support, the Bayesian posterior probability values are shown below the branch.

Table 3.2. Inter and intra-specific pairwise divergence percentages for *COI* (above the Diagonal) and *CYT b* (below the diagonal).

		COI Intra				*CYT b* Inter								
		n	min	max	mean	A.v	b	c	d	e	f	g	T.g	T.v
CYT b Intra	n					1	3	23	14	6	7	5	1	54
	min					-	0	0	0	0	0	0	-	0
	max					-	1.84	3.46	0.69	0.46	0.46	0	-	0
	mean					-	0.82	0.62	0.15	0.20	0.23	0	-	0
COI Inter	A.v	1	-	-	-		26.11	22.6	21.39	20.16	21.3	21.66	20.83	22.58
	b	3	0	1.61	0.72	26.11		23.7	23.92	24	22.91	21.35	23.55	22.27
	c	25	0	3.55	0.96	23.71	24.57		15.28	13.43	18.43	23.11	13.65	23.11
	d	11	0	1.29	0.40	23.14	25.48	10.72		10.61	16.16	20.85	15.68	19.99
	e	2	0	0.97	0.48	25.89	22.66	11.34	13.15		15.65	19.93	15.23	19.93
	f	4	0	0.65	0.30	24.44	23.43	13.9	16.28	16.21		21.66	15.54	18.76
	g	3	0	0.16	0.08	24.52	23.47	14.85	15.69	16.61	12.34		20.61	20.18
	T.g	1	-	-	-	25.65	21.56	21.18	23.43	22.61	23.25	22.04		18.28
	T.v	34	0	2.90	0.40	24.84	21.81	21.59	25.68	23.39	23.22	23.65	18.88	

Note; Abb. A.v: *A. verbasci*, T.g: *T. granarium*, T.v: *T. variabile*

3.3.3 Mitochondrial Gene Analysis

While only the *COI* gene was used to generate the Parsimony and Bayesian trees, the two mitochondrial genes were used to examine the intra and inter-pairwise divergence between clades (Table 3.2). Inter-clade divergence values for *COI* and *CYT b* ranged from 10.72 to 26.11% and 10.61 to 26.11%, respectively. The mean divergence between clades was 1.61%, and in two cases *CYT b* was more divergent than *COI* (clades c and g). The mean intra-clade variation for *CYT b* was 0.29%, almost half that of the *COI* value of 0.48%, with *COI* and *CYT b* ranging from 0 to 3.55% and 0 to 3.46%, respectively. Interestingly, *CYT b* intra-clade variation for the *T. variabile* clade was zero (n=54), while the divergence for *COI* ranged from 0 to 2.9%. Unfortunately, Tamworth7 which was the most divergent specimen within the *T. variabile* clade (2.9%) was not represented in the *CYT b* dataset.

3.3.4 Congruence between Nuclear and Mitochondrial genomes

To determine if the nuclear and mitochondrial genomes were congruent, the *18S* fragment was amplified from 19 individuals, which represented specimens from each clade identified from the *COI* data. The PHT showed congruence between the nuclear gene *18S* and the two mitochondrial genes *COI* and *CYT b*, (p-value = 1 and 0.995 respectively). In contrast, the two mitochondrial genes were less congruent with each other, but not significantly different (p-value = 0.643). With no significant incongruence between the genes trees, the 19 specimens with overlapping sequence data were concatenated to create total evidence Parsimony and ML trees (Figure 3.2E). The parsimony tree revealed the same 9 distinct clades (depicted using a unique letter or reference name) that were again separated by deep branch lengths and supported by high net divergence between clades (from 7.56 to 18.44%; Table 3.3). Most ML bootstrapping values were greater than 84%, apart from the divergence of clade d from e and clade f from g (67% and 59%, respectively). Both the nuclear and mitochondrial genomes supported the nine clades.

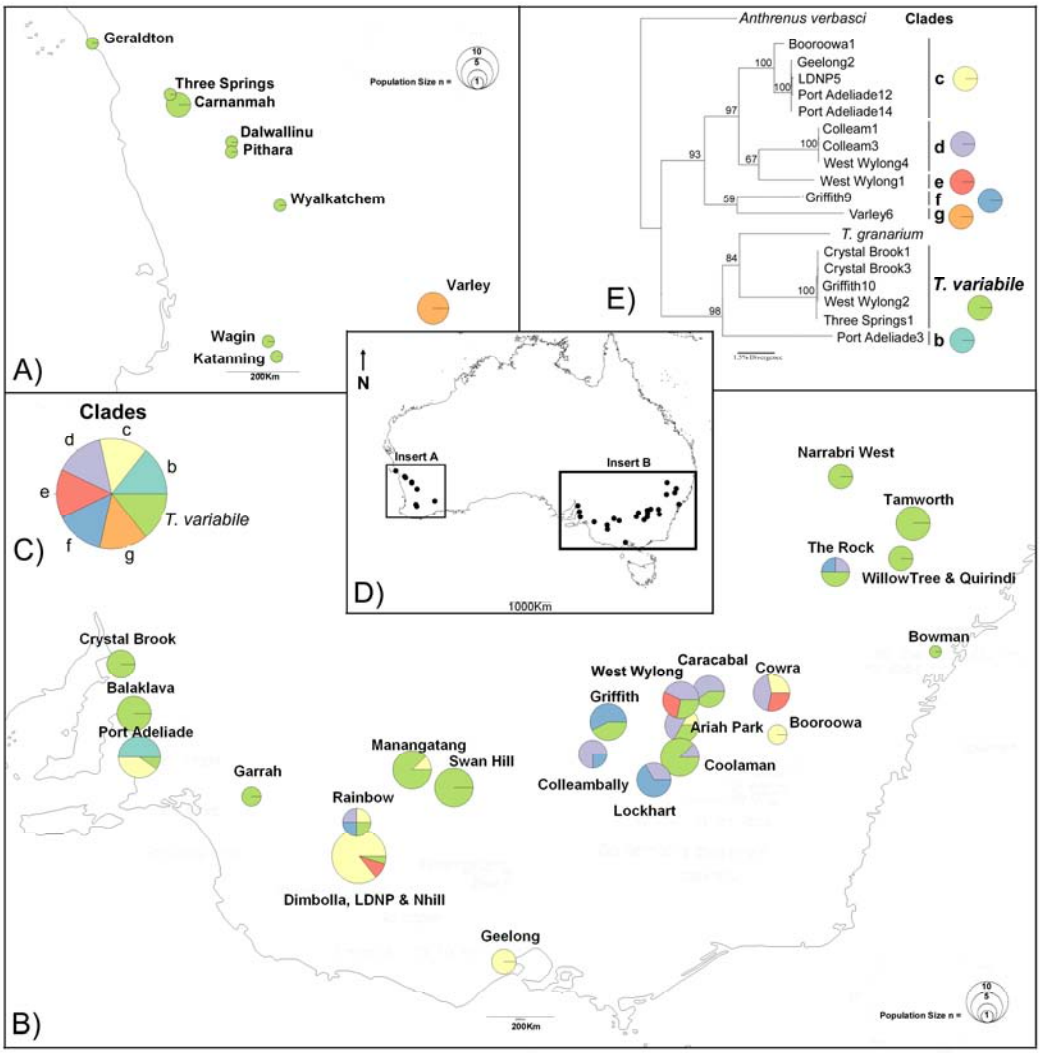

Figure 3.2. Distribution of each clade determined by Parsimony and ML phylogenetic analyses. A) Spatial distribution of each clade within Western Australia. B) Spatial distribution of each clade on the east coast of Australia. C) Clade wheel, each colour representing a clade determined by the phylogenetic trees. D) Sampling areas within Australia. E) Consensus Parsimony tree combining the partial sequences of the two mitochondrial gene regions COI and CYT b and the partial nuclear 18S. The numbers above each of the branches are ML bootstrap values and each letter depicts a clade. The outgroup was *A. verbasci*, an exotic pest established in Australia. The external branches of the three gene parsimony tree are labelled with the location of the specimen and the sample number for that site. These labels can be cross-referenced with the sample collection data found in Table 3.1.

Table 3.3. Net divergence between the eight clades calculated on the three gene fragments (*18S*, *COI* & *CYT b*), the values are nucleotide percentage differences (%). *A. verbasci* (A.v), *T. granarium* (T.g), and *T. variabile* (T.v).

	A.v	b	c	d	e	f	g	T.g
A.v								
b	18.44							
c	15.97	15.17						
d	15.34	15.18	8.35					
e	15.55	14.71	7.56	7.91				
f	15.62	15.76	10.12	10.44	10.06			
g	15.83	15.34	10.27	10.06	10.49	9.08		
T.g	16.89	14.85	14.17	14.73	13.93	14.5	13.72	
T.v	17.28	14.47	14.24	15.51	14.07	13.24	13.99	12.93

The distribution of each clade was visualised on a distribution map (Figure 3.2). The majority of specimens that were placed in clades d and f (95% and 91%, respectively) were mainly observed in the southern part of New South Wales (Table 3.4) and 73% of clade c was found in Victoria. Clade b was only observed at Port Adelaide in South Australia, and clade g was only located at Varley in Western Australia. The specimens that clustered with the *T. variabile* reference specimens were at 78% of the sites sampled (Table 3.1).

Table 3.4. The distribution of each clade per state.

	A. verbasci	*T. variabile*	B	C	D	E	F	G
New South Wales	0	44	0	15	95	67	91	0
South Australia	0	18	100	12	0	0	0	0
Victoria	0	24	0	73	5	33	9	0
Western Australia	100	14	0	0	0	0	0	100
Total Frequency	0.7	46.7	3.3	21.7	13.2	3.9	7.2	3.3

3.3.5 Genetic Structure of *T. variabile* using *COI*

Cytochrome oxidase I was the only gene able to resolve divergence within the *T. variabile* clade (Table 3.2), revealing 9 unique haplotypes (Table 3.5). Network analysis of these haplotypes, revealed one main haplotype (H4) with four direct descendents (H2, H3, H6 & H8). One haplotype (H5) was distantly related to H4 (missing multiple mutational steps) and three haplotypes (H1, H7, and H9) were not connected in the network because their divergence ranged outside the 95% parsimony limit imposed by the program (Figure 3.3C). With no connectivity between H1, H7, and H9 and the rest of the haplotypes, the relatedness between these haplotypes could not be determined. Haplotype 4 was found at all the eastern and western states collection sites where the presence of *T. variabile* was identified (using the molecular data; Table 3.1), except for Willow Tree, that had only H5. Tamworth and Swan Hill both contained additional haplotypes that showed no connectivity the main population.

Table 3.5. Number of *T. variabile* haplotypes per collection site for the mitochondrial markers *CO1*.

Site	State	n	H1	H2	H3	H4	H5	H6	H7	H8	H9
Ariah Park	Vic	2				2					
Balaklava	SA	1				1					
Carnanmah	WA	3				3					
Coolaman	Vic	2				2					
Crystal Brook	SA	3		1		1		1			
Dalwallinu	WA	1				1					
Dimboola	Vic	1				1					
Garrah	SA	2				2					
Geraldton	WA	1				1					
Griffith	NSW	1				1					
Katanning	WA	1				1					
Manangatang	Vic	1				1					
Narrabri West	NSW	3				3					
Pithara	WA	1				1					
Port Adelaide	SA	1				1					
Quirindi	NSW	1				1					
Swan Hill	Vic	4	1		1	1				1	
Tamworth	NSW	5				3			1		1
Three Springs	WA	1				1					
Wagin	WA	1				1					
West Wyalong	NSW	1				1					
Willow Tree	NSW	1					1				
Wyalkatchem	WA	1				1					
Total		39	1	1	1	31	1	1	1	1	1

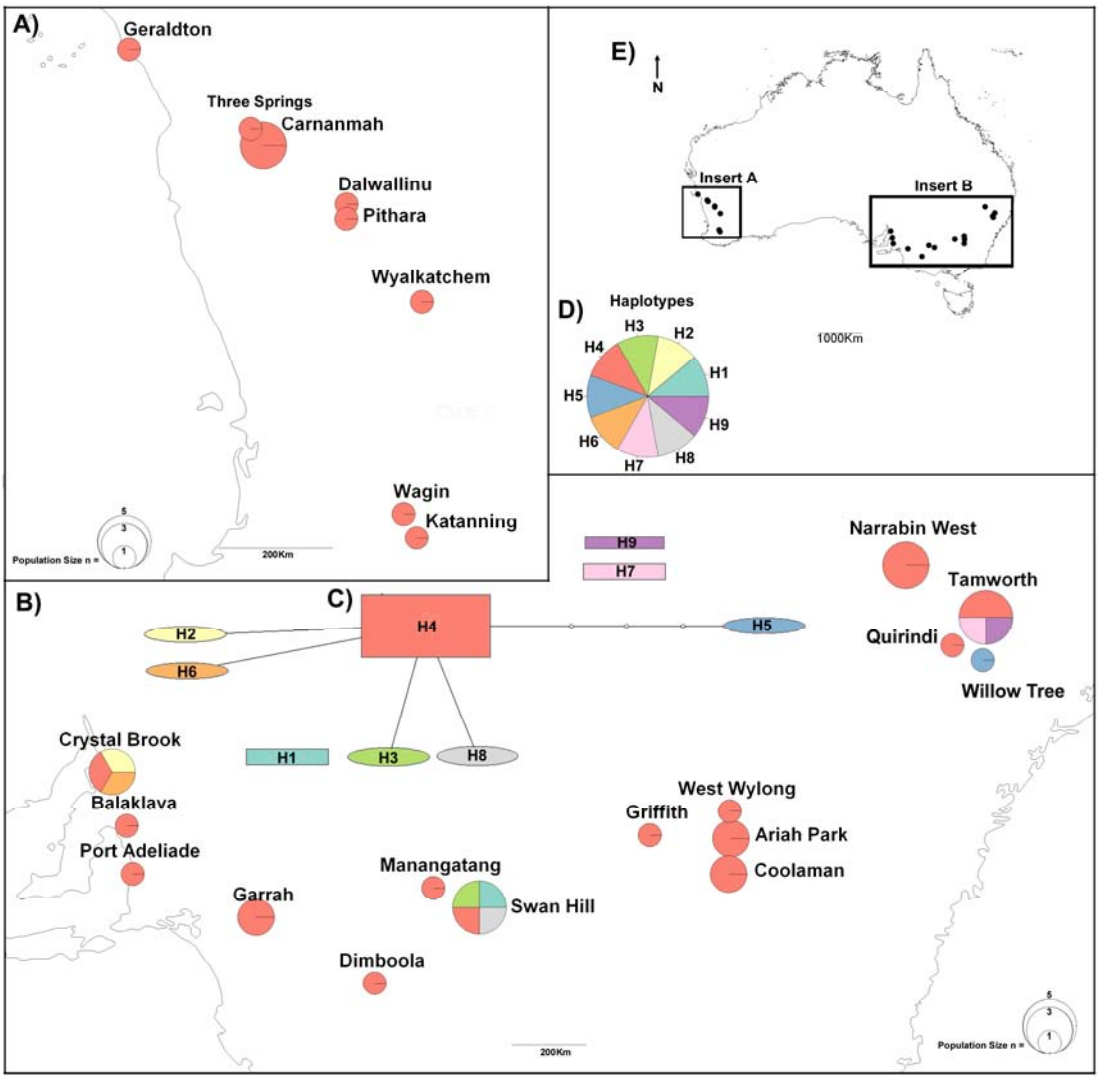

Figure 3.3. Spatial distribution of the nine *T. variabile* haplotypes. A) Spatial distribution of *T. variabile* within Western Australia. B) Spatial distribution of T. variabile within eastern Australia. C) TCS network of the nine haplotypes. D) Haplotype colour key. E) Sampling areas within Australia where the presence of *T. variabile* was confirmed using molecular data.

3.4 Discussion

The first aim of this study was to use molecular techniques to re-examine the *T. variabile* collected in the surveillance traps and compare the results to original morphological identification data. The second aim, if the morphological and molecular results were congruent, was to investigate the population structure and spread of *T. variabile*. The results showed seven deeply distinct clades in both the all gene trees and the *COI* trees, with an average pairwise divergence between clades of 18.5% (S.E. 2.27) and 17.5% (S.E. 1.87%) for *COI* and *CYT b*, respectively. These results, consistent for both nuclear and mitochondrial genes, strongly suggest that seven distinct species were collected, of which only 47% were clustered with the *T. variabile* reference sample (Table 3.4). Population analysis of those specimens identified as *T. variabile* revealed one main haplotype with several derived and three unconnected haplotypes that may be separate incursions. If the samples collected during the *T. variabile* eradication campaign (1977 to 1982) were also misidentified and the true *T. variabile* population the result of multiple incursions could have influenced the decision to stop the campaign.

Of the 153 individuals that were genotyped, only 46.7% could be grouped with the specimens that were morphologically verified as *T. variabile*. Since the specimens were macerated to liberate the DNA, cross-referencing the genetic and morphological patterns was not possible. However, the molecular data strongly suggests that the divergent clades were not *T. variabile* as large pairwise divergence values were observed between clades, implying deep evolutionary divergence and presumptive reproductive isolation. The average inter and intra-specific divergence for the *COI* locus in this study was 18.5% and 0.48%, respectively. In comparison, the inter and intra specific divergence levels for the *COI* locus, previously determined for a multitude of insect pests (Cognato, 2006), are an average of 7.4% (2 – 24%) and 1.75% (0.077 – 26%) respectively. The mean inter-clade divergence observed in this study was more than double this previous report, and the intra-clade divergence was only a fraction of that calculated by Cognato (2006), which further supports the presence of a number of cryptic species in the present study. Furthermore, *T. variabile* and *T. granarium*, which can be differentiated using morphological keys, had a *COI* divergence of 18.8% (Table 3.2), the lowest observed

between *T. variabile* and any other clade. With further research, morphological characters may be resolved that discriminate between the other cryptic species discovered in this study, making field recognition and management more feasible.

The six putative species accounting for 53% (n=82) of the individuals analysed could be attributed to both high trapping yields and the difficulty encountered when identifying samples collected from a trapping program. The traps were placed outside of the grain facilities, possibility to increase the number of *T. variabile* collected (Campbell and Mullen, 2004), but this also potentially increases the number of native dermestids collected, since some natives have been shown to be attracted to the pheromone used in these traps. The number of specimens collected in each trap was likely to have exceeded a few hundred making identification particularly difficult, especially when adopting a quick visual screening approach. Morphological distinction between *T. variabile* and other closely related *Trogoderma* requires dissection and careful microscopic examination of the genitalia (Szito, 2007a), not just the examination of external characteristics. Dissection and preparation of the genitalia for examination requires time and skill, and would be impractical to perform on the number of samples collected at each site. If the eradication campaign adopted a similar approach and only examined the external characters, the likelihood of misidentification would be high and potentially result in efforts directed at non-target species.

The genetic data set generated in this study allowed exploration into possible genetic structure within *T. variabile*. Although sample sizes were low, several interesting trends were observed based upon the distribution and relatedness of haplotypes. The main *T. variabile* haplotype (H4) was found at all eastern and western Australian sites where *T. variabile* was present, except for Willow Tree in NSW. The distance between each of these sites ranged from 9 to 3760 km, and considering that *T. variabile* has a limited dispersal range based on its ecology (Campbell and Mullen, 2004), human-aided transport is the plausible dispersal means. The three non-connected haplotypes observed at Tamworth (NSW) and Swan Hill (Vic) are interesting because they are localised and yet to disperse, suggesting that they are younger – such is the expectation based on coalescent theory (Kingman, 1982, Crandall and Templeton, 1993). Alternatively, the non-connected haplotype observed

at Tamworth could represent another closely related species. This haplotype was 2.9% divergent from the main haplotypes, which was the above the average intra-specific divergence levels observed by Foottit *et al.* (2008) and deWaard *et al.* (2009). However, without morphological support it is difficult to determine if this is another species or intra-specific variation. Interestingly, while *CYT b* was as divergent as COI, no *CYT b* variation was observed in the *T. variabile* clade. A similar observation was made by Tsutsui *et al.* (2001) and again demonstrates the importance of using multiple loci. Future studies will need to increase the sample size both locally and internationally to ascertain if the multiple haplotypes are the result of multiple incursions or a single incursion that contains multiple haplotypes.

The conclusion from this study's findings is that errors were made in the original morphological identification, with 53% of the specimens collected during the 2001 to 2003 *T. variabile* trapping program wrongly identified. On this basis, the rate and extent of spread of *T. variabile* is much less than originally estimated. Furthermore, the intra-specific divergence within the *T. variabile* clade suggests multiple incursions of the pest. If this true, then this finding has significant implications for the eradication campaign, because it shows leakage at the quarantine barrier that could lead to the re-introduction of *T. variabile* even if eradication was initially successful. Future investigations should aim at re-sampling the collection sites that were used in this study in order to obtain intact specimens of the putative species identified and characterise them morphologically to support these results.

Chapter 4

A rapid non-destructive DNA extraction method for insects and other arthropods

Preparation of arthropods for morphological identification often damages or destroys DNA within the specimen. Conversely, DNA extraction methods often destroy the external physical characteristics essential for morphological identification. We have developed a rapid, simple and non-destructive DNA extraction technique for arthropod specimens. This technique was tested on four arthropod orders, using specimens that were fresh, preserved by air drying, stored in ethanol, or collected with sticky or propylene glycol traps. The technique could be completed in 20 min for Coleoptera, Diptera and Hemiptera, and two minutes for the subclass Acarina, without significant distortion, discolouration, or other damage to the specimens. This chapter has been peer reviewed and published in the Journal of Asia-Pacific Entomology and is attached in appendix III.

4.1 Introduction

The identification of arthropod pests typically uses morphological characteristics and is sometimes problematic, potentially failing to recognise a serious pest or to cause alarm over a non-threatening species (Hebert *et al.*, 2003). Difficulties with identification may result from natural phenotypic variation within a species, the involvement of morphologically cryptic taxa, limitations of morphological keys, the need to identify partial specimens that lack important taxonomic characteristics, and the shortage of experienced taxonomists required for identification (Hebert *et al.*, 2003). A combination of morphological identification and DNA barcoding is being developed to overcome such problems (Moritz and Cicero, 2004, Floyd *et al.*, 2009).

Methods used to prepare arthropod samples for morphological identification often prevent the analysis of DNA from the specimen, while the column-based DNA extraction methods most frequently used on arthropods require maceration of the sample, destroying the morphological characteristics required for identification. For this reason, DNA extraction that may cause damage or loss of specimens cannot be performed on Type specimens held in insect collections. Several non-destructive DNA extraction methods have been published, which allow the specimen to be identified using DNA analysis without any obvious alterations to the morphological characteristics (Gilbert *et al.*, 2007, Favret, 2005, Hunter *et al.*, 2008, Rowley *et al.*,

2007, Pons, 2006). However, these methods require the use of toxic or corrosive chemicals (e.g. phenol, chloroform, and guanidine isothiocyanate), are time-consuming through the need for overnight incubation, or risk the loss of DNA through ethanol or isopropanol precipitation. In addition, some methods are expensive and/or inefficient (Gilbert *et al.*, 2007, Favret, 2005, Hunter *et al.*, 2008, Rowley *et al.*, 2007, Pons, 2006). Because of the varying suitability of the different methods, a laboratory dealing with multiple arthropod taxa may need to establish and validate several methods.

To validate a single, broadly applicable method, we have used EDNA HiSpEx (Fisher Biotec, Perth), a high speed DNA extraction method. The method is based on alkaline hydrolysis of proteins and uses ingredients similar to those already routinely used by morphological taxonomists in the preparation of insect and mite samples for dissection and/or slide mounting (Szito, 2007a). It is a simpler and quicker method than most other non-destructive DNA extraction techniques. Extraction of DNA with EDNA was tested on specimens from four arthropod orders, which had been preserved under various conditions. We examined important morphological characteristics before and after DNA extraction. The method was further tested on museum specimens, which had been stored for periods greater than 30 years. In addition, larvae moults are frequently collected by quarantine and biosecurity staff. Subsequently, the efficacy of extracting DNA from larvae moults by immersing them in either water, EDNA or DNeasy Animal Tissue Kit (Qiagen) was examined.

4.2 Materials and Methods

4.2.1 Samples and DNA extraction

Samples were collected and stored in either ethanol, 20% propylene glycol then ethanol, air dried then pinned, or prepared fresh (Table 4.1). Whole specimens from the insect Orders Coleoptera (Dermestidae, Buprestidae and Cerambycidae families), Hemiptera (Aphididae), Diptera (Tephritidae), and the arachnid Order Prostigmata (Eryophyidae) were placed into 0.2-ml microcentrifuge tubes. A master mix of EDNA was made by mixing Solution 1A and Solution 1B together at a ratio of 39:11. For specimens <4mm in length 50 µl of EDNA mix was added to ensure the entire specimen was submerged. For larger specimens the volume of EDNA mix was

increased to 100 µl. Larval specimens were pierced with a micro pin, behind the first abdominal segment, and the pin was immersed in 25 µl of EDNA mix. Eriophyid mites were immersed in 10 µl of EDNA mix. Eriophyid mites were incubated at 99 °C for 2 min and all other specimens were heated at 99ºC for 20 min. After incubation, 0.25 volume of EDNA Solution 2 was added to each tube. Solution 2 and Solutions 1A and 1B were mixed by gentle pipetting and an aliquot was transferred to a sterile microcentrifuge tube, leaving the specimen in the original tube for further preparative steps or to be restored to their original conditions of storage. DNA concentration was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc), and extracts were diluted to 20 ng/µl using PCR-grade water.

Larvae moults from *T. variabile* and *A. verbasci* were used to examine the efficacy of water, EDNA and Qiagen's DNeasy Animal Tissue Kit at extracting DNA (Table 4.2). Larvae moults were placed in 0.2-ml microcentrifuge tubes and immersed in 25 µl of EDNA buffer or water. These specimens were heated at 99 ºC for 20 min. After incubation, 0.25 volume of EDNA Solution 2 was added to each EDNA reaction and mixed by gentle pipetting. The samples that were tested with the DNeasy Animal Tissue Kit (Qiagen) followed the standard DNA extraction method for muscle extraction, excluding the maceration step (Qiagen, 2006).

Table 4.1. Samples used for DNA extraction and amplification, successfully amplified primer sets (primers described in Table 4.2.), and maximum amplified PCR product determined by agarose gel.

No.	Method	Order	*Genus*	*species*	Type	Year	Stored	PrimerSet	Max size (bp)
1	EDNA	Coleoptera	*Orphinus*	*unknown*	Adult	2008	20% propylene glycol then ethanol	1	550
2	EDNA	Coleoptera	*Anthrenocerus*	*australis*	Adult	2008	20% propylene glycol then ethanol	1,2,3	550
3	EDNA	Coleoptera	*Anthrenus*	*verbasci*	Adult	2008	Air dried	1,3	550
4	EDNA	Coleoptera	*Trogoderma*	*variabile*	Adult	1995	Air dried	1,2,3	550
5	EDNA	Diptera	*Bactrocera*	*jarvisi*	Adult	2007	Air dried and mounted	1,2,3	800
6	EDNA	Coleoptera	*Attagenus*	*unknown*	Adult	2006	Air dried and mounted	1,2,3	550
7	EDNA	Diptera	*Bactrocera*	*jarvisi*	Adult	1997	Air dried and mounted	1,2,3	550
8	EDNA	Coleoptera	*Orphinus*	*unknown*	Adult	1994	Air dried and mounted	2	500
9	EDNA	Coleoptera	*Neoanthrenocerus*	*unknown*	Adult	1978	Air dried and mounted	1	550
10	EDNA	Coleoptera	*Anthrenus*	*verbasci*	Adult	1950	Air dried and mounted	-	-
11	EDNA	Coleoptera	*Hylotrupes*	*bajulus*	Larvae	2009	Ethanol	1	550
12	EDNA	Coleoptera	*Hylotrupes*	*bajulus*	Adult	2008	Ethanol	1,4	850
13	EDNA	Hemiptera	*Uroleucon*	*sonchi*	Adult	2008	Ethanol	1,2,7	550
14	EDNA	Coleoptera	*Anthrenus*	*verbasci*	Adult	2007	Ethanol	1,2	550
15	Qiagen	Coleoptera	*Trogoderma*	*granarium*	Adult	2007	Ethanol	1,2,3,5	2000
16	EDNA	Coleoptera	*Trogoderma*	*granarium*	Adult	1993	Ethanol	1	550
17	EDNA	Coleoptera	*Trogoderma*	*granarium*	Larvae	1993	Ethanol	1,2	550
18	EDNA	Coleoptera	*Hylotrupes*	*bajulus*	Larvae	2009	Ethanol pin	1	550
19	EDNA	Coleoptera	*Buprestis*	*novemaculata*	Larvae	2009	Ethanol pin	1	550
20	Qiagen	Coleoptera	*Trogoderma*	*variabile*	Adult	2009	Fresh	1,2,3,6	4000
21	EDNA	Prostigmata	*Aceria*	*tosichella*	Adult	2008	Fresh	2,7	450
22	EDNA	Coleoptera	*Cryptolestes*	*unknown*	Adult	2008	Fresh	1,2,3	550
23	EDNA	Coleoptera	*Sitophilus*	*oryzae*	Adult	2008	Fresh	1,2,3	550
24	EDNA	Coleoptera	*Tribolium*	*castaneum*	Adult	2008	Fresh	1,3	550
25	EDNA	Coleoptera	*Trogoderma*	*variabile*	Adult	2008	Sticky Trap	1,2,3,4	800

Table 4.2. Larvae moult samples used for DNA extraction and amplification.

Genus	Species	Method	Type	n	Amplification Success (%)
Trogoderma	*variabile*	EDNA	Moult	8	88
Anthrenus	*verbasci*	EDNA	Moult	2	0
Trogoderma	*variabile*	Qiagen	Moult	2	0
Anthrenus	*verbasci*	Qiagen	Moult	2	0
Trogoderma	*variabile*	Water	Moult	8	63
Anthrenus	*verbasci*	Water	Moult	2	100
Trogoderma	*variabile*	EDNA	Adult	8	100
Anthrenus	*verbasci*	EDNA	Adult	2	100
Trogoderma	*variabile*	Qiagen	Adult	2	100
Anthrenus	*verbasci*	Qiagen	Adult	2	100
Trogoderma	*variabile*	Water	Adult	8	13
Anthrenus	*verbasci*	Water	Adult	2	0

4.2.2 PCR Amplification

Primer pairs used for DNA amplification targeted single copy nuclear DNA, multiple copy nuclear DNA, or mitochondrial DNA (Table 4.3). Reaction volume was 10 µl, including 1 µl of EDNA extracted DNA as template, 2.5 pmol of each primer, 200 µM of each dNTP (Invitrogen), 0.25 U of Taq Platinum (Invitrogen) and Taq polymerase reaction buffer (Invitrogen). Annealing temperatures and extension times varied as shown in Table 4.3. The thermocycler conditions were: 95 ºC for 5 min; 40 cycles of: 95 ºC denaturation for 30 s, annealing for 30 s, and 72 ºC for extension; with a single final extension period of 72 ºC for 5 min. PCR products were analysed by electrophoresis (Chapter 2.3). Samples refractory to amplification were analysed by PCR using 0.2 U of Phusion High-Fidelity DNA polymerase with Phusion HF buffer (Finnzymes) in a volume of 10 µl with 1µl of template, 4 pmol of each primer, 200 µM of each dNTP (Invitrogen).

4.2.3 Sample Preparation

After DNA extraction, specimens belonging to the families Aphididae and Eriophyidae required additional preparation steps for mounting on glass microscope slides prior to microscopic examination. Clearing and slide mounting of aphid species has been described previously (Favret, 2005)

Table 4.3. Amplification primers used in this study.

Primer No.	Target Gene	Primer Name	Primer Sequence 5`-3`	Aneal Tm	Ext Time (s)	Exp Size (bp)	Type	Reference
1	Multiple Nuclear	*18S*F1 *18S*R1	TGCTTGTCTCAAAGATTAAG GTATCGTTATTTTTCGTCAC	43.7	30	550	Universal	This Study
2	Mitochondrial	C1-J-1751 C1-N-2183	GGATCACCTGATATAGCATTYCC CAACATTTATTTTGATTTTTTGG	43.7	30	500	Universal	(Simon *et al.*, 1994)
3	Single Nuclear	Wg578F Wg1032R	ACYTCGCAGCACCARTGGAA TGCACNGTGAARACYTGCTGGATGCG	43.7	30	450	Specific	(Moreau *et al.*, 2006)
4	Multiple Nuclear	ITS1 ITS2	TCCGTAGGTGAACCTGCGG GCTGCGTTCTTCATCGATGC	50	30	800	Universal	(White *et al.*, 1990)
5	Mitochondrial	N4-F-8772 CB-R-10904	CAGGAGCCTCTACATGAGCTTT TTTGATCCTGTRTGGTGWAGAA	50	300	4000	Specific	*[1]
6	Mitochondrial	C3-F-5393 N4-R-9004	TTCCCAAATTCACCACTTCG TTGGTTCCCTTATTATTTTTGATTG	50	120	2000	Specific	*[1]
7	Mitochondrial	CB-J-10933 CB-N-11367	TATGTACTACCATGAGGACAAATATC ATTACACCTCCTAATTTATTAGGAAT	43.7	30	500	Universal	(Simon *et al.*, 1994)

*** Note 1 - These primers were created from known *T. variabile* sequences.**

4.3 Results

Amplification products from DNA extracted by the EDNA procedure from Coleoptera, Hemiptera, and Acarina, are shown in Figure 4.1, together with results for specimens from which DNA was extracted by the Qiagen column-based method (DNeasy). The successful amplification of products from 400 to 550bp demonstrated that the EDNA method is successful with specimens that have been stored for more than 20 years. Amplification was not achieved for the oldest sample, which was collected in 1950. Individual primer-pairs did not yield amplification products from all specimens tested. However, specimens from all four orders showed amplification with at least one primer pair (Figure 4.1).

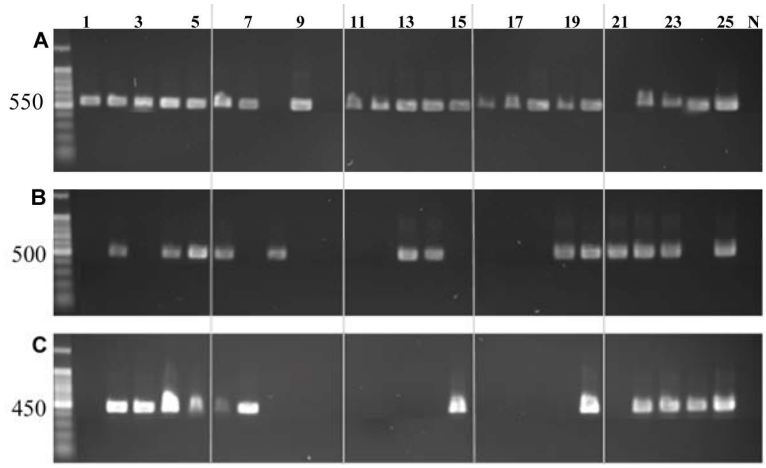

Figure 4.1. PCR amplification products from (A) 18S rRNA gene primers (B) mitochondrial DNA primers (C) Wingless gene primers. Samples 1 to 25 correspond to the specimens shown in Table 4.1. Lane 1: 100 bp ladder. Samples 15 and 20: DNA extracted using a Qiagen DNeasy Animal Tissue Kit.

Figure 4.2 shows aphids, mites and beetles that were cleared and slide mounted without EDNA extraction (A1, B1, and C1) and with EDNA extraction (A2, B2, and C2). Samples treated with EDNA showed no discernable morphological changes. For example, the prodorsal shield pattern (A2.1) and the featherclaw (A2.2) of *Aceria tosichella (*Acarina: Eriophyidae) and other Eriophyid mites are important diagnostic features for classification (Whitmoyer *et al.*, 1972, Halliday and Knihinicki, 2004) and Figure 4.2B2.2 shows that EDNA extraction did not alter the delicate structures. Although EDNA extraction achieved some degree of clearing, this

alone was insufficient for microscopic analysis when the slide was mounted in the non-clearing medium Euparal. However, Hoyer's medium produced satisfactory clearing for specimens with and without EDNA treatments.

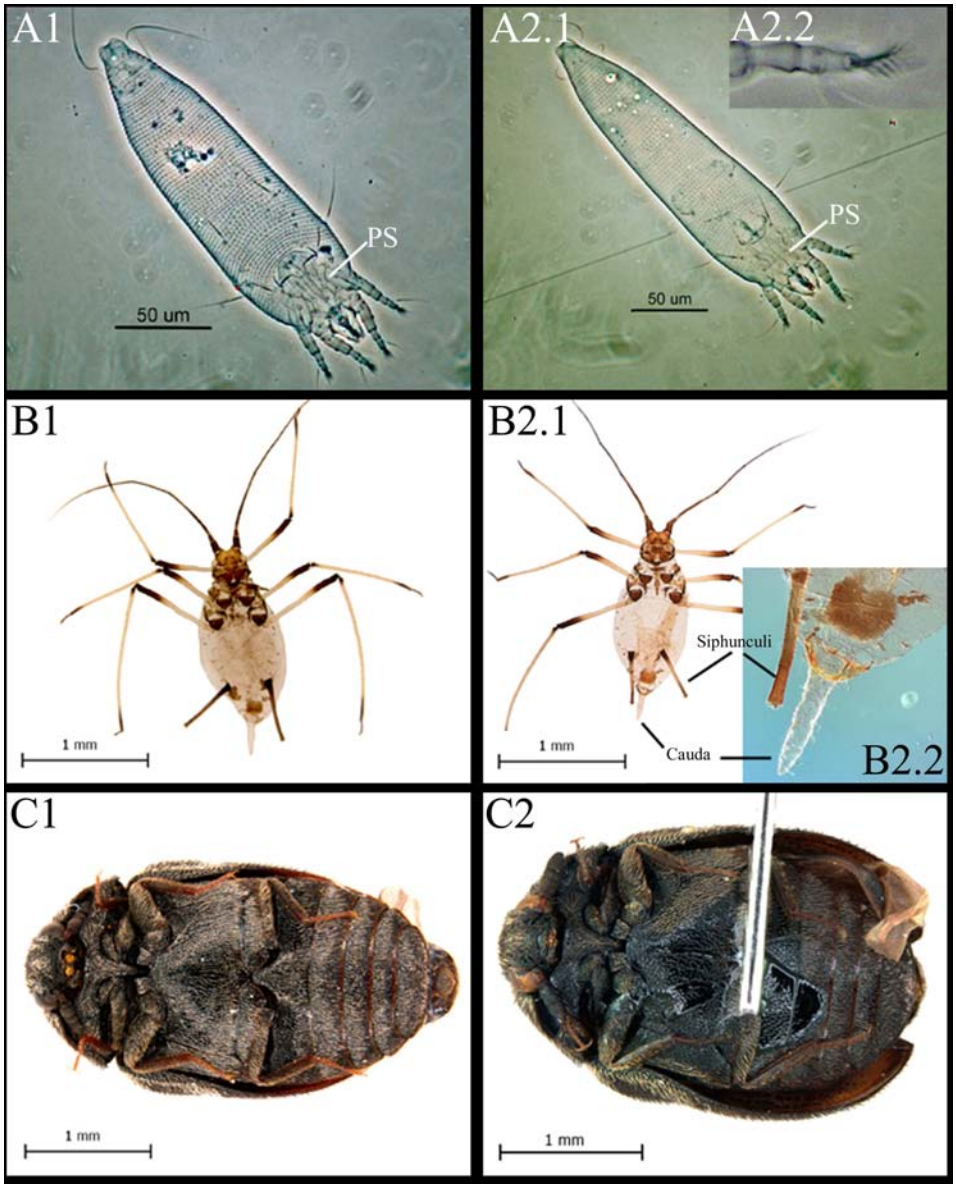

Figure 4.2. Effect of EDNA extraction on external morphological characteristics. Panels on the left show the samples that were slide mounted without EDNA treatment (A1 and B1) or prior to EDNA extraction (C1), on the right, those extracted with EDNA. (A) Eriophyid mite (sample 21), (B) Aphididae (sample 13), (C) Coleoptera (sample 2). A2.1 and A2.2 illustrate the prodorsal shield (PS) pattern and featherclaw that are used to characterise *A. tosichella*. A2.2 shows the integrity of the delicate featherclaw after DNA extraction.

DNA extraction using EDNA showed no detectable effect on the morphological characteristics of *Uroleuchon sonchi* (Hemiptera: Aphididae), leaving the fine body setae intact. The weak clearing action on the specimens during DNA extraction allowed the subsequent KOH clearing process to be performed in a shorter time than normal and abrogated the need for piercing of the lateral abdominal wall to obtain adequate clearing. In *Bactrocera* species (Diptera: Tephritidae), the wing integrity was lost due to shrivelling. No damage to any other morphological features was recorded in this genus. For all coleopteran specimens the EDNA method produced no detectable loss of scales or setae, or damage to the male genitalia. Conveniently, EDNA provided sufficient clearing of fatty tissue in the genital region to make morphological examination easier. In Coleoptera, the EDNA treatment was observed to soften joints, allowing easier manipulation of the body parts in the identification process and subsequent mounting.

Analysis of Qiagen's DNeasy Animal Tissue kit on larvae moults showed that this method failed to extract amplifiable DNA (Table 4.2). EDNA was highly successful (88%) at extracting DNA from *T. variable* moults, yet failed to extract DNA from the *A. verbasci* moults. While water failed to extract amplifiable DNA from intact adult specimens, it was equally successful as EDNA at extracting DNA from larvae moults.

4.4 Discussion

Many previously described methods for DNA extraction from arthropods require the use of multiple steps, toxic or corrosive chemicals, or expensive components (Gilbert *et al.*, 2007, Favret, 2005, Hunter *et al.*, 2008, Rowley *et al.*, 2007, Pons, 2006). In contrast, EDNA extractions are inexpensive, time and labour efficient, and of low toxicity, allowing the extraction of DNA for amplification without damage to the external characteristics. The wing shrivelling observed in *Bactrocera* spp. can be corrected with critical point drying (Brown, 1993). Three genomic regions were amplified to evaluate the extraction process, because they represent different copy-numbers per genome. This allowed the effects of differing abundance of target DNA in the genome to be assessed. Perhaps more importantly, amplifying three different targets allowed us to show that a lack of amplification in a particular reaction was

due to reaction-specific factors, such as primer mismatch, and was not caused by general inhibition of PCR by residual extraction materials or DNA degradation. DNA extracted from a 50 year old, air dried sample was refractory to amplification. This may have resulted from partial depurination of the DNA, which often occurs in older specimens, leading to extensive DNA fragmentation in the alkaline conditions in EDNA (Sherman and Loeb, 1977). We concluded that the Finnzymes product, PhusionTaq, was the most suitable enzyme for this PCR application because it successfully amplified DNA from samples that could not be amplified using Roche Fast Start Taq polymerase. Phusion Taq has been shown previously to be less sensitive to residual inhibitors (Yang and André, 2007). However, PCR analysis of DNA from the 50 year old sample failed with both polymerases, supporting the conclusion that the sample lacked DNA of suitable quality.

The non-destructive DNA extraction and mounting method developed by Favret (2005) is time-consuming and requires the use of a minuten pin to liberate DNA from the specimen, which could potentially lead to sample cross contamination (Rowley *et al.*, 2007). The results showed that incision or perforation of the insect is not required to extract DNA using EDNA. Rather, as indicated by a previous study (Gilbert *et al.*, 2007) the extraction buffer liberates DNA from the interior through openings such as the mouth, anus and spiracles. Removing the puncturing step further reduces handling that may damage the exoskeleton or cuticle, which are required for morphological characterisation. Following EDNA extraction of DNA, slide mounting of some of the specimens requires the normal clearing step in KOH, but the process can be completed in less time.

The minute, delicate, and soft body of the Eriophyid mites has previously prevented a researcher's ability to characterise an individual specimen using both morphological and DNA analysis. Previously, in order to extract DNA, individual specimens were destroyed (Carew *et al.*, 2009). In addition, specimens must be cleared then slide mounted before being identified using light microscopy and are thus unable to be used for DNA extraction. Consequently, assumptions must be made about the species under examination before DNA extraction occurs. Full analysis of individual Eriophyid specimens is important because different species can co-exist on the same plant host. Mite specimens required only 2 min at 99°C for release of

DNA and showed no visible damage to the cuticle or body structure. In contrast, the previously described non-destructive method to extract DNA from terrestrial arthropods (Rowley *et al.*, 2007) required more than 4 hours. Previous studies on the preservation of mites showed that cryoconservation in water was an efficient medium for short term storage (Rey *et al.*, 2004). This study confirmed that amplification was successful when mites were preserved in water. However, if the samples were frozen prior to DNA extraction, morphological examination revealed that the specimens became fragmented during the heating process. This was possibly due to the formation of ice crystals during the freezing process which, when rapidly heated, resulted in rupture of the cuticle. As an alternative, it was found that short term storage of the mites prior to DNA extraction and mounting, was best achieved by live preservation in zip lock bags filled with host plant material (up to 1 month), or longer term storage in 70% - 100% ethanol.

DNA extraction from a single fresh adult coleopteran specimen yielded DNA sufficient for at least 100 PCR analyses, without visual discolouration or loss of setae. Previously reported problems such as flaccid claws, buckled disks or the apical tarsal segments becoming droopy after extraction with guanidine isothiocyanate (Rowley *et al.*, 2007) were not observed with this method. The EDNA method was found to soften joints, which allowed easier positioning and manipulation of specimens' legs, antennae and elytra prior to resetting.

The extracted DNA was useful only for amplification of targets up to approximately 800 bp. Column-based extracts (Qiagen), which require maceration of the sample, allowed amplification of targets >2,000 bp in preserved specimens and up to 4,000 bp in fresh specimens (Table 4.1). While Qiagen DNeasy is quite often used as a non-destructive technique, a search of the literature revealed no published evaluation of this method. Furthermore, the use of Qiagen often requires several hours of incubation and is a relatively expensive extraction method, therefore may not be suitable for diagnostic facilities.

The preservation method used did not appear to influence the results, with most coleopteran specimens producing DNA that could be amplified. However, specimens stored in ethanol for more than a decade gave variable results. The preparation of

coleopteran specimens for dissection, slide mounting or card mounting, often involves cleaning with Decon 90, limonene, citricide, Dissolv-it or ultrasound (Szito, 2007a). Electrophoresis of the extracted DNA revealed that these methods, plus storage in ethanol, caused degradation of the DNA. To be usable for analysis, DNA extraction by EDNA or any other process should be performed prior to such preparative steps.

Extraction of DNA from Dipteran specimens, even by soaking in water, caused shrivelling and changes to the shape of the wings (Rowley *et al.*, 2007) and EDNA caused similar changes. While morphological wing characteristics are extremely important in their taxonomy, it may not be feasible to stop the changes to the shape of the wing, although critical point drying can restore wing shape (Brown, 1993). Alternatively it may be possible to extract DNA from the bodies of Diptera using EDNA, without immersion of wings. Alternatively, DNA could be obtained from a single leg if the loss of integrity did not pose taxonomic problems. For unique or taxonomically important specimens, the body part removed for DNA extraction could be retained intact with the original specimen. This would be particularly important for a unique specimen or type specimen that had to be used for DNA barcoding. For any taxonomic group it is essential to be able to make taxonomic comparisons using both established and new morphological taxonomic criteria.

When used directly on larval specimens the EDNA process was highly destructive, causing various degrees of structural collapse. The alkaline components in EDNA solubilise lipids, break down adipose tissue and significantly disfigure the larvae. However, we found that by inserting a pin through the ventral abdominal wall and performing EDNA extraction on the material adhering to the pin, sufficient DNA material could be obtained for amplification, without significantly damaging the larval physical characteristics. Thus, we are of the opinion that this method could be applied to other soft cuticle specimens such as Lepidoptera.

Using water to extract DNA from larvae moults was equally as successful as EDNA, despite the adult specimens failing to amplify. The failure to obtain amplifiable DNA from adult specimens could be attributed to intact cell membranes, which would require lysis to liberate the DNA. The failure of the DNeasy Animal Tissue Kit

(Qiagen) to extract DNA was ascribed to the mechanisms involved with DNA binding to the silica membrane. While the efficacy of binding DNA to a silica membrane is high, problems occur when DNA concentrations are <5ng (Qiagen, 2006). Despite the small sample size, water may be a suitable substance that can be used to extract DNA from larvae moults.

By incorporating the EDNA DNA extraction method into the diagnostic protocols used here, the specimens were able to be cleaned, joints softened, fats and soft tissue cleared from within the internal cavities (often without the need for piercing), and softening of the abdomen for easy removal of the genitalia. This offers a valuable tool for the traditional morphological taxonomist and may bridge the gap between morphological taxonomy and molecular taxonomy without the destruction of important type specimens. The increased ease and rapidity of species identification will facilitate applications such as biomonitoring for crop pests, identification of invasive species, and monitoring the effects of climate change on arthropod populations. Rapid, non-destructive DNA isolation will help to reduce costs, speed up such processes as DNA barcoding, and allow the increased integration of new technologies in the development of accurate and rapid taxonomic identification.

Chapter 5

Molecular Phylogeny Supports the Paraphyletic Nature of the Genus *Trogoderma* (Coleoptera: Dermestidae) collected in the Australasian Ecozone

To date, a molecular phylogenetic approach has not been used to investigate the evolutionary structure of Trogoderma and closely related genera. Using two mitochondrial genes, Cytochrome Oxidase I and Cytochrome B, and the nuclear gene 18S, the reported polyphyletic positioning of Trogoderma was examined. Paraphyly in Trogoderma was observed, with one Australian Trogoderma species reconciled as sister to Anthrenus, Anthrenocerus and Orphinus genera. In addition, time to most recent common ancestor for Trogoderma and four closely related genera from the Megatomini and Attagenini tribes was calculated. Estimation of Dermestidae origins exceeded 175 million years, placing the origins of this family in Pangaea. This chapter has been accepted and published in the journal of Bulletin of Entomology and is appended tin appendix IV.

5.1 Introduction

The Family Dermestidae is a group of mostly xerophilic, necrophagic beetles that scavenge for dried, proteinaceous materials such as dried fish, meat, and hides (Kiselyova and McHugh, 2006, Zhantiev, 2009, Peacock, 1993). This type of material is consumed mainly during the larval stages, whereas the majority of adults feed on plant material, generally nectar and pollen, to help with egg maturation (Kiselyova and McHugh, 2006, Zhantiev, 2009). Exceptions do occur; for example, adults within the *Dermestis* genus do not feed on plant material and some species of *Trogoderma* can complete their entire lifecycle on plant material, making them pests of stored products (Mroczkowski, 1968, Hinton, 1945).

The modern classification of the Dermestidae recognises five subfamilies: Dermestinae, Attageninae, Trinodinae, Thylodriinae, and Anthreninae (Hinton, 1945). Of particular interest to this study is the subfamily Anthreninae and within it the tribe Megatomini. This subfamily is widespread, with most of the 27 extant genera occurring in the Palaearctic (11), Australian (9), and Oriental regions (4;(Peacock, 1993, Mroczkowski, 1968). One of the largest genera within this subfamily and Megatomini tribe is the *Trogoderma* genus containing approximately 130 described species (Hava, 2003). *Trogoderma* species are geographically widely distributed and found in most zoogeographical regions defined by Mroczkowski

(1968), excluding Antarctica and Ethiopian regions. The majority of species inhabit Australia (52 species not taking into account the proposed synonymies, A. Szito pers. comm..) and the Neotropics (30 species;(Mroczkowski, 1968). Dispersal routes for *Trogoderma* have been proposed to have followed a similar path to mammals, migrating from North America to South America and into Australia at about 30 to 65 million years ago (MA; Kiselyova and McHugh, 2006). However, this scenario does not explain how several *Trogoderma* species come to exist in the Palaearctic region, of which a few species are major pests of stored grains (Hinton, 1945). The two main pest species are *T. granarium* and *T. variabile*. Interestingly, the native Australian *Trogoderma* species are not pests of stored products, yet some species belonging to other Megatomini genera, such as *Anthrenocerus australis* Hope and *Orphinus fulvipes* Guérin-Méneville, have been reported as minor pests to stored products other than grain (Hinton, 1945).

The latest phylogenetic study on the dermestid family, using larval morphological characteristics, was conducted by Kiselyova and McHugh (2006). The authors observed different *Trogoderma* species scattered throughout the crown of the tree and concluded that the genus *Trogoderma* was polyphyletic, but were unable to ascertain whether this was a result of inadequate larval data or due to the incomplete state of dermestid taxonomy (Kiselyova and McHugh, 2006). Furthermore, in three instances the relationship between *Trogoderma* and closely related genera were unable to be resolved and led to polytomy of the phylogenetic trees.

As the dermestid fossil record is poor (Mroczkowski, 1968, Kiselyova and McHugh, 2006), a molecular clock in conjunction with the biogeographical distribution of several dermestid genera was used to investigate the origins of the Family Dermestidae. Since minimal molecular studies have been performed on this family, molecular clock data generated by Hunt *et al.* (2007) will be used as a basis for this study. Dermestids, belong to the series Bostrichiformia (Crowson, 1981) whose origins have been estimated at 219 MA (Hunt *et al.*, 2007). Assuming the Dermestid ancestors fed on angiosperm pollen, it is plausible to suggest that their origins are earlier than the oldest angiosperm pollen fossils which have been dated at 130 MA (Brenner, 1996). However, molecular data suggests that the radiation of angiosperms was somewhat earlier, 190 MA (Sanderson and Doyle, 2001). Beal and Zhantiev

(2001) did suggest that the Dermestidae origins were Pangean when a new species of *Egidyella*, only thought to occur in the deserts of central Asia, was found inhabiting the sand dunes of California. Therefore, on the balance of current available evidence, it is possible that the Dermestidae origins are Pangean and >175 MA.

The objective of this study was to evaluate the existing morphologically based polyphyletic positioning of *Trogoderma* (in species both exotic and endemic to Australia) using a molecular phylogenetic approach. The second objective was elucidation of the molecular relationships between five related genera: *Anthrenocerus* Arrow, *Anthrenus* Müller, *Attagenus* Latreille, *Dermestes* Linnaeus, *Orphinus* Motschulsky and *Trogoderma* Dejean, in order to better understand evolutionary processes that may have led to their current biogeographical distributions, specifically, to investigate the relationship between the pest *Trogoderma* species and the Australian native *Trogoderma*. To study these relationships *COI*, *CYT b* and the entire 18S nuclear gene were selected. The entire 18S was selected as it might be required to resolve deep polyphyletic structuring that was inferred in the earlier morphology study of Kiselyova and McHugh (2006).

5.2 Materials and Methods

5.2.1 Samples

More than 3,000 dermestid samples were collected during the Australian spring and summer (September to February) for three seasons from 2007 to 2010. Specimens were collected by either sweep netting or using a pheromone based trap. The traps were a Gypsy Moth bell trap (APTIV Inc) baited with a Khapra beetle kairomone attractant lure (Trécé Inc). The traps were hung from low-lying tree branches (<1m from the ground) then filled with 20% propylene glycol. Every 2 months, the propylene glycol was transferred into plastic bottles and a new pheromone lure was added to the trap. A sieve with a 1mm gauge was used to remove the specimens from the propylene glycol. The specimens were then rinsed with sterile water, followed by a 70% ethanol rinse, and finally stored in 95% ethanol at -20 °C. One specimen each of *T. variabile*, *T. granarium*, *Anthrenocerus*, *Attagenus* and *A. verbasci* were obtained from the Department of Agriculture and Food Western Australian insect

collection, and were used as representative reference samples to help demarcate lineages.

5.2.2 Morphological Identification

Due to the large number of specimens the first step towards identification was to separate specimens into morphospecies, groups that shared similar morphological features. These features included club size and number, setae colour, and body shape. One to four specimens were taken from each group and used in the subsequent study (Table 5.1). Due to the absence of morphological keys for most of the Australian native Megatomini, all the specimens were identified to genus level only, using descriptors outlined by Peacock (1993) and Booth *et al.* (1990). The genera *Orphinus*, *Trogoderma* and *Anthrenocerus* were differentiated from other genera within the tribe Megatomini by their well defined antenna cavity (Figure 5.1.) and separation of the middle coax (Peacock, 1993). The antenna cavity, or 'fossa', is either open or closed at the posterior end. *Orphinus* can be easily identified because the apical antenna segment is broader and longer than the penultimate (forming a distinct two-segmented antenna club), with an open fossa (Peacock, 1993). The Genus *Trogoderma* has a three-eight segmented antennal club (joined more or less symmetrically along the central axis), with an open fossa (Peacock, 1993). *Anthrenocerus* is identified by its well-differentiated three-segmented antenna club and closed fossa (Booth *et al.*, 1990). *Attagenus* was distinguished by the length of the hind tarsus (whereby the first segment is half or less than half the length of the second) and three segmented antenna club (Booth *et al.*, 1990). Specimens of exotic *T. granarium*, *T. variabile* and *A. verbasci* were identified to species level using the morphological keys described by Peacock (1993).

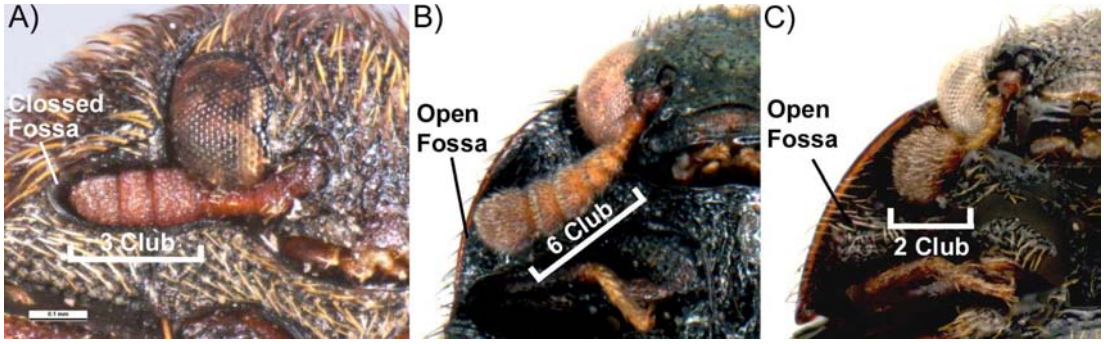

Figure 5.1 Photographs of A) *Anthrenocerus*, B) *Trogoderma* (Sample 107), and C) *Orphinus* (Sample 116), highlighting the open and closed fossa, and antenna clubs. These morphological features are used to differentiate these genera. The Anthrenocerus photograph courtesy of the Pest and Disease Image Library (PaDIL; http://www.padil.gov.au/).

5.2.3 DNA extraction

DNA was extracted from the samples using EDNA HiSpEx the non-destructive method described in Chapter 2.1.

5.2.4 Amplification

Amplification of *COI*, *CYT b* and *18S* gene was performed using primer pairs C1-J-2183 and TL2-N-3014, CB-J-10933 and CB-N-11367, all four *18S* primers, and the *T. granarium* primers Tg-J-Cytb and Tg-N-Cytb (Chapter 2 Table 2.1). The methodology for the PCR reaction mixtures, PCR cycling parameters, electrophoresis, DNA purification, and sequencing are described in Chapter 2.

Table 5.1. Specimens used in the phylogenetic study, including collection method, location and genetic loci analysed.

ID	Origin	Collection Sites		*Genus*	*species*	Collection	Lat	Long	*COI*	*CYT b*	*18S*
		State	Region								
0	Palaearctic	WA	Australia	*Anthrenus*	*verbasci*	Sweep	-31.9830	115.8619	✓	✓	✓
71	Palaearctic	WA	Australia	*Anthrenus*	*verbasci*	Sweep				✓	✓
73	Australia	WA	Australia	*Anthrenocerus*		Sweep	-33.8386	117.1511	✓	✓	✓
80	Unknown	WA	Australia	*Attagenus*		Sweep	-32.1080	115.8658	✓		✓
96	Australia	WA	Australia	*Trogoderma*		Wet Trap	-31.9830	115.8619	✓	✓	✓
97	Australia	Qld	Australia	*Trogoderma*		Wet Trap	-27.4000	153.0667	✓	✓	✓
98	Australia	Qld	Australia	*Trogoderma*		Wet Trap	-27.4000	153.0667	✓	✓	
105	Australia	Tasmania	Australia	*Trogoderma*		Wet Trap	-41.2600	147.5825	✓		✓
106	Australia	Tasmania	Australia	*Trogoderma*		Wet Trap	-41.2600	147.5825	✓		✓
107	Australia	ACT	Australia	*Trogoderma*		Wet Trap	-35.1627	149.0640	✓	✓	✓
115	Australia	Qld	Australia	*Orphinus*		Wet Trap	-27.4000	153.0667		✓	
116	Australia	Qld	Australia	*Orphinus*		Wet Trap	-27.4000	153.0667	✓	✓	✓
119	Australia	Qld	Australia	*Orphinus*		Wet Trap	-27.4000	153.0667	✓	✓	✓
120	Australia	Qld	Australia	*Orphinus*		Wet Trap	-27.4000	153.0667	✓	✓	✓
121	Australia	Qld	Australia	*Trogoderma*		Wet Trap	-27.4000	153.0667	✓		✓
124	India		Sri Lanka	*Trogoderma*	*granarium*	Quarantine			✓	✓	✓
130	Unknown	WA	Australia	*Anthrenus*		Sweep	-32.5128	118.4861		✓	✓
132	Australia	WA	Australia	*Trogoderma*		Sweep	-32.3259	118.9491	✓	✓	✓
134	Australia	WA	Australia	*Trogoderma*		Sweep	-32.3259	118.9491	✓	✓	✓
135	Australia	WA	Australia	*Trogoderma*		Sweep	-32.3259	118.9491		✓	
137	Australia	WA	Australia	*Trogoderma*		Sweep	-34.3193	118.7918		✓	✓
138	Australia	WA	Australia	*Trogoderma*		Sweep	-34.3193	118.7918	✓	✓	✓
142	Australia	WA	Australia	*Trogoderma*		Sweep	-33.3965	119.9092	✓	✓	✓
146	Australia	WA	Australia	*Trogoderma*		Sweep	-34.3918	118.0629	✓	✓	✓
147	Australia	WA	Australia	*Trogoderma*		Sweep	-34.3918	118.0629	✓		
153	Australia	NSW	Australia	*Orphinus*		Wet Trap	-35.0577	147.3580		✓	✓
157	Australia	WA	Australia	*Trogoderma*		Wet Trap	-34.4069	117.9218	✓	✓	✓
160	Australia	WA	Australia	*Trogoderma*		Wet Trap	-34.3330	118.1208	✓	✓	
165	Australia	WA	Australia	*Trogoderma*		Wet Trap	-34.3330	118.1208		✓	
168	Australia	WA	Australia	*Trogoderma*		Wet Trap	-33.5019	117.4084		✓	✓
170	Australia	WA	Australia	*Trogoderma*		Wet Trap	-34.3925	118.0665	✓	✓	✓
176	Australia	WA	Australia	*Trogoderma*		Wet Trap	-34.3512	118.0690	✓		
179	Australia	WA	Australia	*Trogoderma*		Wet Trap	118.0867	-34.3789	✓	✓	
184	Palaearctic	Tas	Australia	*Trogoderma*	*variabile*	Quarantine			✓	✓	✓

*** Note: Quarantine intercepts are made at the border in Australia and exact origins overseas is uncertain.**

5.2.5 Data Analyses

All sequences were edited using CodonCode Aligner 3.0.3 (CodonCode Corporation), and the sequences associated with each locus were aligned using the built-in version of MUSCLE (Edgar, 2004). MEGA 4 (Tamura *et al.*, 2007) was used to calculate the Maximum Likelihood (ML) distances between individuals. The model used was Tamura-Nei, as it takes into account different substitution rates between nucleotides and corrects for multiple hits. The ML distance values were plotted against the number of transition and transversion to examine the levels of saturation at each codon position. The phylogenetic signal of each mitochondrial gene and the *18S* were evaluated in PAUP 4.0 (Swofford, 2003) using the Partition Homogeneity Test (PHT). The parameters include hsearch, randomseed=0 and nrep=1000.

To examine the polyphyletic positioning of *Trogoderma*, the sequences from all three genes regions were concatenated to create an all-gene dataset. The Dermestidae lineage was rooted with *Tribolium castaneum* Herbst, of the series Cucujiformia, for both Parsimony and Bayesian analysis. Bayesian trees were generated using Mr Bayes 3.1.2 (Ronquist & Huelsenbeck, 2003) following the evolution model of GTR+I+Γ which was determined by jModelTester using both AIC and BIC (Posada, 2008). The mitochondrial protein-coding genes (*COI* and *CYT b*) were divided into six partitions, each partition corresponding to the first, second and third codon positions, and the *18S* was not partitioned. The program was run for gen=40,000,000, sample freq=1000, sump burnin=25,000 and sumt burnin=25,000. PAUP 4.0 was also used to generate Parsimony trees to compare against the Bayesian trees. Due to saturation at the third codon position, several approaches were adopted to generate the best tree. First, for each gene, the first, second and third codon positions were analysed separately; then, each tree was compared with majority rule to determine which position was forcing particular tree topologies. Second, total evidence trees were generated whereby we either weighted the first position at two or weighted the transition and transversions (1:2). The Parsimony trees were calculated under the following conditions: hsearch, addseq=random, nrep=10,000, swap=TBR, MaxTrees=Auto, with the additional parameters nchuck=5 and chuckscore=1 added for the bootstrapping analysis.

To estimate time to most recent common ancestor (TMRCA), Bayesian evolutionary analysis by sampling trees (BEAST: Drummond & Rambaut, 2007) was used. This allows for the nodes within the phylogenetic tree to be dated, which may help determine origins. The same concatenated dataset that generate the Parsimony and Bayesian trees was used. The Dermestidae lineage was rooted with *Tribolium castaneum* Herbst, of the series Cucujiformia. The root height parameters were set to the following: prior distribution normal, mean 235 MA (std dev 10 MA). These dates correspond to origins of the series Cucujiformia, with the low boundary incorporating the origins of Bostrichiformia (series of Dermestidae: Hunt *et al.*, 2007). The *Dermestes* and *Trinodies* samples were clustered and the root height parameters were set to the following: prior distribution lognormal mean=2, SD=0.5, and offset lower bounds of 65 MA, which corresponds to the earliest fossil record (Cockerell, 1917). Using the strict clock, both GTR and HKY evolution models were evaluated, gen=50,000,000 and log=500. Tree Annotator v1.5.4 was used to generate the consensus tree, using the following parameters: burnin at 10,000, posterior probability 0.95 and mean heights (Drummond & Rambaut, 2007).

5.3 Results

5.3.1 Mitochondrial sequence variation

The 743 bp of the mitochondrial *COI* was highly polymorphic, with 323 parsimony-informative nucleotide sites from 23 individuals. The average sequence divergence, calculated using ML distance, was 0.26 and the highest level of variation between Dermestid sequences was 0.36 (Table 5.2). Cytochrome b (430 bp) was similarly polymorphic with 188 parsimony-informative sites from 17 individuals. Despite both genes having variation occurring at 43% of their sites, *CYTb* was considerably more variable with a divergence range of 0.25 to 0.46, the highest between two *Trogoderma* species collected in Australia (T96 and T134). The polymorphisms observed, within the mitochondrial genes, resulted in 63 (25%) and 34 (23%) non-synonymous changes in the amino acid sequence. No insertions or deletions were observed in the mitochondrial gene fragments. In addition, no stop codons were observed when the sequences were translated, suggesting that there were no NUMTs sequence within the dataset.

Table 5.2. Locus information including variable sites, pairwise divergence (%) values and transition/transversion ratios.

Gene	Codon position	Conserved	Informative sites	ML distance (Tamura-Nei)
COI	All	366	323	0.12–0.35 (0.26)
	1	157	73	
	2	201	33	
	3	8	217	
CYT b	All	217	188	0.25–0.46 (0.3)
	1	92	43	
	2	120	16	
	3	5	129	
18S		1761–1777	78	0.002–0.038 (0.019)

5.3.2 Saturation curves

The number of transitions and transversions for each codon position and mtDNA gene were plotted against ML distances to examine sequence saturation (Figure 5.2). For both genes, transition saturation was evident at the third position, illustrated by the plateauing of the polynomial curve. No transition saturation was observed at the first and second positions, but rather an upward directional curve. Similarly, transversion saturation was not observed at the first and second positions, only at the third positions.

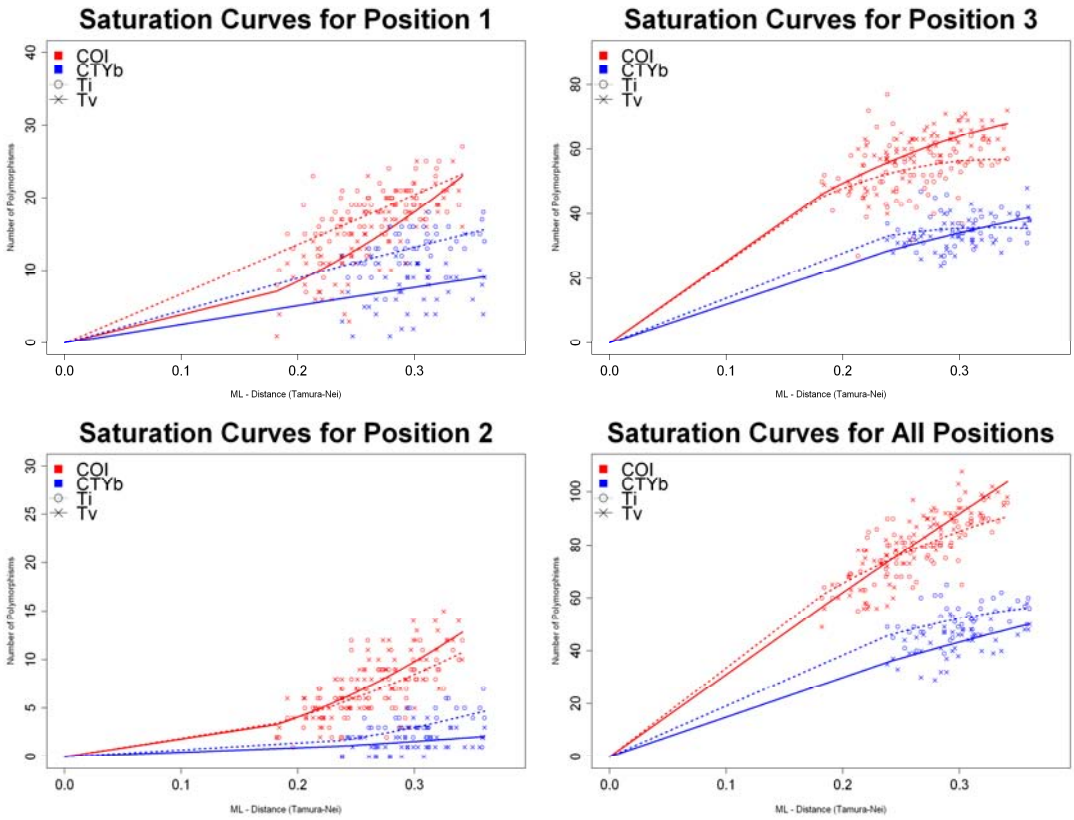

Figure 5.2. Saturation Plots for the three codon positions.

5.3.3 Analyses of 18S

The *18S* ranged from 1839 to 1855 bp in length in which there were 78 parsimony-informative sites (Table 5.2) from 26 individuals. These polymorphisms included 134 single nucleotide polymorphisms and 39 indels. As expected, *18S* showed evidence of much slower evolution than *COI* and *CYT b*, with average sequence divergence of only 0.019, in comparison to 0.26 and 0.3, respectively. Limited variation was observed between the variable domains of the *18S*, with only 21 polymorphic sites, of which five were parsimony-informative. Variation within the seven variable domains ranged from two to 46 polymorphic sites. Three variable domains were highly variable (V2, V4 and V7) containing 40, 46 and 44 polymorphic sites, respectively. Sequence length variation was only observed within the V2, V4, V7 and V9 domains and was the result of increased sequence length of the *Attagenus* sample and Australian *Trogoderma* sample 96.

5.3.4 Partition homogeneity test

The different number of polymorphic sites observed within all of the variable domains indicated that each domain was under varying evolutionary constraints (Weider *et al.*, 2005). To determine if this may lead to incongruence between the nuclear and mitochondrial genomes, a PHT was performed to compare the variable domains to each other as well as the mitochondrial genes. The PHT revealed that there was no significant incongruence between any of the gene fragments. Therefore, the gene fragments were concatenated and used to generate a total evidence tree.

5.3.5 Bayesian and parsimony trees

The primer pairs did not successfully amplify all specimens; therefore, there are differing numbers of total *COI*, *CYT b* and *18S* sequences available in this dataset. The high levels of variation within the mitochondrial genes suggested possible mismatch in the priming sites as a cause for failure to amplify. Seven pairs of alternative primers were trialled but amplification was refractory. Samples with missing sequences were included in the analysis, but the sequence data was treated as missing. Maximum likelihood distances (Tamura-Nei) were calculated for the combined *COI* and *CYT b*, and *18S* for the 23 individuals (Table 5.3).

The Bayesian phylogenetic tree identified *Trogoderma* as paraphyletic with respect to *Dermestes*, *Trinodes*, *Attagenus*, *Anthrenus*, *Orphinus* and *Anthrenocerus* (Figure 5.3a). The majority of the *Trogoderma* and *Anthrenocerus* species form a well-supported sister group to the *Orphinus* sample (posterior probability 1). The only exception was the Australian *Trogoderma* sample 96 that was sister to the rest of the ingroup. This sample was morphologically re-examined and confirmed to be part of the *Trogoderma* genus. The exotic *Trogoderma* species, originally from the Palaearctic and Indian regions, were strongly supported sister taxa (posterior probability 1.00; Figure 5.3a). This exotic *Trogoderma* clade was sister Australian *Trogoderma* clade and the *Anthrenocerus* samples. The two *Anthrenocerus* samples were nested deep within the Australian *Trogoderma* clade and sister to T170. While the Bayesian analysis was able to resolve all but the deepest internal node, the majority of internal branches within the discordant zone (grey box; Figure 5.3) were short.

Table 5.3. Maximum Likelihood distances between the 23 individuals that were used to construct the phylogenetic trees; below the diagonal is *COI* and *CYT b* combined, and above the diagonal is the *18S*.

			Dermestes			***Attagenus***				**Exotic *Trogoderma***			**Western Australia *Trogoderma***						**Eastern *Trogoderma***			***Anthrenocerus***	
	T. castan..	T96	*T. rufes*	*D. laniar*	*D. muri*	*A. unkn*	*A. pellio*	*A. v*	*Orphinus*	T103	T184	T124	T142	T134	T138	T132	T170	T157	T97	T105	T107	A72	A73
T. castanuem		0.017	0.028	0.030	0.027	0.032	0.027	0.028	0.032	0.029	0.029	0.029	0.031	0.030	0.030	0.030	0.029	0.029	0.030	0.030	0.030	0.033	0.032
T96	0.31		0.024	0.026	0.021	0.029	0.022	0.029	0.030	0.029	0.028	0.029	0.029	0.029	0.028	0.026	0.026	0.027	0.027	0.027	0.027	0.029	0.030
T. rufescens	0.36	0.34		0.031	0.029	0.034	0.028	0.032	0.038	0.035	0.035	0.035	0.036	0.035	0.036	0.034	0.033	0.035	0.034	0.035	0.035	0.035	0.033
D. laniarius	0.31	0.29	0.30		0.006	0.032	0.027	0.028	0.034	0.033	0.032	0.034	0.032	0.029	0.032	0.031	0.031	0.030	0.032	0.031	0.031	0.034	0.034
D. murinus	0.28	0.29	0.27	0.12		0.029	0.024	0.025	0.030	0.029	0.028	0.030	0.028	0.026	0.029	0.027	0.027	0.027	0.028	0.027	0.027	0.030	0.030
A. Unknown	0.31	0.25	0.31	0.27	0.26		0.011	0.027	0.031	0.029	0.026	0.029	0.030	0.025	0.030	0.027	0.028	0.028	0.027	0.027	0.029	0.029	0.031
A. pellio	0.33	0.28	0.35	0.29	0.30	0.21		0.021	0.026	0.026	0.023	0.026	0.027	0.023	0.027	0.023	0.024	0.026	0.024	0.024	0.025	0.026	0.027
A.verbasci	0.32	0.32	0.30	0.30	0.28	0.27	0.29		0.021	0.022	0.022	0.024	0.023	0.023	0.023	0.022	0.023	0.022	0.023	0.024	0.021	0.023	0.023
Orphinus	0.35	0.32	0.34	0.27	0.28	0.24	0.25	0.33		0.009	0.009	0.012	0.015	0.012	0.014	0.009	0.011	0.014	0.012	0.011	0.009	0.012	0.011
T103	0.32	0.34	0.32	0.26	0.28	0.26	0.30	0.32	0.29		0.004	0.004	0.014	0.009	0.013	0.008	0.007	0.012	0.008	0.007	0.009	0.011	0.010
T184	0.31	0.32	0.29	0.26	0.26	0.28	0.29	0.29	0.27	0.24		0.01	0.014	0.008	0.013	0.008	0.006	0.012	0.008	0.007	0.008	0.009	0.009
T124	0.33	0.37	0.34	0.28	0.30	0.30	0.31	0.31	0.28	0.24	0.23		0.013	0.008	0.012	0.005	0.004	0.013	0.005	0.004	0.006	0.009	0.008
T142	0.33	0.35	0.30	0.29	0.31	0.29	0.31	0.35	0.28	0.29	0.26	0.28		0.011	0.001	0.010	0.011	0.004	0.011	0.010	0.011	0.012	0.013
T134	0.37	0.39	0.30	0.28	0.27	0.27	0.32	0.32	0.32	0.29	0.29	0.32	0.28		0.011	0.007	0.007	0.011	0.006	0.006	0.008	0.008	0.009
T138	0.33	0.35	0.33	0.27	0.30	0.27	0.27	0.29	0.28	0.25	0.27	0.27	0.23	0.25		0.009	0.009	0.003	0.009	0.009	0.009	0.011	0.012
T132	0.34	0.33	0.28	0.25	0.24	0.24	0.25	0.30	0.28	0.25	0.25	0.25	0.24	0.26	0.22		0.001	0.011	0.002	0.002	0.002	0.006	0.006
T170	0.31	0.29	0.27	0.27	0.27	0.25	0.22	0.28	0.23	0.26	0.22	0.27	0.27	0.25	0.22	0.23		0.011	0.002	0.002	0.002	0.006	0.006
T157	0.35	0.31	0.31	0.27	0.27	0.26	0.27	0.29	0.24	0.27	0.26	0.28	0.27	0.25	0.24	0.22	0.22		0.011	0.010	0.011	0.012	0.014
T97	0.36	0.31	0.27	0.28	0.24	0.25	0.29	0.30	0.29	0.31	0.27	0.32	0.27	0.29	0.24	0.27	0.24	0.25		0.001	0.004	0.007	0.008
T105	0.31	0.36	0.34	0.28	0.27	0.28	0.32	0.30	0.28	0.28	0.26	0.26	0.26	0.25	0.24	0.23	0.25	0.24	0.22		0.004	0.007	0.008
T107	0.32	0.30	0.30	0.26	0.25	0.26	0.24	0.30	0.26	0.27	0.26	0.31	0.26	0.27	0.23	0.25	0.21	0.27	0.24	0.24		0.007	0.006
A72	0.34	0.37						0.29	0.32	0.30	0.29	0.27	0.30	0.32	0.27	0.26	0.23	0.28	0.26		0.30		0.002
A73	0.33	0.32	0.33	0.25	0.26	0.26	0.28	0.30	0.26	0.29	0.28	0.28	0.27	0.28	0.22	0.24	0.24	0.24	0.26	0.25	0.24	0.15	

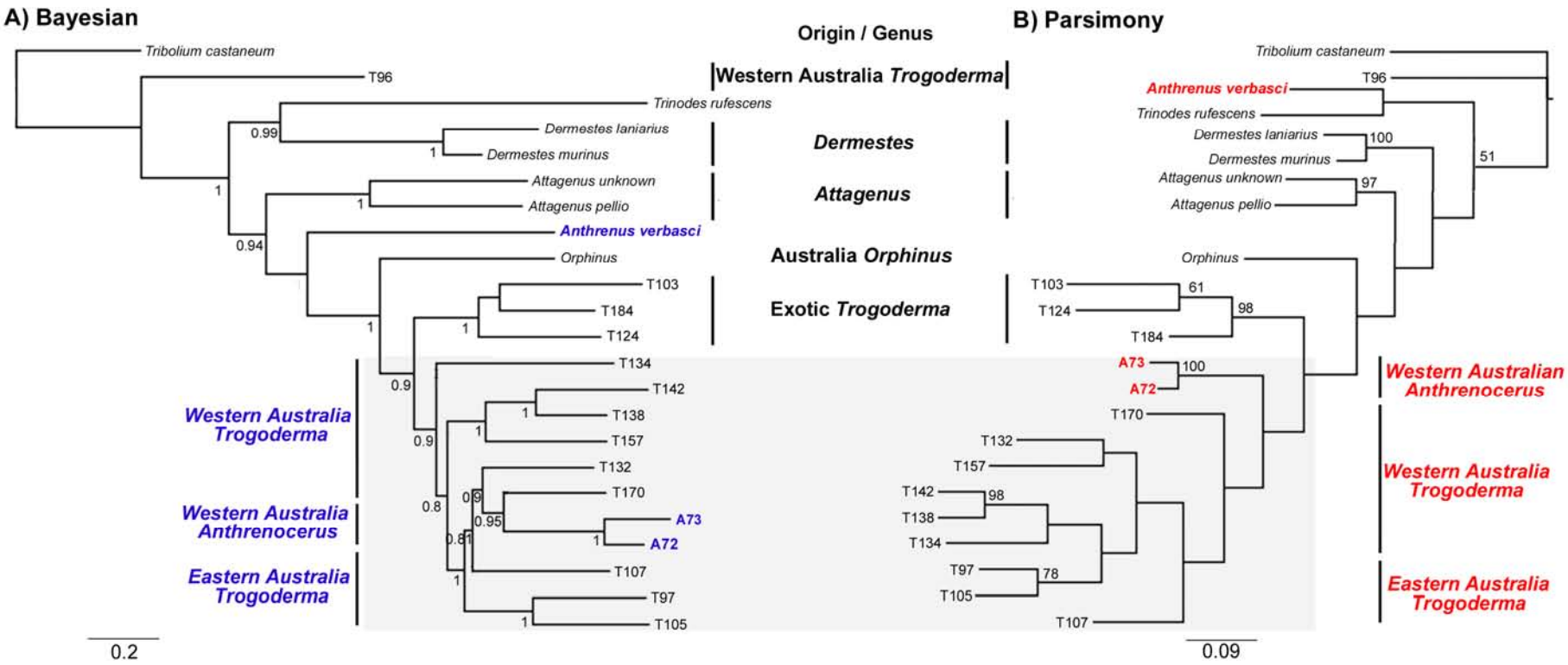

Figure. 5.3. (a) Bayesian and (b) Parsimony trees based on the combined *18S*, *COI* and *CYT b* datasets. The value below the branch of the Bayesian tree represents the Bayesian posterior probably. The values below the line represent of the Parsimony tree represents bootstrapping values. Nodes with <50% bootstrapping support were not shown on the tree. The outgroup for both trees was *T. castaneum*.

Due to saturation at third position, each codon position was concatenated with the *18S* and analysed separately using Parsimony (data not shown). All parsimony trees supported the same deep nodes as the Bayesian analysis, but major structural change was observed below the *Attagenus* clade and within the discordant zone. The first codon position supported the placement of the exotic *Trogoderma* but positioned *Anthrenocerus* as sister to the Australian *Trogoderma* and clustered *A. verbasci* with *T. rufescens.* The second position nested the exotic *Trogoderma* and *Orphinus* deep within the Australian *Trogoderma* clade. Interestingly, the only three polymorphisms between *T. granarium* (T124), T107 and T134 were observed. The placement of *Anthrenocerus* mirrored that of the Bayesian analysis. The third position also placed the exotic *Trogoderma* and *Anthrenocerus* deep within the Australian *Trogoderma* clade. This placement of the exotic *Trogoderma* was rejected as it was neither biologically or biogeographically sensible nor phylogenetically robust.

Each position was reanalysed with the transversions/transititions weighted 2:1. This made minimal difference and the same Parsimony trees were returned. Finally, the first position was weighted twice that of the other positions (2:1). Parsimony trees generated under these parameters generated four equally parsimonious trees, each with a score of 3857 (Figure. 3b), which was the closest match to the Bayesian tree (Fig. 5.3a). The main differences were *A. verbasci*, forming a clade with *T. rufescens*, and *Anthrenocerus* clade, which was sister to the Australian *Trogoderma*. Minimal bootstrapping support was generated on each node. Apart from the clades that had bootstrapping support >50%, all samples within the discordant zone shared different common ancestors.

5.3.6 BEAST analysis

Estimates of TMRCA were generated using BEAST (Drummond & Rambaut, 2007). Both GTR and HKY substitution models predicted effective sample sizes (ESS) >200, indicating a moderate degree of support for each node. GTR ESS values were slightly higher and, subsequently, this model was used to estimate TMRCA. The split between the series Bostrichiformia and Cucujiformia was estimated at 237 MA, with the ingroup estimated at 209 MA (Figure 5.4, Table 5.4). The TMRCA for Attagenus was calculated at 194 MA, while the common ancestor for *A. verbasci* was

181 MA. The mean estimate for the *Orphinus* common ancestor was 157 MA. The proposed common ancestor for exotic *Trogoderma* was 102 MA, while the split from the Australian *Trogoderma* and *Anthrenocerus* was 141 MA. The common ancestors for the Australian *Trogoderma* samples, excluding sample 96, ranged from 128 to 66 MA.

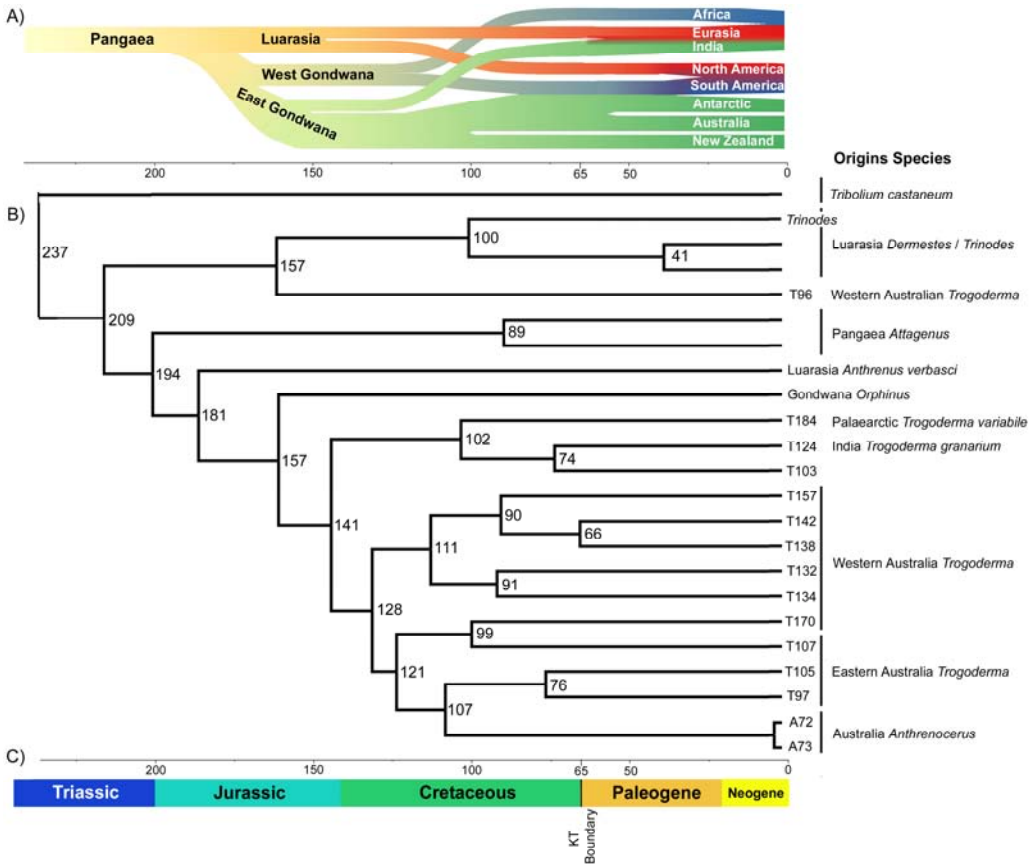

Figure 5.4. Phylogenetic history of the Dermestidae genera *Attagenus*, *Anthrenus*, *Anthrenocerus*, *Dermestes*, *Trogoderma* and *Orphinus*. (a) Timeline of the continental breakup and drift from 250 MA to present (based on Li & Powell, 2001). (b) Bayesian tree with estimations of time to most recent common ancestor (MA). Mean time estimates are shown after the node. Scale bars are located above and below the phylogenetic tree and represent millions of years. The outgroup *T. castaneum* is from the superfamily Curcjiformia. (c) The geologic timeline from the Triassic to the present.

Table 5.4. Time to most recent common ancestor using BEAST.

Genus	Mean (MA)	95% HPD lower	95% HPD upper
Root	237	224	247
Dermestes	209	198	217
Attagenus	194	180	205
Anthrenus	181	168	192
Orphinus	157	146	165
Trogoderma / Anthrenocerus	141	137	159
Australian *Trogoderma / Anthrenocerus*	128	122	143
Exotic *Trogoderma*	102	96	119

5.4 Discussion

Despite discordance between Bayesian and Parsimony approaches, the genus *Trogoderma*, according to this study and excluding *Trogoderma* sample 96, is paraphyletic with respect to *Anthrenocerus*. The current taxonomic nomenclature of the Megatomini tribe, as described by Peacock (1993) and Booth *et al.* (1990), uses two main morphological characteristics to differentiate *Anthrenocerus* and *Trogoderma*. The main morphological characteristics used are the number of antenna segments that form the club and the open or closed fossa. The *Anthrenocerus* genus is identified by the three segmented clubs joined eccentrically. In comparison, *Trogoderma* have three to eight segmented clubs, joined more or less symmetrically along the central axis. However, several species of *Trogoderma*, held in the Department of Agriculture and Food Western Australian insect collection, have well-defined three-segmented clubs that are eccentrically joined. The only definitive difference between these two genera is the closed antennal fossa found in *Anthrenocerus* species. Our study, similar to the morphological study conducted by Kiselyova & McHugh (2006), suggests that *Anthrenocerus* is possibly an Australian *Trogoderma* variant.

The morphological characteristics used to delineate *Trogoderma*, as well as the results presented here, suggest that the body plan of *Trogoderma* has been retained throughout history (parallelism) rather than arising independently (convergent evolution: Wiens *et al.*, 2003). This may give the impression that morphologically similar species are more closely related than they are genetically, a common phenomenon (Bossuyt & Milinkovitch, 2000). The number of morphological characteristics used to delineate between the three *Trogoderma* groups (sample 96, the exotic *Trogoderma* and the Australian native *Trogoderma*; Figure 5.3b) are minimal. One of the main external differences between the exotic and Australian native *Trogoderma* are the lighter-coloured patterns in the integument of the elytra of the exotic *Trogoderma* and setae colour. Interestingly, morphological comparisons between the Australian native *Trogoderma* and *Trogoderma* sample 96 were negligible, yet the phylogenetic trees grouped these individuals at opposite ends of the tree and a ML distance >0.29. The conservation of morphological characters is not uncommon and has been observed in frogs (Bossuyt & Milinkovitch, 2000;

Chiari *et al.*, 2004), bacteria (Degnan *et al.*, 2005) and other insects (Buckley *et al.*, 2009). This suggests parallelism of the *Trogoderma* characteristics and that, even with major morphological taxonomic revision, the paraphyly in *Trogoderma* may not be easily resolved.

A high degree of branch swapping was observed when the Parsimony and Bayesian approaches were compared, despite Bayesian generating trees with high levels of posterior probability support on all nodes. This lack of agreement between Bayesian and Parsimony trees is not uncommon and is a phenomena previously observed (Shoup & Lewis, 2003; Khan *et al.*, 2008). To explain this lack of congruence between the two phylogenetic approaches, three commonly attributed causes were investigated: homoplasy, gene divergence rates and lineage sorting (O'Huigin *et al.*, 2002; Cantarel *et al.*, 2006; Rasmussen & Kellis, 2007; McCormack *et al.*, 2009).
The saturation plots illustrated that transitions and transversion saturation had been reached at the third position, suggesting that homoplasy is present within our dataset. The presence of homoplasy at the third position may make it difficult for Parsimony to generate accurate trees as no nucleotide substitution models are employed and the tree with least number of steps is preferred (Page & Holmes, 1998). Furthermore, at the second position, only three polymorphisms between *T. granarium* (T124), T107 and T134 were observed, making these species look more closely related than they are at other base sites. In this study, both the presence of homoplasy and invariable sites may have contributed to inaccuracies between trees.

The ML distances for *COI* and *CYTb* were considerably high, with the *Trogoderma* distance only 0.21 (between T170 and T107; Table 5.3). Cytochrome b, which has previously been reported as a slower evolving gene (Castresana, 2001), in our study, mutated at a greater rate than *COI.* In comparison, minimal divergence was observed within the *18S.* Previous studies examining the effect that divergence has on tree accuracy concluded that accuracy and divergence are proportional (Cantarel *et al.*, 2006; Rasmussen & Kellis, 2007). Rapidly evolving genes can write over the phylogenetic signals, whereas slowly evolving genes may have insufficient informative sites. In our case, either increasing the number of closely related *Trogoderma* species or a gene evolving slower than *COI* and *CYT b* would likely help increase support within the discordant zone.

Alternatively, the polytomy observed in the Parsimony trees and the short internal branch lengths within the Bayesian trees may represent hard polytomy or lineage sorting. Lineage sorting is a process whereby three or more species radiate rapidly such that an ancient polymorphism is retained in all descendents and by chance each species may share a set of alleles at different loci (McCracken & Sorenson, 2005). Therefore, the Parsimony polytomy and short internal nodes from the Bayesian analysis may signify rapid radiation (hard polytomy) and lineage sorting, not in accuracies in the phylogenetic models or selected genes.

The interpretation of the biogeographical distributions and TMRCA (BEAST) for *Anthrenus*, *Anthrenocerus*, *Attagenus*, *Dermestes*, *Orphinus* and *Trogoderma* allows postulations about the origins of the Dermestidae family. Current biogeographical distribution of *Attagenus* species includes: the southern Palaearctic (45%), Ethiopian (35%), Oriental (7%) and Nearctic (5%) regions (Mroczkowski, 1968), which were all connected as part of the supercontinent Pangaea (Li & Powell, 2001). For *Dermestes*, the majority of the 73 species are found in the Palaearctic (50%) and Nearctic (26%; Mroczkowski, 1968). Approximately 175 MA, Pangaea began to break-up and a rift formed between North America and Africa, resulting in the continents Laurasia and Gondwana (Li & Powell, 2001). With a high proportion of *Attagenus* and *Dermetes* found on both Laurasia and Gondwana, the origins must be earlier than this vicariant event. Estimation of date of origin by BEAST placed the *Attagenus* and *Dermestes* samples near the base of the tree (Figure 5.4), a similar position to that in the morphological study by Kiselyova & McHugh (2006), with a value of 209 MA (±33 MA). Furthermore, molecular data suggests angiosperms evolved at, or after, 190 MA (Sanderson & Doyle, 2001), which together suggests that the origin of the Dermestidae family is the mid-early Jurassic (>175 MA) on the Pangean supercontinent.

Orphinus appear to have a Gondwanan distribution, with the majority of species found in the Australian, Oriental and Ethiopian regions (Mroczkowski, 1968). Based on biogeographical data, it seems reasonable to suggest that the origins of *Orphinus* are Ethiopian and prior to the west Gondwana (Africa, Arabia and South America) and east Gondwana (Australia, Antarctica, Madagascar and India) split, approximately 160 MA (Li & Powell, 2001). This placement of origin was supported

by BEAST that estimated the common ancestor of *Orphinus* at 157 MA. Future studies will need to include *Orphinus* from multiple biogeographical locations to determine if this group is monophyletic.

The origins of the *Trogoderma* and *Anthrenocerus* clade, excluding sample T96, were estimated at 141 MA. Using this date, the scenario, proposed by Kiselyova & McHugh (2006), that *Trogoderma* followed the dispersal of marsupial mammals (carcasses of which are a likely food source for larvae) from North America to South America and in to Australia around 30 to 65 MA is doubtful. Rather, it is plausible to suggest that *Trogoderma*'s origins are East Gondwanan, whereby they inhabited both the Indian and Australian continents. The exotic *Trogoderma*, inhabited the Indian continent which collided with Asia approximately 55 MA (Klootwijk *et al.*, 1992), allowing this genus to spread east into the Palaearctic. Rather than *Trogoderma* migrating from North America to South America and then to Australia, it is more likely that migration was in the opposite direction. We base this scenario on a number of factors. First, the number of *Trogoderma* species decreases as we move away from Australia. Second, the lack of species present within Ethiopian region and the date estimated by BEAST suggest that *Trogoderma* were not present prior to the Gondwana split. Third, several nodes within the *Trogoderma* clade are greater than 100 MA. Finally, this scenario provides a link between the Australian and Exotic *Trogoderma*. To further confirm this scenario, North and South American samples are required, and this could be a potential avenue of future work.

Despite the phylogenetic clustering of *T. variabile* and *T. granarium*, this may not reflect a similar capability of infesting stored products. While *T. granarium* is a major pest of stored products, this can attributed mostly to circumstance and possibly explained by the following evolutionary scenario. Agricultural activities require the storage of grains, which are frequently attacked by non-Dermestid insect pests (Hinton, 1945). As the population of non-Dermestid insect pests increases, so does the number of dead insects, which is a suitable food source for all Dermestids. As the population of Dermestids increase some individuals maybe more inclined to consume stored products. When this scenario is applied to the birth places of agriculture, such as the Palaearctic region (23,000 ya; Piperno *et al.*, 2004), Northern America, Mexico and South America (5,000 to 9,000 ya; Piperno *et al.*, 2009) it is

interesting to note that these areas are endemic to *Trogoderma versicolor*, *Trogoderma simplex*, *Trogoderma glabrum*, and *Trogoderma ornatum*, all minor pests of stored products (Hinton, 1945). In Australia, there has been no stored grain before the large scale agriculture started around mid-to late 19th century and the Australian flora encourage the native ants to collect their seeds. However, with similar behavioural and food sources between the dermestids species, it is conceivable that they all have the ability to become pests of stored products and why some species within *Anthrenus*, *Anthrenocerus*, and *Orphinus* are pests of stored products other than grain.

The phylogenetic analysis provided an interesting insight in to the evolution of the *Trogoderma* genus as well as five closely related genera from the Megatomini and Attagenini tribes. While this study was unable to resolve the polyphyly reported using the current taxonomic nomenclature, this study did highlight three interesting *Trogoderma* groups. The species within these groups need to be further scrutinised to determine if unique characteristics have been overlooked, which has resulted in the clumping of a number of species. However, it seems possible that *Trogoderma* is a common ancestor to a proportion of the Dermestidae genera, especially given the limited morphological characteristics that differ between *Anthrenocerus*, *Orphinus* and *Trogoderma,* and the minimal dietary differences between *Trogoderma* and all other dermestids. This study also highlights the importance of a well sampled dataset, especially if polytomy is caused by a lack of closely related samples as the addition of new species may alter the topography of the phylogenetic trees and lead to inaccurate assumptions.

Chapter 6

Dermestidae Species differentiated using a novel approach to Quantitative PCR

Accurate identification of T. granarium *is fundamentally important, however, morphology, DNA barcoding and traditional polymerase chain reaction have limitations in providing an accurate identification. To overcome the inherent limitations of these existing methods, a quantitative PCR multiplex test was designed using species specific primers and probes, an amplification control, followed by melt curve analysis. Each component of the multiplex reduced the incidence of both false positives and false negatives, thereby increasing specificity and sensitivity of the test.*

6.1 Introduction

Accurate identification of *T. granarium* is important to the Australian agricultural economy, as the potential cost of establishment of the pest in Australia is estimated to be $AU 1.8 billion annually (Mack *et al.*, 2000, Cook, 2003). In general, three methods are commonly used to identify a suspect sample: morphology, DNA barcoding and traditional PCR. However, each of these methods has limitations in providing an accurate identification. For *T. granarium*, adults and larvae that are in good morphological condition, morphological keys are reliable and accurate when used by a trained and skilled diagnostician. Difficulties arise when keys are geographically limited, are only effective for a particular life stage, or when specimens are damaged and missing the required diagnostic features (Hebert *et al.*, 2003). DNA barcoding can be reliably accurate when based on an exhaustive reference database (Meyer and Paulay, 2005), but requires medium to high quality DNA and can take days for a result to be generated. Traditional PCR methods including RAPDs, AFLPs, RFLPs, and species specific markers can also be problematic (Armstrong, 2010, Vignal *et al.*, 2002, Groth and Wetherall, 2000). RAPD markers are typically undesirable because of their dominant nature, high numbers of artifactual bands resulting from PCR induced mutations, and reproducibility between laboratories is frequently low (Groth and Wetherall, 2000, Vignal *et al.*, 2002). While RFLPs and AFLPs are reproducible and variable, they are time consuming (Vignal *et al.*, 2002, Groth and Wetherall, 2000). Species specific markers are highly sensitive when analysing a known population, but when a sample

is from an unknown population the test can fail (Meyer and Paulay, 2005). While each method is, to a certain extent, accurate it can be difficult to detect false positives and negatives.

For DNA barcoding, false positives can occur when an individual's genetic variation exceeds the reference group's intra-specific divergence level, resulting in the identification of a novel taxa (commonly referred to as 'splitting'; Meyer and Paulay, 2005). False negatives are inaccurate identifications that occur when the unknown specimen clusters within a reference group (lumping), which is caused by low intra-specific variation (Meyer and Paulay, 2005). Furthermore, the time and cost of sequencing is impractical in certain diagnostic situations, such as large scale surveillance programs. For species-specific PCR based methods, false negatives can result from the lack of amplification due to inhibitors of the PCR, damaged DNA or variation within the priming site's sequence (particularly at the 3` end; Peist *et al.*, 2001, Nolte, 2004). False positives can occur when there is a lack of primer specificity or cross contamination. It is important to identify and prevent both types of false diagnoses in pest identification programs, as misidentification may result in late initiation of containment or an eradication programs.

Specimens can be identified by quantitative PCR (or Real-time PCR), which measures the increase in amplified product in real time, thus removing the need for downstream analysis (Lyon and Wittwer, 2009, Wittwer *et al.*, 1997). The amplified product is measured by the addition of a fluorophore (either an intercalating dye or hydrolysis probe) to the PCR reaction. Intercalating dyes are generally aromatic cations that bind specifically to the minor grove of the DNA duplex. This arrangement creates a large increase in the fluorescence signal relative to the free dye in solution (Li *et al.*, 2003, Deligeorgieva *et al.*, 2009). Alternatively, a hydrolysis probe consists of a short oligonucleotide sequence with two linkers attaching a fluorophore to the 3` end and a quencher to the 5` end. When the probe anneals to a unique sequence during the PCR, the fluorophore is cleaved by the 5`-3` exonuclease action of the polymerase, resulting in a increase in fluorescence signal relative to the intact probes.

In addition to measuring amplification of a PCR product, as described above, species can also be identified through the use of melt curve analysis (MCA). After the PCR reaction is complete, the temperature of the reaction is increased to denature the dsDNA product. Once the dsDNA product reaches it's denaturation point, which is dependent upon its specific nucleotide content and concentration of salt ions in the solution, the intercalating dye can no longer interact with the DNA and fluorescence is lost (Li *et al.*, 2003). This technique has been shown to be highly sensitive at detecting sequence variation, to the extent of differentiating two sequences that differ by a single base pair mutation (Chang *et al.*, 1994), Therefore, the amplified PCR products from two species that differ in nucleotide composition or length will generate distinctly different melt curves.

In this study, a PCR based method is established using species specific primers and probes to detect and distinguish *T. variabile* and *T. granarium*. In addition, the multiplexing of species specific primers and probes, in conjunction with an internal amplification control, is assessed. Melt curve analysis is then applied to test for false positives and negatives, so that the incidence of these can be accounted for and reduced. This technique of using different methods acting as internal checks within the one procedure would create a novel Real-Time PCR multiplex test, which can be used to distinguish important *Trogoderma* species.

6.2 Materials and Methods

6.2.1 Samples

Two datasets were used to evaluate the efficacy of the *Trogoderma* multiplex test. Dataset one contained 116 samples, a sub sample of the 151 specimens analysed from the *T. variabile* phylogenetic study (Chapter 3). Phylogenetic analysis of these specimens revealed that there was a mixture of six unidentified species, denoted as species B to G, as well as *T. variabile*. The second dataset (dataset two) consisted of 40 specimens that were morphologically identified as: *T. variabile, T. granarium, T. versicolor*, *T. angustum*, *T. glabrum*, *T. ornatum*, and *T. megatomoides*. These specimens were either intercepted by United States Customs and Border Protection, Australian Quarantine and Inspection Service (AQIS) or obtained from the Berlin museum collections. Reference specimens of *T. variabile*, *T. granarium*, and a native

Orphinus species were taken from the Department of Agriculture and Food Western Australian insect collection, identified using traditional morphological keys (Peacock, 1993), and were subsequently defined as representative reference samples used for this study.

6.2.2 DNA extraction

For samples from dataset one, the method used to extract DNA is described in Chapter 3.2.2.

For samples from dataset two and the reference samples, DNA was extracted using the non-destructive method described in Chapter 2.1

6.2.3 Development of Diagnostic Primers

The primers and probes used to create the *Trogoderma* multiplex test were designed using Beacon Designer 7.6 (www.premierbiosoft.com). The amplification control primers were designed from the *18S* sequences generated during the CSIRO *T. variabile* study (Genbank Accession numbers HM243239 - HM243268; Chapter 3). These sequences were aligned using CLUSTALW and primers were designed to anneal to the conserved regions that surrounded the variable part of the V4 domain. As previously shown (Chapter 3; Table 3.2) the 18S fragment, which contains the V4 domain, has no intra-specific variation, but, inter-specific variation. For *T. granarium* and *T. variabile,* species specific primers and hydrolysis probes were designed using the *COI* sequences obtained from the CSIRO *T. variabile* study (Genbank Accession numbers HM243383 - HM243470; Chapter 3). Beacon Designer 7.6 generated and rated each primer pair. Primers that were unlikely to form primer dimers and duplex formations were selected and subsequently used in the multiplex test. The hydrolysis probes were synthesised by Biosearch Technologies (www.biosearchtech.com); *T. variabile* 5' CAL Fluor Red 610 and 3' BHQ-2, and *T. granarium* 5' CAL Fluor Orange 560 and 3' BHQ-1 (Table 6.1).

Table 6.1. Details for the primers used to create the *Trogoderma* multiplex test (amplification control and species specific primers).

Primer Name	Sequence `5- 3`	Region	Size (bp)	Type
18SF	GCACTCTAATTTGTTCAAAG	Nuclear	250	Amplification
18SR	GTAATTCCAGCTCCAATAG			Control
COI Tv probe	TCCTCCATGACACATACTACGT			Species
COI Tv F	CTGGAGTTGTTCTAGCTA	Mitochondrial	200	Specific
COI_Tv_R	GGGAAAGAATGTTAGGTTT			(*T. variabile*)
COI Tg probe	CGTGGTATTCCTCTCAGTCCTAAG			Species
COI Tg F	CTATCAATAGGAGCAGTATTC	Mitochondrial	200	Specific
COI Tg R	ATGTTGTGTAAGCATCTG			(*T. granarium*)

6.2.4 Optimisation of PCR primers

Specificity of the species specific primers and the universality of the amplification control primers were confirmed by trialling each primer set in a singleplex assay. The primers were tested on the *T. granarium* and *T. variabile* reference specimens. Cycling conditions were optimised by gradient PCR, with annealing temperatures ranging from 51-57 ºC at intervals of 2 ºC. The PCR reaction mix was the standard mixture described in Chapter 2.2.2, except that the Taq polymerase was substituted for Faststart Taq polymerase (Roche). The PCR conditions were: 95 ºC for 10 min; 40 cycles of: 95 ºC for 10 s, gradient temperature for 10 s, 72 ºC for 15 s; with a single final extension period of 72 ºC for 5 min. The PCR reactions were analysed for quality and quantity via electrophoresis (Chapter 2.3).

6.2.5 Multiplex Real-time PCR

The PCR cycling and solution conditions required further optimisation so that the two specific primer pairs, two probes, amplification control primers, and intercalating dye could be combined to form the *Trogoderma* multiplex test. Different combinations of primer pairs, probes and dye were used to determine if any of the components would cross react in the multiplex reaction. Quantitative PCR was performed on a Roche Lightcycler 480 (LC480) series II. The reaction volumes were 20 µl, including 2 µl of extracted DNA (1:10 dilution DNA:Water). The DNA tested

included the reference specimens, *T. granarium* and *T. variabile*. The PCR reaction mix comprising of 1 × Polymerase buffer (Roche), 0.4 µM of each primer (Table 6.1), ¼ volume of the suggested 1 × Resolight saturating dye (Roche), 1.8 mM $MgCl_2$, 200 µM of each dNTP, and 0.5 U of Faststart Taq polymerase (Roche). The LC480 conditions were: 95°C for 10 min; 40 cycles of: 95 ºC for 10 s, 55 ºC for 10 s, and 72 ºC for 15 s; a high resolution melt starting at 70 °C for 10s; a ramp rate of 0.01 °C/ sec (25 acquisitions/ °C) to 95 °C. A large range of melting temperatures was required to include both the species specific and amplification control products. The point of exponential increase of fluorescence above the background was recorded as a Cp value, which denotes Crossing Point and is equal to the Ct (Cycle Threshold) value. Fluorescence produced by the intercalating dye and the two hydrolysis probes was monitored in three channels, green (λ 465-510), orange (λ533-580), and red (λ618-660). The intercalating dye was used to detect the amplification of the species specific primers and the amplification control and subsequently MCA of the PCR products to support the probe result (Figure 6.1). The amplification control was detected with the intercalating dye and tests the DNA quality and the presence of PCR inhibitors. Melt curve analysis of the amplification control should provide a unique temperature profile for each species. A colour compensation file was generated to reduce the background fluorescence associated with the overlapping emissions of the two probes.

6.2.6 Detection and Specificity

Specificity of the *Trogoderma* multiplex test was determined on the two datasets. The validity of the multiplex reaction was determined in multiple ways: (i) successful amplification, (ii) detection of *T. variabile* probe, (iii) detection of *T. granarium* probe, (iv) MCA of the species specific primers to confirm the sequence that the probe anneals to, and (v) MCA of the amplification control. Three-fold serial dilutions of *T. granarium* and *T. variabile* DNA were utilised to create a concentration gradient that ranged from 375pg to 0.6pg. The sensitivity of the species specific primers and probes was also tested on a range of mixed templates. The mixed templates had different ratios (1:1, 1:2, 1:10, 1:100) of *T. variabile* or *T. granarium* DNA to *Orphinus* DNA.

6.2.7 Amplification and Sequencing of *COI*

Amplification of the 3` end of the *COI* was performed using PCR primer pairs C1-J-2183 and TL2-N-3014 (Chapter 2 Table 2.1) using the reaction volumes, PCR cycling conditions, DNA purification, electrophoresis and sequencing described in Chapter 2. Samples included the *T. variabile* and *T. granarium* reference samples and the *T. variabile* samples that generated false negatives results (Table 6.3).

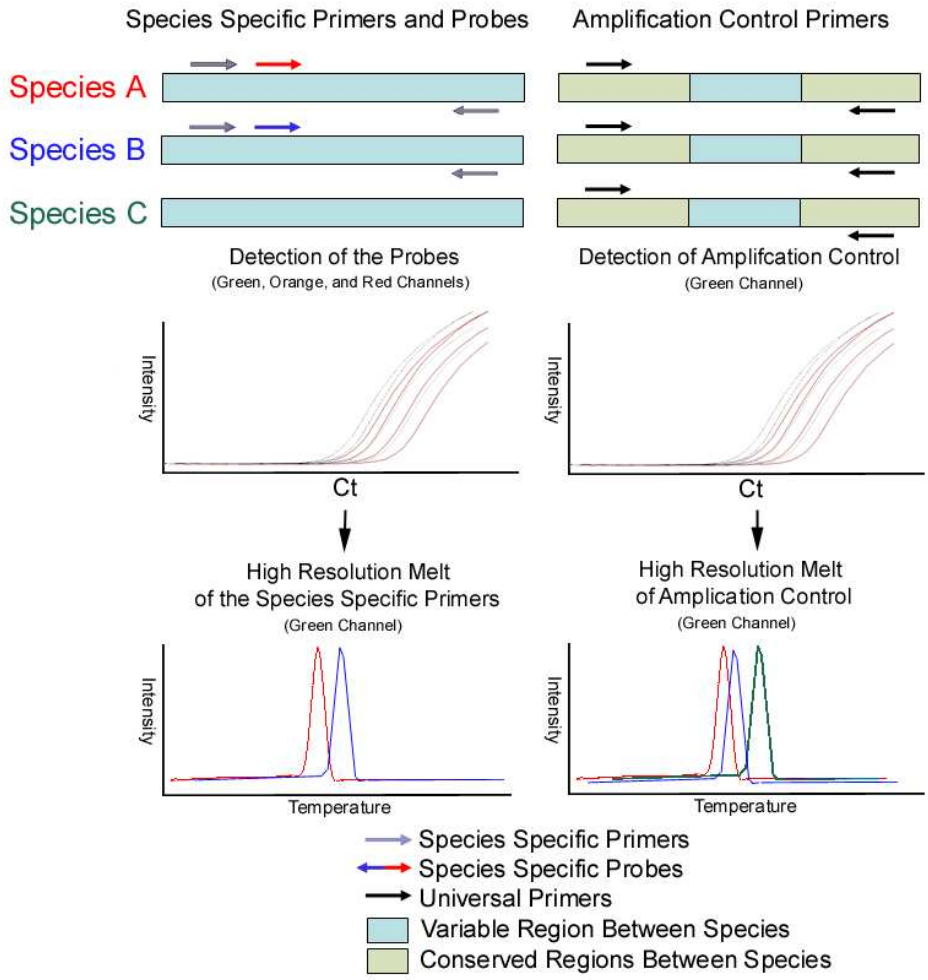

Figure 6.1. Details of the *Trogoderma* multiplex test. Species specific primers and probes were designed to anneal to *COI*, which is variable between species. Detection of the probes is dependent upon the emission spectrum of the probes, in this case the red and orange channels. Amplification of the specific primers is quantified using the intercalating dye. Inclusion of the intercalating dye allows for melt curve analysis to be performed on the species specific PCR product to confirm the target. The amplification control primers were designed to anneal to the conserved regions of *18S* that flank the V4 domain. Detection of the amplification control, using the intercalating dye, tests the DNA quality and the presence of PCR inhibitors. Melt curve analysis of the amplification control should provide a unique temperature profile for each species.

6.2.8 Data Analysis

The *COI* barcoding sequences were viewed and edited using CodonCode Aligner 3.0.3 (CodonCode Corporation) and aligned using the built-in version of CLUSTALW. Neighbour-joining (NJ) phylogenetic trees were generated using Mega 4 (Tamura *et al.*, 2007) under the following conditions: Maximum Composite Likelihood Model, d: Transitions + Transversions, Uniform rates, complete deletion for gaps, random seed and 10,000 bootstrap replicates.

6.3 Results

6.3.1 Multiplex Performance

No cross reaction between the *T. variabile* and *T. granarium* species specific primers and probes was observed, suggesting that they are highly specific. Inclusion of the intercalating dye had no effect on the cleavage, specificity or fluorescence of the hydrolysis probes. The fluorescence of each probe was observed in the expected channels and amplification of the species specific primers and amplification control were observed in the green channel (Figure 6.2). Multiplexing the specific primers and probes with the amplification control had no effect on the probes Cp value. As expected, MCA of the multiplex reaction yielded two products per species. The first product to melt was generated by the species specific primers. The PCR product amplified by the *T. variable* specific primers melted at 79.2 ºC (79.1 to 79.3 ºC) and the amplification control at 92.7 ºC (92.7 to 92.8 ºC; Figure 6.2A). In comparison, the *T. granarium* specific product melted at 81.3 ºC (81.2 to 81.5 ºC) and the amplification control at 92.1 ºC (92.2 to 92.5 ºC). A slight variation in melt peaks was detected between replicates of the same sample, but no overlap was observed between the two species. A smaller peak was observed at 84 ºC, half the intensity of the amplification control. Electrophoresis of the multiplex reaction was performed to determine if this third peak was the result of primer dimers forming. Interestingly, two PCR products were observed and no primer dimers. The detection limits for both probes was 0.6pg of the target DNA, both detected with a Cp value of 31 (Figure 6.2C & D).

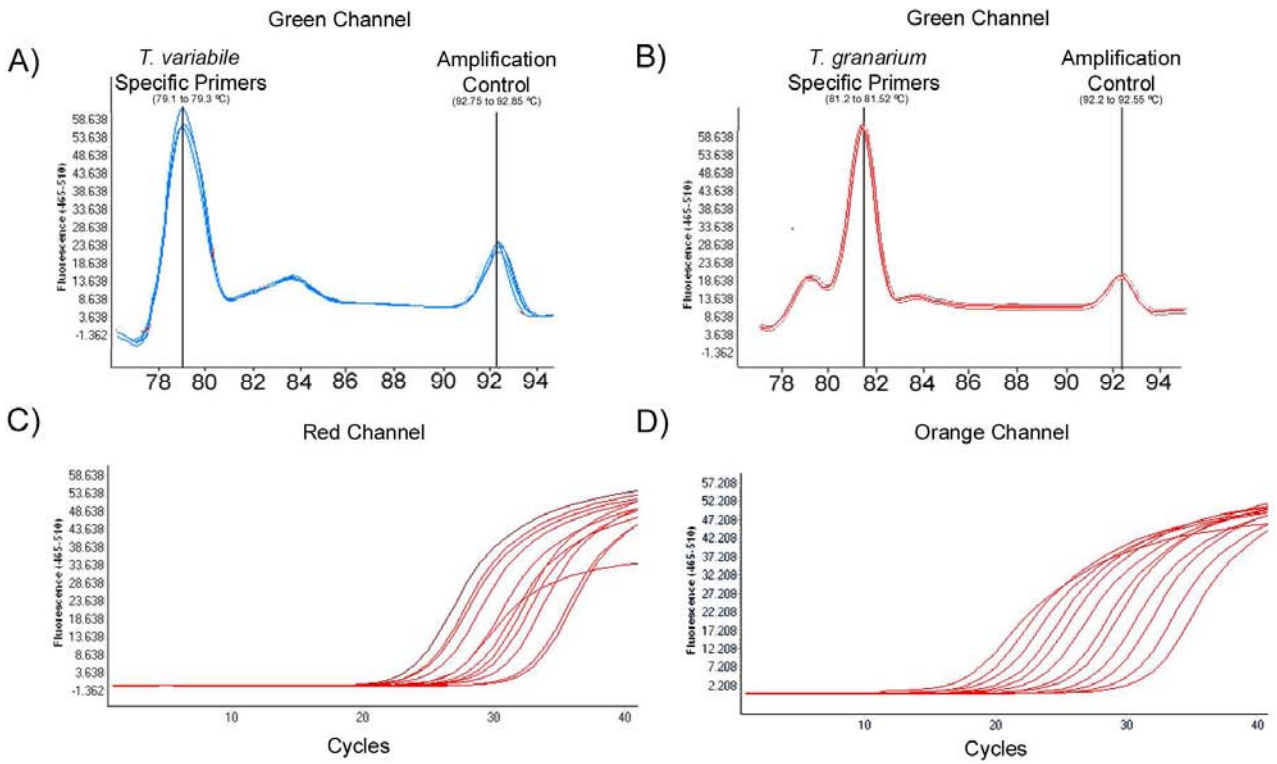

Figure 6.2. Performance of the *Trogoderma* multiplex test on *T. granarium* and *T. variabile*. A) Melt curve analysis of species specific primers and amplification primers for the *T. variabile* reference (green channel λ 465-510). B) Melt curve analysis of species specific primers and amplification primers for the *T. granarium* (green channel λ 465-510). C) Detection limits of the *T. variabile* specific probes with DNA concentrations ranging from 375pg to 0.6 pg with the probe detected in red channel (λ618-660). D) Detection limits of the *T. granarium* specific probes with DNA concentrations ranging from 375pg to 0.6 pg with the probe detected in orange channel (λ533-580).

6.3.2 Validation of the *Trogoderma* Multiplex Test on Dataset One

The positive controls that were randomly distributed throughout the plates were all successfully amplified (Cp values ranged from 18.2 to 22.1 and 22.7 to 25.8, *T. variabile* and *T. granarium* respectively; Table 6.2.). Each probe successfully bound to its target DNA and fluorescence was recorded in the specific channel. Amplification was observed in 87% of the samples tested. The Cp values recorded in the green channel ranged from 17.5 to 35. Twenty-five samples amplified after the 35^{th} cycle, within the last 5 cycles of the analysis. These samples were considered negative because they could represent cross contamination or amplification problems within that specific well.

Only the six *T. granarium* reference samples were detected in the orange channel, indicating the sensitivity and specificity for the *T. granarium* specific probes was 100%, irrespective of the inclusion of the amplification control (Table 6.3). In comparison, in the absence of an amplification control, the *T. variabile* specific test generated nine false negatives and five false positives results. The sensitivity and specificity was shown to be 93% and 85% respectively. When the amplification control was included, no sample returned a false negative result, increasing the sensitivity to 100% (Table 6.3). Melt curve analysis of the species specific amplification products was a highly sensitive approach at detecting both *T. variabile* and *T. granarium* (Table 6.3.). All the *T. granarium* samples melted between 81.1 ºC and 81.5 ºC (Table 6.3). In comparison, the *T. variabile* samples generated products with a melt temperature between 79.1 ºC and 79.3 ºC. While the specificity for detecting *T. granarium* was 100%, it was considerably lower (63%) for *T. variabile* (Table 6.3).

Table 6.2. Samples in Dataset one used to validate the *Trogoderma* multiplex test, number of samples tested (n), successful amplification of the internal control (%), number of samples identified as *T. granarium* using the specific primer probe (T.g Spec), number of samples identified as *T. variabile* using the specific primer probe (T.v Spec), MCA of the species-specific markers, and MCA of the amplification control, and Cp values.

Species	n	Amplification of Internal Control (%)	T.g Spec (n)	T.v Spec (n)	MCA Species Specific primers (C°; min to max)	MCA Amplification Control (C°; min to max)	Cp values (min to max)
B	5	100				92.57 (92.4 to 92.7)	32.1 to 35
C	27	85		4	79.2 (79.2 to 79.2)	92.52 (92.1 to 92.6)	28 to 35
D	14	79		1	79.2 (79.2 to 79.2)	92.31 (92.1 to 92.5)	24.8 to 30.96
E	5	80		1	79.2 (79.2 to 79.2)	92.65 (92.4 to 92.8)	28.3 to 35
F	10	90		1	79.1 (79.1 to 79.1)	92.52 (92.4 to 92.7)	26.2 to 30.12
G	4	100				92.63 (92.6 to 92.7)	23.9 to 30.55
T. variabile	49	84		42	79.18 (79.1 to 79.3)	92.49 (92.4 to 92.6)	20 to 35
T. granarium reference	8	100	8		81.3 (81.2 to 81.5)	92.32 (92.2 to 92.5)	24.3 to 26.21
T. variabile reference	10	100		10	79.22 (79.2 to 79.3)	92.63 (92.5 to 92.8)	18.2 to 22.95
	132	87	8	59			

Table 6.3. Sensitivity and Specificity for the Species Specific Probes and Melt Curve Analysis of the Species Specific Primers for Dataset One.

Species Specific Probe Results

		T. granarium Probe	No Detection				*T. variabile* Probe	No Detection		
Without Amplification Control	*T. granarium*	8	0	Spec	100%	*T. variabile*	50	9	Spec	93%
	Non *T. granarium*	0	124	Sens	100%	Non *T. variabile*	5	68	Sens	85%
With Amplification Control	*T. granarium*	8	0	Spec	100%	*T. variabile*	50	0	Spec	89%
	Non *T. granarium*	0	87	Sens	100%	Non *T. variabile*	Am5	40	Sens	100%

Melt Curve Analysis Results

		>81.2<81.6	<81.2				>79.1<79.3	>79.3		
Melt Curve Analysis of specific Primers	*T. granarium*	8	0	Spec	100%	*T. variabile*	42	0	Spec	63%
	Non *T. granarium*	0	59	Sens	100%	Non *T. variabile*	5	8	Sens	100%

Melt curve analysis of the amplification control generated highly variable temperature profiles for each of the putative species indentified in Chapter 3 and the reference samples (Figure 6.3). On average, each species had a melt curve temperature range of 0.4 ºC. Melt curve analysis of the *T. granarium* reference samples ranged from 92.2 ºC to 92.5 ºC and the *T. variabile* reference samples ranged from 92.5 ºC to 92.8 ºC. Each of the putative species as well as the *T. variabile* verification samples generated a melt curve temperature that overlapped with the reference samples. Melt curve analysis of the amplification control proved uninformative at delineating species, but reduced the number of false negatives associated with failed amplification.

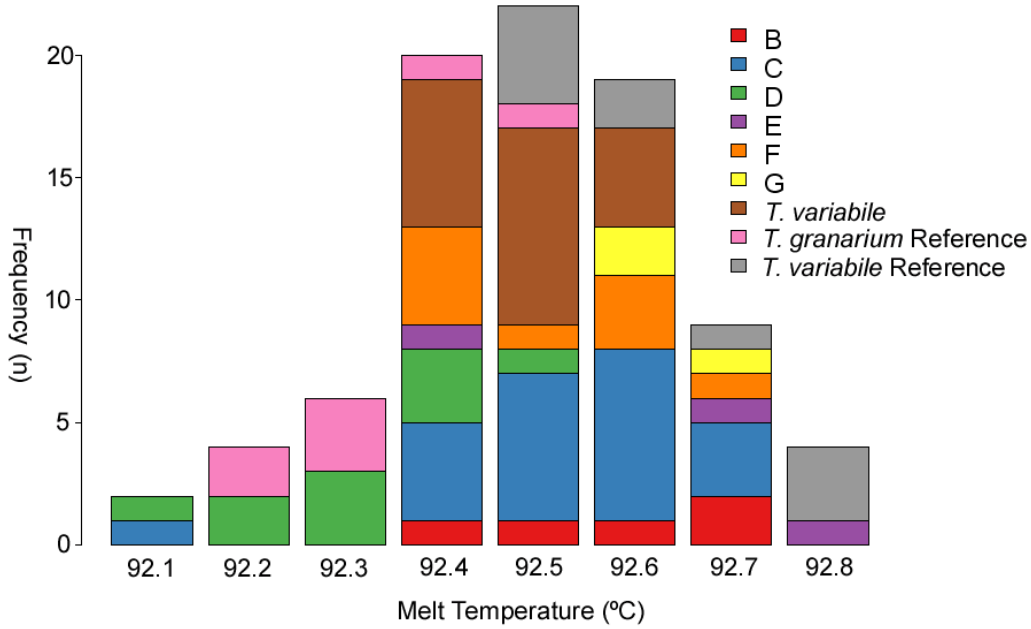

Figure 6.3. Temperature profiles for each species in dataset one, generated by MCA of the amplification control. The labels B to G represent the six putative species discovered in Chapter 3.

6.3.3 Validation of the *Trogoderma* Multiplex Test on Dataset Two

Forty specimens, made up of seven species of exotic *Trogoderma*, obtained either from interceptions or from the Berlin museum collections were analysed using the "test". The amplification control showed that 25% of the samples failed to amplify (Table 6.4). Inclusion of the amplification control reduced the number of *T.*

granarium false negative results from four to one (Table 6.5). Melt curve analysis of the remaining *T. granarium* specimen revealed that the *T. granarium* specific primers annealed, but the probe was not cleaved. The melt curve temperature of the PCR product was 81.6 ºC, well within the melting range of the other *T. granarium* included in this dataset, suggesting that this sample was indeed *T. granarium*.

Without the amplification control the *T. variabile* specific probes provided low sensitivity and high specificity, 60% and 92% respectively. Even with the addition of an amplification control, there were four false negative and two false positive results, increasing the sensitivity and decreasing the specificity (69% and 88% respectively). The two false positives were samples morphologically identified as *T. angustum* and *T. versicolor*. Melt curve analysis of these two samples produced a melt curve temperature >79.3 ºC, which was above the *T. variabile* melt temperature range (Table 6.3). While the probe annealed, the increased melt temperature indicated that the nucleotide composition of the amplified product was different to that of the *T. variabile* reference samples. This suggests that these specimens are not *T. variabile*. The four false negatives were not resolved by MCA, suggesting either misidentification of the specimens, mutations in either a primer or probe binding site or a high degree of variability within *T. variabile* that needs further investigation.

Table 6.4. Details of the samples used to test the *Trogoderma* multiplex test, number of samples tested (n), successful amplification of the internal control (%), number of samples identified as *T. variabile* using the specific primer probe, number of samples identified as *T. granarium* using the specific primer probe, and MCA of the species-specific markers.

Genus	Species	n	Amplification (%)	*T. granarium* Specific Probe (n)	*T. variabile* Specific Probe (n)	Melt Curve Analysis Specific Markers (C°; min to max)
Trogoderma	*angustum*	4	75		1	79.3
Trogoderma	*glabrum*	2	50			
Trogoderma	*granarium*	10	70	6		81.68 (81.6 to 81.7)
Trogoderma	*megatomoidies*	4	25			78.14 (78.1 to 78.1)
Trogoderma	*ornatum*	2	100			
Trogoderma	*variabile*	15	86		9	79.37 (79.1 to 79.8)
Trogoderma	*versicolor*	2	100		1	79.3
	Total	40	75	6	10	

Table 6.5. Sensitivity and Specificity for the Species Specific Probes and Melt Curve Analysis of the Species Specific Primers for Dataset Two.

Species Specific Probe Results

			T. granarium Probe	No Detection			*T. variabile* Probe	No Detection		
Without Control	Amplification	*T. granarium*	6	4	Spec	100%	*T. variabile*	9	6	Spec 92%
		Non *T. granarium*	0	30	Sens	60%	Non *T. variabile*	2	23	Sens 60%
With Control	Amplification	*T. granarium*	6	1	Spec	100%	*T. variabile*	9	4	Spec 88%
		Non *T. granarium*	0	23	Sens	86%	Non *T. variabile*	2	15	Sens 69%

Melt Curve Analysis Results

		>81.2<81.74	<81.2				>79.1<79.3	>79.3		
Melt Curve Analysis of specific Primers	*T. granarium*	7	0	Spec	100%	*T. variabile*	7	2	Spec	100%
	Non *T. granarium*	0	15	Sens	100%	Non *T. variabile*	0	12	Sens	78%

6.3.4 Barcoding Analysis of the *T. variabile* False Negatives

Phylogenetic analysis clustered all the morphologically identified *T. granarium* together (Figure 6.4). Two out of the four *T. variabile* false negatives that were detected in the MCA test (samples 10 & 11; Table 5.5.), clustered with the *T. variabile* reference samples. The other two false negatives (samples 5 & 6) that didn't amplify with the species specific primers nor were detected by the probes formed a novel clade, sister to *T. variabile*. However, no bootstrapping support was placed on this node. Pairwise divergence between the novel clade and the *T. variabile* and *T. granarium* reference samples was 17.83% and 19.61% respectively (Table 6.6), suggesting that these individuals may have been morphologically misidentified.

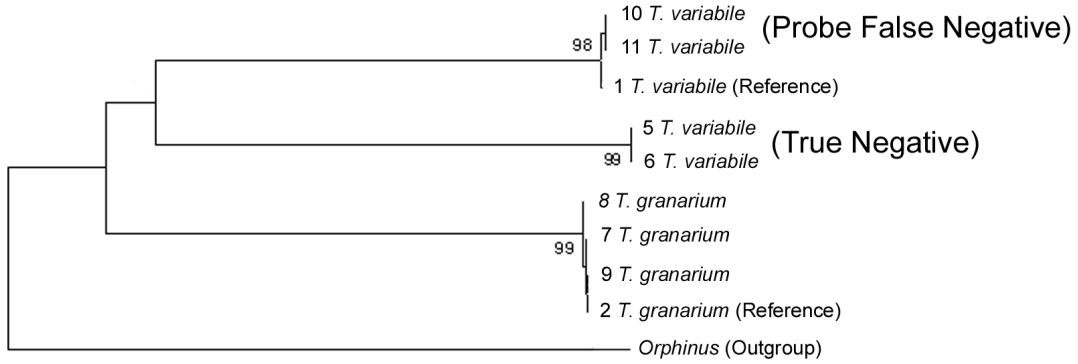

Figure 6.4. Phylogenetic Analysis of the *T. variabile* false negatives generated by the *Trogoderma* Multiplex Test.

Table 6.6 Pairwise divergence between the reference samples (*T. variabile* and *T. granarium*) and the quarantine intercepts.

		4	5	6	*Orph*	7	8	9	10	11
4	*T. variabile* Reference									
5	*T. variabile*	17.83								
6	*T. variabile*	17.83	0.00							
	Orphinus Outgroup	23.50	23.34	23.34						
7	*T. granarium*	18.64	19.45	19.45	23.18					
8	*T. granarium*	18.31	19.13	19.13	22.85	0.32				
9	*T. granarium*	18.48	19.45	19.45	23.18	0.16	0.49			
10	*T. variabile*	1.78	17.67	17.67	23.34	18.80	18.48	18.64		
11	*T. variabile*	1.78	17.67	17.67	23.34	18.80	18.48	18.64	0.00	
12	*T. granarium* Reference	18.64	19.61	19.61	23.34	0.32	0.65	0.16	18.80	18.80

6.4 Discussion

The goal of this study was to develop a high throughput diagnostic test and reduce the false negatives associated with PCR inhibition and poor DNA quality. Rather than using a plasmid or synthesised sequence (Nolte, 2004, Rosenstraus *et al.*, 1998), the V4 domain within the *18S* as an amplification control was examined. Including the amplification control increased the sensitivity of the multiplex test by 15% to 26%, dataset one and two respectively (Tables 6.3 & 6.5). This suggested that the false negatives were attributed to failed amplification and that the amplification control was a worthwhile addition to the multiplex test.

The species specific primers and probes were highly specific at detecting *T. granarium*, but less specific for *T. variabile* (Tables 6.3 & 6.5). The *T. variabile* specific primers and probes were too specific resulting in the test only identifying the Australian population and failing to identify samples 10 and 11. DNA barcoding of samples 10 and 11 showed that these samples were 1.78% divergent from the *T. variabile* reference sample and that the probes failed because due to the following: First, a single SNP in the binding site of the T. *variabile* probe; Second, two SNPs were observed in the *T. variabile* forward priming site, which may have reduced the likelihood of the primer binding to the target site, thus causing the amplification to fail.; Finally, a single SNP in the *T. granarium* forward priming position increased this primer's specificity, causing the forward *T. granarium* primer and the reverse *T. variabile* primer to amplify and generate a spurious melt profile. Without the intercalating dye and subsequent MCA, this error may have been missed. This highlights the problems associated with species specific assays, especially when only part of the targets population has been analysed. This problem was only exposed because two methods (morphology and DNA barcoding) were used to support the identification.

Melt curve analysis of the amplification control proved to be an unreliable means to differentiate species. While HRM analysis is designed to delineate between single nucleotide polymorphisms, it was found that the melt curve reproducibility within species and within replicates was low, with the melt curve temperature for most species varying up to 0.4°C (Figure 6.3). The lack of reproducibility was attributed to

the folding of the amplified *18S* fragment. Folding predictions of the amplicon using DINAMelt Server (http://dinamelt.bioinfo.rpi.edu), generated a range of secondary and tertiary structures, each with a unique melting temperature. This finding could explain why two amplification control peaks were observed during melt curve analysis as well as variation observed within a species, but only one visible band on an agarose gel. While the V4 domain performed well as an amplification control, it was unable to adequately meet the requirement as a positive control for the probes.

The false positives for the *T. variabile* probe (Tables 6.2 & 6.3) that were observed during the validation of dataset one were probably the result of historical cross contamination. These samples were stored in glass vials in 70% ethanol for a period of 7 to 8 years, and each vial contained a range of different species. Shokralla *et al.* (2010) performed a series of experiments on the storage of specimens in ethanol and found that a specimens' *COI* could be amplified and sequenced from the ethanol, even after 7 years of storage. Therefore, it is probable that the false positives are the result of *T. variabile* cross contaminating the non-target samples.

The false positives observed during the validation of dataset two were the amplification of *T. angustum* and *T. versicolour* by the *T. variabile* specific probes. These false positives were attributed to the cross reaction of primers on closely related species. These samples had highly degraded DNA and were unable to be genotyped with the *COI* primers. Therefore, phylogenetic comparisons between these species and the reference samples were not possible. Sequencing of the *T. angustum* and *T. versicolour* amplified products would not have been particularly informative, as the sequences would have contained the species specific primers and not their true sequence. Moreover, the species specific product size was only 200 bp (of which 20% would be primer), which could make it difficult to determine whether differences between the new sequences and the references are intra or inter-specific variation. This shows the importance of sequencing closely related species before designing species specific markers.

DNA barcoding proved to be a vital part of delineating species when dealing with samples 5 and 6 (Table 6.6; Figure 6.4). These two samples were morphologically identified as *T. variabile*, yet the diagnostic probes and MCA of the species specific

primers failed to support the morphological classification. Phylogenetic analysis of these samples showed that they formed a novel clade that was 17.83% and 19.61% divergent from the *T. variabile* and *T. granarium* reference samples respectively (Table 6.6). Using barcoding as a way of identification for quarantine material can be undesirable. First, DNA barcoding analysis can take several days if the samples need to be processed offsite. Second, if the intra-specific boundaries are unknown it can be difficult to ascertain whether a novel clade is truly a new species. For example, intra-specific divergence has been shown to range from 0.67% to 10% (Foottit *et al.*, 2008, Cognato and Sun, 2007). Third, a barcoding approach requires an extensive database with corresponding supportive morphological taxonomy (Armstrong, 2010). Nevertheless, in this case it has proved to be a powerful additional tool when species specific tests and morphology are incongruent.

In this chapter, the multiplexing of species specific primers and probes with an amplification control, and including an intercalating dye to perform MCA on the species specific amplified products generated a useful methodology for reducing the incidence of false positives and false negatives. The inclusion of the intercalating dye was the most important addition in the test because it allowed for MCA of the species specific product. In all cases where false positives were generated, MCA of the species specific product allowed the result to be securitised and eventually excluded based on their melt curve profile.

Chapter 7

General Discussion

This thesis investigated the phylogeny of *Trogoderma* and closely related species to support the hypothesis "Molecular markers can differentiate *Trogoderma granarium* from *Trogoderma variabile* and other closely related species". Accurate morphological identification of the main *Trogoderma* pest species (*T. variable* and *T. granarium*) is possible because each species has morphological features that are diagnostically unique and are alone sufficient to identify adults of the species. However, when dealing with damaged specimens, immature stages, or the existence of cryptic species, accurate identification using morphological characteristics is difficult (Walter, 2003, Schutze *et al.*, 2006, Hebert *et al.*, 2003) and often these are the types of specimens that are intercepted by quarantine agencies.

To overcome the difficulties associated with morphological identification, studies investigating invasive insect pests have turned to molecular phylogenetics (e.g. (DeSalle, 2006, Foottit *et al.*, 2008, Gwiazdowski *et al.*, 2006, Loxdale and Lushai, 1998, Carew *et al.*, 2009). The use of a phylogenetic approach to aid with the identification of a species should ideally satisfy the following five criteria:

1. For an invasive species that has successfully colonised (invaded) a new geographical location, investigate the population structure because significant differences between individuals may indicate multiple incursions, points of entry, or the presence of cryptic species (Loxdale and Lushai, 1998, Mikac and Clarke, 2006).
2. Compare the new population to other known populations and home biogeographical range. This may allow determination of the point of origin of the invasive species in its new range (e.g. Tsutsui *et al.*, 2001). For quarantine services, such a diagnostic capability has advantages in tracing intercepted organisms to their source.
3. Combine phenotypic and genotypic data to determine the species boundaries and reduce the incidence of false positives and false negatives.

4. Investigate the phylogenetic relationship between the closely related native species and the exotic pest species. This is particularly pertinent as the Australian native *Trogoderma* are frequently collected around grain storage structures and likely to be intercepted in a shipment of grain.
5. Use multiple genes from different parts of the genome.

This thesis shows that failure to apply these five points to a research project would in some cases lead to misidentification.

The *T. variabile* adult samples that were collected during the CSIRO national trapping program from 2001 to 2003 were to be used to examine the species boundaries and model the dispersal mechanisms of *T. granarium.* Rather than a single population of *T. variabile,* the study revealed eight deeply diverged lineages, of which only 53% of the specimens grouped with the *T. variabile* reference sample. When compared to the *T. variabile* reference sample, the pairwise divergence ranged from 18.88 to 25.68%, the lowest of these values was a comparison with the *T. granarium* reference sample (Chapter 3). Unfortunately a destructive DNA extraction method was used to extract DNA, whereby entire beetles were macerated to liberate the DNA, making it impossible to review the morphological characters. It is clear that the original morphological identification was not rigorous because of the time constraints in screening many larvae samples.

While these clades are likely to represent Australian native dermestids, without morphological confirmation they must remain as putative species. Despite the definition of "a molecular species" being individuals that cluster together by either lineage or through a character based analysis (Balakrishnan, 2005, Sites and Marshall, 2003), to suggest that six clades represented six species, without the support of the morphological data, could lead to incorrect inference. Studies using *COI* have shown that a species boundary can vary dramatically depending on the organism and dispersal range (Cognato, 2006, Cognato and Sun, 2007, deWaard *et al.,* 2009, Foottit *et al.,* 2008, Meyer and Paulay, 2005). In particular Foottit *et al.* (2008) showed that the divergence between two species of aphid was as little as 0.61%, whereas Cognato and Sun (2007) reported a species with an intra-specific variation of 10%. Both studies based these estimates of the species boundary on

morphological and molecular data; therefore, it is risky to suggest that each clade identified in this research represents a species. Furthermore, without combining the *T. variabile* phenotypic data with the pairwise divergence it is difficult to determine if the upper boundary for the *T. variabile* species was 2.9%. This value is above the mean intra-specific divergences calculated by deWaard *et al.* (2009) and Foottit *et al.* (2008). Thus it is impossible to draw any strongly supported conclusion about the number of incursion events. This outcome highlights the importance of non-destructive sampling to (i) recheck the specimen's morphological classification when intra-specific divergence exceeds the previously known boundary, and (ii) to recheck when molecular and morphological diagnoses are incongruent.

While several non-destructive sampling methods for DNA extraction from insects have been published (Gilbert *et al.*, 2007, Favret, 2005, Hunter *et al.*, 2008, Rowley *et al.*, 2007, Pons, 2006) these methods didn't satisfy the needs of this project. These methods require the use of toxic or corrosive chemicals (e.g. phenol, chloroform, and guanidine isothiocyanate), are time-consuming through the need for overnight incubation, or risk the loss of DNA through ethanol or isopropanol precipitation. Rather this project required a non-destructive approach that was rapid, inexpensive, non-toxic and capable of high throughput to deal with the numerous sample numbers (>100 per trap) common in trapping programs. EDNA Hi SpEx, an alkaline hydrolysis based method satisfied the desired criteria by allowing the extraction of amplifiable DNA in 20 min without distorting or discolouring the cuticle (Chapter 4). The simplicity of this method allowed it to be incorporated into the first step of any diagnostic protocol that requires both morphological and molecular identification. With this type of methods future *Trogoderma* trapping programs may be able to screen numerous samples and match each of the clades indentified in Chapter 3 to a physical specimen.

.

The aim of the phylogenetic study, particularly the *COI* data, was to generate a reference database that could serve as a frame of reference for the identification of quarantine samples. The value in using such an approach has been shown by Armstrong and Ball (2005), Armstrong (2010), and Meyer and Paulay (2005). Furthermore, the use of a DNA database can be >96% accurate when the genus or

family has been thoroughly genotyped (Meyer and Paulay, 2005). The phylogenetic studies showed that *T. granarium* and *T. variabile* were sister taxa that were separated by deep branch lengths and >18% pairwise divergence. Likewise, the Australian *Trogoderma* and closely related genera were easily delineated from the two exotic *Trogoderma* species by >19% pairwise divergence and long branch lengths (Chapter 5). Despite these findings the *Trogoderma* genus was paraphyletic, the trees were only supported by Bayesian analysis, Parsimony analysis suffered from polytomy, and the putative species identified in Chapter 3 were not rediscovered using this method. Without a well constructed database, defining species' boundaries can be difficult (Armstrong, 2010) and can result in false positives and false negatives due to an individual being grouped with another species, or split from another species resulting in an apparent novel taxon (Meyer and Paulay, 2005). This was evident when a specimen intercepted by AQIS was morphologically identified as *T. variabile*, yet, its pairwise divergence was outside *T. variable's* boundary and thus formed a novel clade (Chapter 6).

The problems with large samples size, likely to be collected from a trapping program, highlighted the need for a PCR-based test. Despite the fact that species specific markers can be prone to false positives and false negatives (Peist *et al.*, 2001, Nolte, 2004, Armstrong, 2010), it was believed that this problem could be overcome by multiplexing hydrolysis probes, amplification controls, and including a melt curve analysis step (Chapter 6). The multiplex test that was developed in this project was highly successful. Seventy six percent of the false negatives were attributed to DNA quality or PCR inhibition. Despite *T. angustum* and *T. versicolor* testing positive as *T. variabile* via the species specific probes, the melt curve analysis resulted in these two samples falling outside the melting range of *T. variabile*. The only *T. granarium* sample that failed the species specific probe test was confirmed as *T. granarium* via melt curve analysis. Including the hydrolysis probes, amplification controls, and including melt curve analysis provided an internal checking system that reduced the incidence of false positives and negatives. However, melt curve analysis of the amplification control proved to be an unreliable means to differentiate between species. The amplification control was designed to provide a melt temperature that was conserved within a species but variable between species, thus highlighting a "failed" test due to a new intra-species variant. Future work should investigate

alternative genes to the *18S* that have a highly variable region flanked by highly conserved regions that can be used as PCR priming sites. However, the likelihood of finding such a region is low without an extensive phylogenetic study using multiple gene regions.

The hypothesis that "Molecular markers can differentiate *T. granarium* from *T. variabile* and other closely related species" has been supported by the research reported herein and when markers are used correctly they can be a valuable tool to confirm morphological identifications. Conversely, the results also show that to correctly identify *Trogoderma* species using exclusively molecular methods, especially a single gene, would prove unreliable, especially if the *Trogoderma* genus is the result of rapid radiation (Chapter 5). Ancient polymorphisms could be retained in all descendent lineages and by chance each species may share a set of alleles at different loci (McCracken and Sorenson, 2005), misleading the identification of a specimen when using a phylogenetic tree approach. Likewise, to rely solely on morphological features to identify a species is also dangerous as: morphological characters vary continuously (Lee, 2004), morphological characteristics can be retained throughout history (parallelism) or evolve separately (convergent evolution; Wiens *et al.*, 2003) giving the impression that morphologically similar species are more closely related than they are genetically, and when larvae or damaged specimens are missing diagnostically significant characteristics. To accurately indentify a *Trogoderma* species requires both approaches to provide different parts of the puzzle to ensure accurate identification of pest species.

7.1 Future Directions

Without the inclusion of representative individuals from a range of species and different populations within a species it is difficult to ascertain where the species boundary lies. Accurate identification of the species boundary will in turn increase the accuracy of a DNA-based barcoding or species specific test. Future studies will need to examine *Trogoderma* species from the Palaearctic (*Trogoderma versicolour* Normand, *Trogoderma megatomoides* Reitter) and American native species (*Trogoderma angustum* Solier, *Trogoderma anthernoids* Blackburn, *T. glabrum, Trogoderma ornatum* LeConte, and *Trogoderma simplex* Jayne). In addition,

Attagenus, *Anthrenocerus*, *Anthrenus* and *Orphinus* specimens will need to be included. Including these additional species may help resolve the polytomy observed in the phylogenetic trees. Furthermore, some of these samples are pests of stored products and are likely to be intercepted by quarantine agencies in Australia and elsewhere. Only once these species have been phylogenetically and morphologically analysed can all *Trogoderma* pest species be identified with confidence.

References

ALVAREZ, A. & WENDEL, J. F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29**,** 417-434.

ALVAREZ, W., KAUFFMAN, E. G., SURLYK, F., ALVAREZ, L. W., ASARO, F. & MICHEL, H. V. 1984. Impact Theory of Mass Extinctions and the Invertebrate Fossil Record. *Science,* 223**,** 1135-1141.

ARMSTRONG, K. 2010. DNA barcoding: a new module in New Zealand's plant biosecurity diagnostic toolbox. *EPPO Bulletin,* 40**,** 91-100.

ARMSTRONG, K. F. & BALL, S. L. 2005. DNA barcodes for biosecurity: invasive species identification *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 1813–1823.

AVERS, C. J. 1989. *Process and pattern in evolution,* New York Oxford University Press.

BALAKRISHNAN, R. 2005. Species Concepts, Species Boundaries and Species Identification: A View from the Tropics. *Systematic Biology,* 54**,** 689-693.

BANKS, H. J. 1977. Distribution and Establishment of Trogoderma granarium Everts (Coleoptera: Dermestidae): Climatic and other influences. *Journal Stored Products Research,* 13**,** 183-202.

BEAL, R. S., JR. & ZHANTIEV, R. D. 2001. A New California Species of Egidyella Reitter (Coleoptera: Dermestidae), a Genus Previously Unknown in the New World. *The Coleopterists Bulletin,* 55**,** 70-74.

BEHURA, S. K. 2006. Molecular marker systems in insects: current trends and future avenues. *Molecular Ecology Notes,* 15**,** 3087–3113.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2001. *Biochemistry,* New York, W. H. Freeman and Company.

BOORE, J. 1999. Animal mitochondrial genomes. *Nucl. Acids Res.,* 27**,** 1767-1780.

BOOTH, R. G., COX, M. L. & MADGE, R. B. 1990. *IIE Guides to Insects of Importance to Man: No. 3: Coleoptera,* Wallingford, CAB International.

BOSSUYT, F. & MILINKOVITCH, M. C. 2000. Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 6585-6590.

BOWER, J. E., COOPER, R. D. & BEEBE, N. W. 2009. Internal Repetition and Intraindividual Variation in the rDNA ITS1 of the Anopheles punctulatus Group (Diptera: Culicidae): Multiple Units and Rates of Turnover. *Journal Molecular Evolution,* 68**,** 66-79.

BRENNER, G. J. 1996. *Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel,* New York, Chapman and Hall.

BROWER, A. V. Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 6491-6495.

BROWN, B. V. 1993. A further chemical alternative to critical- point-drying for preparing small (or large) flies. *Fly Times,* 11.

BROWN, W. M., MATTHEW, G. & WILSON, A. C. 1979. Rapid evolution of animal mitochondrial DNA. *Genetics,* 76**,** 1967-1971.

BUCKLEY, T. R., ATTANAYAKE, D. & BRADLER, S. 2009. Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. *Proceedings of the Royal Society B: Biological Sciences,* 276**,** 1055-1062.

BURGES, H. D. 1962. Studies on the dermestid beetle Trogoderma grunarium Everts V-Reactions of diapause larvae to temperature. *Bulletin of Entomological Research,* 53**,** 193-213.

CACCONE, A., GARCIA, B. A. & POWELL, J. R. 1996. Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. *Insect Molecular Biology,* 5**,** 51-59.

CAMPBELL, J. F. & MULLEN, M. A. 2004. Distribution and Dispersal Behavior of Trogoderma variabile and Plodia interpunctella Outside a Food Processing Plant. *Journal of Economic Entomology,* 97**,** 1455-1464.

CANTAREL, B. L., MORRISON, H. G. & PEARSON, W. 2006. Exploring the Relationship between Sequence Similarity and Accurate Phylogenetic Trees. *Molecular Biology and Evolution,* 23**,** 2090-2100.

CAREW, M., SCHIFFER, M., UMINA, P. A., WEEKS, A. & HOFFMANN, A. 2009. Molecular markers indicate that the wheat curl mite Aceria tosichella Keifer may represent a species complex in Australia. *Bulletin of Entomological Research*.

CASTRESANA, J. 2001. Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals. *Molecular Biology and Evolution,* 18**,** 465-471.

CATERINO, M. S., CHO, S. & SPERLING, F. A. H. 2000. The Current State Of Insect Molecular Systematics: A Thriving Tower of Babel. *Annual Review of Entomology,* 45**,** 1-54.

CHANG, G. J., TRENT, D. W., VORNDAM, A. V., VERGNE, E., KINNEY, R. M. & MITCHELLAND, C. J. 1994. An Integrated Target Sequence and Signal Amplification Assay, Reverse Transcriptase-PCR-Enzyme-Linked Immunosorbent Assay, To Detect and Characterize Flaviviruses. *Journal of Clinical Microbiology,* 32**,** 477-483.

CHIARI, Y., VENCES, M., VIEITES, D. R., RABEMANANJARA, F., BORA, P., RAMILIJAONA RAVOAHANGIMALALA, O. & MEYER, A. 2004. New evidence for parallel evolution of colour patterns in Malagasy poison frogs (Mantella). *Molecular Ecology,* 13**,** 3763-3774.

CLARY, D. O. & WOLSTENHOLME, D. R. 1985. The Mitochondrial DNA Molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code *Journal Molecular Evolution,* 22**,** 252-271.

CLEMENT, M., POSADA, D. & CRANDALL, K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology Notes,* 9**,** 1657-1660.

COCK, M. J. W. 2003. Biosecurity and Forests: An Introduction - with particular emphasis on forest pests. FAO Forest Health and Biosecurity. *In:* DEPARTMENT, F. (ed.). Rome: Working Paper FBS/2E.

COGNATO, A. I. 2006. Standard Percent DNA Sequence Difference for Insects Does Not Predict Species Boundaries. *Journal of Economic Entomology,* 99**,** 1037-1045.

COGNATO, A. I. & SUN, J. H. 2007. DNA based cladograms augment the discovery of a new Ips species from China (Coleoptera: Curculionidae: Scolytinae). *Cladistics,* 23**,** 539-551.

COOK, D. 2003. Prioritising Exotic Pest Threats to Western Australia Plant Industries. *In:* AUSTRALIA, D. O. A. W. (ed.). Department of Agriculture Western Australia

CRANDALL, K. A. & TEMPLETON, A. R. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics,* 134**,** 959-969.

CROWSON, R. A. 1981. *The Biology of the Coleoptera,* London, Academic Press.

CRUZ, D. D. O., JORGE, D. M. D. M., PEREIRA, J. O. P., TORRES, D. C., SOARES, C. E. A., FREITAS, B. M. & GRANGEIRO, T. B. 2006. Intraspecific variation in the first internal transcribed spacer (ITS1) of the nuclear ribosomal DNA in Melipona subnitida (Hymenoptera, Apidae), an endemic stingless bee from northeastern Brazil. *Apidologie,* 37**,** 376-386.

DANFORTH, B. N. & JI, S. 1998. Elongation Factor-1a Occurs as Two Copies in Bees: Implications for Phylogenetic Analysis of EF-1a Sequences in Insects. *Molecular Biology and Evolution,* 15**,** 225-235.

DEGNAN, P. H., LAZARUS, A. B. & WERNEGREEN, J. J. 2005. Genome sequence of Blochmannia pennsylvanicus indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Research,* 15**,** 1023-1033.

DELIGEORGIEVA, T. G., KALOYANOVAA, S. & VAQUEROB, J. J. 2009. Intercalating Cyanine Dyes for Nucleic Acid Detection. *Recent Patents on Materials Science,* 2**,** 1-26.

DESALLE, R. 2006. Species Discovery versus Species Identification in DNA Barcoding Efforts: Response to Rubinoff. *Conservation Biology,* 20**,** 1545-1547.

DESALLE, R. 2007. Phenetic and DNA taxonomy; a comment on Waugh. *BioEssays,* 29**,** 1289-1290.

DEWAARD, J. R., LANDRY, J.-F., SCHMIDT, B. C., DERHOUSOFF, J., MCLEAN, J. A. & HUMBLE, L. M. 2009. In the dark in a large urban park: DNA barcodes illuminate cryptic and introduced moth species. *Biodiversity and Conservation,* 18**,** 3825-3839.

DRUMMOND, A. J. & RAMBAUT, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* 7.

EICKBUSH, T. H. & EICKBUSH, D. G. 2007. Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics,* 175**,** 477-485

FAVRET, C. 2005. A New Non-Destructive DNA Extraction and Specimen Clearing Technique for Aphids (Hemiptera). *Proc. Enotomol. Soc. Wash.,* 107**,** 469-470.

FLOYD, R. M., WILSON, J. J. & HEBERT, P. D. N. 2009. DNA barcodes and insect biodiversity. *In:* FOOTTIT, R. G. & ADLER, P. H. (eds.) *Insect Biodiversity. Science and Society.* Wiley-Blackwell.

FOOTTIT, R. G., MAW, H. E. L., VON DOHLEN, C. D. & HERBERT, P. D. N. 2008. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Molecular Ecology Resources,* 8**,** 1189-1201.

FRITZ, G. N., CONN, J., COCKBURN, A. & SEAWRIGHT, J. 1994. Sequence Analysis of the Ribosomal DNA Internal Transcribed Spacer 2 from Populations of Anopheles nuneztovari (Diptera: Culicidae). *Molecular Biology and Evolution,* 11**,** 406-416.

FUNK, D. 1999. Molecular systematics of cytochrome oxidase I and 16S from Neochlamisus leaf beetles and the importance of sampling. *Molecular Biology and Evolution,* 16**,** 67-82.

FUNK, D. J., FUTUYMA, D. J., ORTI, G. & MEYER, A. 1995. Mitochondrial DNA sequences and multiple data sets: a phylogenetic study of phytophagous beetles (Chrysomelidae: Ophraella). *Mol Biol Evol,* 12**,** 627-640.

FUNK, D. J. & OMLAND, K. E. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics,* 34**,** 397-423.

FUTUYMA, D. J. 1998. *Evolutionary Biology,* Sunderland, Sinauer Associates.

GHEERBRANT, E. 2009. Paleocene emergence of elephant relatives and the rapid radiation of African ungulates. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 10717-10721.

GILBERT, M. T. P., MOORE, W., MELCHIOR, L. & WOROBEY, M. 2007. DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage. *PloS One,* 2**,** e272.

GILLESPIE, J. J., TALLAMY, D. W., RILEY, E. G. & COGNATO, A. I. 2008. Molecular phylogeny of rootworms and related galerucine beetles (Coleoptera: Chrysomelidae). *Zoologica Scripta,* 37**,** 195-222.

GROTH, D. M. & WETHERALL, J. D. 2000. *"Molecular " Tools in Epidemiological Investigations,* London, Arnold Press.

GWIAZDOWSKI, R. A., VAN DRIESCHE, R. G., DESNOYERS, A., LYON, S., WU, S.-A., KAMATA, N. & NORMARK, B. B. 2006. Possible geographic origin of beech scale, Cryptococcus fagisuga (Hemiptera: Eriococcidae), an invasive pest in North America. *Biological Control* 39**,** 9-18.

HALLIDAY, R. B. & KNIHINICKI, D. K. 2004. The occurrence of Aceria tulipae (Keifer) and. Aceria tosichella (Keifer) in Australia (Acari: Eriophyidae). *International Journal of Acarology,* 30**,** 113-118.

HANSEN, T. A. 1988. Early Tertiary Radiation of Marine Molluscs and the Long-Term Effects of the Cretaceous-Tertiary Extinction. *Paleobiology,* 14**,** 37-51.

HARRIS, D. J. & CRANDALL, K. A. 2000. Intragenomic Variation Within ITS1 and ITS2 of Freshwater Crayfishes (Decapoda: Cambaridae): Implications for Phylogenetic and Microsatellite Studies. *Molecular Biology and Evolution,* 17**,** 28-291.

HARTLEY, D. J. & GREENING, H. G. 1983. Review of the warehouse beetle control programme. *In:* WALES, D. O. A. N. S. (ed.). Sydney.

HARTWELL, L., HOOD, L., GOLDBERG, M. L., REYNOLDS, A. E., SILVER, L. M., RUTH., V. & HARTWELL, H., GOLDBERG, REYNOLDS, SILVER, VERES,, , PAGE 688, 2004. *Genetics - from genes to genomes,* Boston, McGraw Hill Higher Education.

HAVA, J. 2003. *World Catalogue of the Dermestidae (Coleoptera),* Prague, Brandýse nad Labem.

HAY, J. M., SUBRAMANIAN, S., MILLAR, C. D., MOHANDESAN, E. & LAMBERT, D. M. 2008. Rapid molecular evolution in a living fossil. *Trends in Genetics,* 24**,** 106-109.

HEBERT, P. D. N., CYWINSKA, A., BALL, S. L. & DEWAARD, J. R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences,* 270**,** 313-321.

HINTON, H. E. 1945. *A Monograph of the Beetles Associated with Stored Products,* London, British Museum.

HOELZEL, A. R., FLEISCHER, R. C., CAMPAGNA, C., LE BOEUF, B. J. & ALVORD, G. 2002. Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology,* 15**,** 567-575.

HUMPHRIES, E. M. & WINKER, K. 2010. Working through polytomies: Auklets revisited. *MOLECULAR PHYLOGENETICS AND EVOLUTION,* 54**,** 88-96.

HUNT, T., BERGSTEN, J., LEVKANICOVA, Z., PAPADOPOULOU, A., JOHN, O. S., WILD, R., HAMMOND, P. M., AHRENS, D., BALKE, M., CATERINO, M. S., GÓMEZ-ZURITA, J., RIBERA, I., BARRACLOUGH, T. G., BOCAKOVA, M., BOCAK, L. & P., V. A. 2007. A

Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science,* 318**,** 1913-1916.
HUNTER, S. J., GOODALL, T. I., WALSH, K. A., OWEN, R. & DAY, J. C. 2008. Nondestructive DNA extraction from blackflies (Diptera: Simuliidae): retaining voucher specimens for DNA barcoding projects *Molecular Ecology Notes,* 8**,** 56-61.
HWANG, U.-W. & KIM, W. 1999. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *The Korean Journal of Parasitology,* 37**,** 215-228.
HWANG, U.-W., REE, H. I. & WON, K. 2000. Evolution of Hypervariable Regions, V4 and V7, of Insect 18S rRNA and Their Phylogenetic Implications. *Zoological Science,* 17**,** 111-121.
JENKINS, T. M., JONES, S. C., LEE, C.-Y., FORSCHLER, B. T., CHEN, Z., LOPEZ-MARTINEZ, G., GALLAGHER, N. T., BROWN, G., NEAL, M., THISTLETON, B. & KLEINSCHMIDT, S. 2007. Phylogeography illuminates maternal origins of exotic Coptotermes gestroi (Isoptera: Rhinotermitidae). *Molecular Phylogenetics and Evolution,* 42**,** 612-621.
KALINOWSKI, S. T., MUHLFELD, C. C., GUY, C. S. & COX, B. 2010. Founding population size of an aquatic invasive species. *Conservation Genetics,* 11**,** 2049-2053.
KHAN, H. A., ARIF, I. A., BAHKALI, A. H., FARHAN, A. H. A. & HOMEIDAN, A. A. A. 2008. Bayesian, Maximum Parsimony and UPGMA Models for Inferring the Phylogenies of Antelopes Using Mitochondrial Markers. *Evolutionary Bioinformatics,* 2008**,** 263.
KINGMAN, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications,* 13**,** 235-248.
KISELYOVA, T. & MCHUGH, J. V. 2006. A phylogenetic study of Dermestidae (Coleoptera) based on larval morphology. *Systematic Entomology,* 31**,** 469-507.
KLOOTWIJK, C. T., GEE, J. S., PEIRCE, J. W., SMITH, G. M. & MCFADDEN, P. L. 1992. An early India-Asia contact: Paleomagnetic constraints from Ninetyeast Ridge, ODP Leg 121. *Geology,* 20**,** 395-398.
KUROSE, N. M., R. AND TATARA M. 2005. Fecal DNA Analysis for Identifying Species and Sex of Sympatric Carnivores: A Noninvasive Method for Conservation on the Tsushima Islands, Japan. *Journal of Heredity,* 96**,** 688-697.
LABANDEIRA, C. C., JOHNSON, K. R. & WILF, P. 2002. Impact of the terminal Cretaceous event on plant–insect associations. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 2061-2066.
LAMBERT, D. M. 1983. A Population Genetical Study of the African Mosquito Anopheles marshallii (Theobald). *Evolution,* 37**,** 484-495.
LAMBERT, D. M., RITCHIE, P. A., MILLAR, C. D., HOLLAND, B., DRUMMOND, A. J. & BARONI, C. 2002. Rates of Evolution in Ancient DNA from Adelie Penguins. *Science,* 275**,** 2270-2273.
LEE, H.-C., YANG, M.-M. & YEH, W.-B. 2008. Identification of Two Invasive Cacopsylla chinensis (Hemiptera: Psyllidae) Lineages Based on Two Mitochondrial Sequences and Restriction Fragment Length Polymorphism of Cytochrome Oxidase I Amplicon. *Journal Economic Entomology,* 101.
LEE, M. S. Y. 2004. The molecularisation of taxonomy. *Invertebrate Systematics,* 18**,** 1-6.
LI, W., XI, B., YANG, W., HAWKINS, M. & SCHUBART, U. K. 2003. Complex DNA melting profiles of small PCR products revealed using SYBR® Green I. *BioTechniques,* 35**,** 702-706.
LI, Z. X. & POWELL, C. M. 2001. An Outline of the palaeogeographical evolution of the Australian region since the beginning of the Neopreterozoic. *Earth-Science Reviews,* 53**,** 237-277.
LORENZ-LEMKWE, A. P., MUSCHNER, V. R. C., BONATTO, S. L., CERVI, A. C., SALZANO, F. M. & FREITAS, L. B. 2005. Phylogeographic Inferences Concerning Evolution of Brazilian Passiflora actinia and P. elegans (Passifloraceae) Based on ITS (nrDNA) Variation. *Annals of Botany,* 95**,** 799-806.
LOWE, S., BROWNE, M., BOUDJELAS, S. & DEPOORTER, M. 2000. *100 of the World's Worst Invasive Alien Species: A Selection from the Global Invasive Species Database,* Auckland, Invasive Species Specialist Group (ISSG).
LOXDALE, H. D. & LUSHAI, G. 1998. Molecular markers in entomology. *Bulletin of Entomological Research,* 88**,** 577-600.
LUNT , D. H., ZHANG , D. X., SZYMURA , J. M. & HEWITT , G. M. 1996. The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology,* 5**,** 153-165.
LYON, E. & WITTWER, C. T. 2009. LightCycler Technology in Molecular Diagnostics. *Journal of Molecular Diagnostics,* 11**,** 93-11.

MACK, R. N., SIMBERLOFF, D., LONSDALE, W. M., EVANS, H., CLOUT, M. & A., B. F. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications,* 10**,** 689-710.

MADDISON, W. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics,* 5**,** 365-377.

MARINI, M. & MANTOVANI, B. 2002. Molecular Relationships among European Samples of Reticulitermes (Isoptera, Rhinotermitidae). *Molecular Phylogenetics and Evolution,* 22**,** 454-459.

MCCRACKEN, K. G. & SORENSON, M. D. 2005. Is Homoplasy or Lineage Sorting the Source of Incongruent mtDNA and Nuclear Gene Trees in the Stiff-Tailed Ducks (Nomonyx-Oxyura)? *Systematic Biology,* 54**,** 35-55.

MEYER, C. P. & PAULAY, G. 2005. DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol,* 3**,** e422.

MICHEL, H., BEHR, J., HARRENGA, A. & KANNT, A. 1998. CYTOCHROME C OXIDASE: Structure and Spectroscopy. *Annual Review of Biophysics and Biomolecular Structure,* 27**,** 329-356.

MIKAC, K. M. & CLARKE, G. C. 2006. Tracing the geographic origin of the cosmopolitan parthenogenetic insect pest Liposcelis bostrychophila (Psocoptera: Liposcelididae). *Bulletin of Entomological Research,* 96**,** 1-8.

MIKAC, K. M. & FITZSIMMONS, N. F. 2010. Genetic structure and dispersal patterns of the invasive psocid Liposcelis decolor (Pearman) in Australian grain storage systems. *Bulletin of Entomological Research,* in press.

MILLER, D. R. & ROSSMAN, A. Y. 1995. Systematics, Biodiversity, and Agriculture. *BioScience,* 45**,** 680-686.

MOREAU, C. S., BELL, C. D., VILA, R., ARCHIBALD, S. B. & E., P. N. 2006. Phylogeny of the Ants: Diversification in the Age of Angiosperms. *Science,* 312**,** 101 - 104.

MORITZ, C. & CICERO, C. 2004. DNA Barcoding: Promise and Pitfalls. *PLoS Biology,* 2**,** 1529-1531.

MROCZKOWSKI, M. 1968. Distribution of the Dermestidae (Coleoptera) of the world with a catalogue of all known species. *Annales Zoologici.* Warszawa.

MUN, J., BOHONAK, A. J. & RODERICK, G. K. 2003. Population structure of the pumpkin fruit fly *Bactrocera depressa* (Tephritidae) in Korea and Japan: Pliocene allopatry or recent invasion? *Molecular Ecology,* 12**,** 2941-2951.

MURAJI, M. & NAKAHARA, S. 2001. Phylogenetic relationships among fruit flies, *Bactrocera* (Diptera, Tephritidae), based on the mitochondrial rDNA sequences. *Insect Molecular Biology,* 10**,** 549-559.

NADEL, R. L., SLIPPERS, B., SCHOLES, M. C., LAWSON, S. A., NOACK, A. E., WILCKEN, C. F., BOUVET, J. P. & WINGFIELD, M. J. 2009. DNA bar-coding reveals source and patterns of Thaumastocoris peregrinus invasions in South Africa and South America. *Biological Invasions,* 12**,** 1067-1077.

NOLTE, F. S. 2004. Novel Internal Controls For Real-Time PCR Assays. *Clin Chem,* 50**,** 801-802.

O'BRIEN, S. J., WILDT, D. E., BUSH, M., CARO, T. M., FITZGIBBON, C., AGGUNDEY, I. & E., L. R. 1987. East African cheetahs: evidence for two population bottlenecks? *Proceedings of the National Academy of Science USA,* 84**,** 508-511.

O'HUIGIN, C., SATTA, Y., TAKAHATA, N. & KLEIN, J. 2002. Contribution of Homoplasy and of Ancestral Polymorphism to the Evolution of Genes in Anthropoid Primates. *Molecular Biology and Evolution,* 19**,** 1501-1513.

PAGE, R. D. M. & HOLMES, E. C. 1998. *Molecular Evolution A Phylogenetic Approach,* Victora, Blackwell Publishing.

PAMILO, P. & NEI, M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution,* 5**,** 568-583.

PEACOCK, E. R. 1993. *Adults and Larvae of Hide, Larder and Carpet Beetles and their Relatives (Coleoptera: Dermestidae) and of Derodontid Beetles (Coleoptera: Derodontidae),* London, The Natural History Museum.

PEIST, R., HONSEL, D., TWIELING, G. & LÖFFERT, D. 2001. PCR inhibitors in plant DNA preparations. *QIAGEN*

PIMENTEL, D., ZUNIGA, R. & MORRISON, D. 2004. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics***,** 1-16.

PIPERNO, D. R., RANERE, A. J., HOLST, I., IRIARTE, J. & DICKAU, R. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 5019-5024.

PIPERNO, D. R., WEISS, E., HOLST, I. & NADEL, D. 2004. Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature,* 430**,** 670-673.

PONS, J. 2006. DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers *Molecular Ecology Notes,* 6**,** 623-626.

POSADA, D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution,* 25**,** 1253-1256.

POSADA, D. & CRANDALL, K. A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics,* 14**,** 817-818.

PRESA, P., PARDO, B. G., MARTı´NEZ, P. & BERNATCHEZ, L. 2002. Phylogeographic Congruence Between mtDNA and rDNA ITS Markers in Brown Trout. *Molecular Biology and Evolution,* 19**,** 2161-2175.

PROVINE, W. B. 2004. Ernst Mayr: Genetics and speciation. *Genetics,* 167**,** 1041-6.

PUILLANDRE, N., DUPAS, S., DANGLES, O., ZEDDAM, J. L., CAPDEVIELLE-DULAC, C., BARBIN, K., TORRES-LEGUIZAMON, M. & SILVAIN, J. F. 2008. Genetic bottleneck in invasive species: the potato tuber moth adds to the list. *Biological Invasions,* 10**,** 319-333.

QIAGEN 2006. DNeasy® Blood & Tissue Handbook. *In:* QIAGEN (ed.). Qiagen.

RASMUSSEN, M. D. & KELLIS, M. 2007. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. *Genome Research,* 17**,** 1932-1942.

REES, D. P., STARICK , N. & WRIGHT, E. J. 2003. Current status of the warehouse beetle Trogoderma variabile (Coleoptera: Dermestidae) as a pest of grain storage in Australia. *Stored grain in Australia*.

REY, I., DORDA, B. A. & VALDECASAS, A. G. 2004. Traditional water mite fixatives and their compatibility with later DNA studies. *Experimental and Applied Acarology,* 34**,** 59-65.

ROKAS, A. & CARROLL, S. B. 2005. More Genes or More Taxa? The Relative Contribution of Gene Number and Taxon Number to Phylogenetic Accuracy. *Molecular Biology and Evolution,* 22**,** 1337-1344.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

ROS, V. I. D. & BREEUWER, J. A. J. 2007. Spider mite (Acari: Tetranychidae) mitochondrial COI phylogeny reviewed: host plant relationships, phylogeography, reproductive parasites and barcoding. *Exp Appl Acarol,* 42**,** 239-262.

ROSENSTRAUS, M., WANG, Z., CHANG, S.-Y., DEBONVILLE, D. & SPADORO, J. P. 1998. An Internal Control for Routine Diagnostic PCR: Design, Properties, and Effect on Clinical Performance. *J. Clin. Microbiol.,* 36**,** 191-197.

ROWLEY, D. L., CODDINGTON, J. A., GATES, M. W., NORRBOM, A. L., OCHOA, R. A., VANDENBERG, N. J. & GREENSTONE, M. H. 2007. Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes,* 7**,** 915-924.

RUGMAN-JONES, P. F., HODDLE, M. S. & 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 Economic Entomology,* 103**,** 877-886.

SANCHIS, A., MICHELENA, J. M., LATORRE, A., QUICKE, D. L. J., GARDENFORS, U. & BELSHAW, R. 2001. The Phylogenetic Analysis of Variable-Length Sequence Data: Elongation Factor-1 alpha Introns in European Populations of the Parasitoid Wasp Genus Pauesia (Hymenoptera: Braconidae: Aphidiinae). *Molecular Biology and Evolution,* 18**,** 1117-1131.

SANDERSON, M. J. & DOYLE, J. A. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany,* 88**,** 1499-1516.

SCHEFFER, S. J., LEWIS, M. L. & JOSHI, R. C. 2006. DNA Barcoding Applied to Invasive Leafminers (Diptera: Agromyzidae) in the Philippines. *Annals of the Entomological Society of America,* 99**,** 204-210.

SCHUTZE, M. K., MATHER, P. B. & CLARKE, A. R. 2006. Species status and population structure of the Australian Eucalyptus pest Paropsis atomaria Olivier (Coleoptera: Chrysomelidae). *Agricultural and Forest Entomology,* 8**,** 323-332.

SHERMAN, C. W. & LOEB, L. A. 1977. Depurination decreases fidelity of DNA synthesis in vitro. *Nature,* 270**,** 537-538.

SHOKRALLA, S., SINGER, G. A. C. & HAJIBABAEI, M. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *BioTechniques,* 48**,** 305-306.

SHOUP, S. & LEWIS, L. A. 2003. Polyphyletic Origin of Parallel Basl Bodies in Swimming Cells of Chlorophycean Green Algae (CHLOROPHYTA). *Journal of Phycology,* 39**,** 789-796.

SIMON, C., FRATI, F., BECKENBACH, A., CRESPI, B., LIU, H. & FLOOK, P. 1994. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Annals of the Entomological Society of America,* 87**,** 651-701.

SITES, J. W. J. & MARSHALL, J. C. 2003. Delimiting species: a Renaissance issue in systematic biology. *TRENDS in Ecology and Evolution,* 18**,** 462-470.

SMITH, P. T., KAMBHAMPATI, S. & ARMSTRONG, K. A. 2003. Phylogenetic relationships among *Bactrocera* species (Diptera: Tephritidae) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 26**,** 8-17.

SUAREZ, A. V., HOLWAY, D. A. & TSUTSUI, N. D. 2008. Genetics and Behavior of a Colonizing Species: The Invasive Argentine Ant. *The american naturalis,* 172**,** 72-84.

SWOFFORD, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 4 ed. Sunderland, Massachusetts: Sinauer Associates.

SZITO, A. 2007a. Diagnostic Protocols for Regulated Pests Trogoderma granarium. *In:* FOOD, D. O. A. A. (ed.).

SZITO, A. 2007b. Global Invasive Species Database. Invasive Species Specialist Group.

TAMURA, K., DUDLEY, J., NEI, M. & KUMAR, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution,* 24**,** 1596-1599.

TANG, J., TOE, L., BACK, C. & UNNASCH, T. R. 1996. Intra-specific Heterogeneity of the rDNA Internal Transcribed Spacer in the Simulium damnosum (Diptera: Simuliidae) Complex. *Molecular Biology and Evolution,* 13**,** 244-252.

TAUTZ, D., ARCTANDER, P., MINELLI, A., THOMAS, R. H. & VOLGER, A. P. 2003. A plea for DNA taxonomy. *Trends in Ecology and Evolution,* 18**,** 70-74.

TAYLOR, M., MCKECHNIE, S., PIERCE, N. & KREITMAN, M. 1993. The lepidopteran mitochondrial control region: structure and evolution. *Molecular Biology and Evolution,* 10**,** 1259-1272.

TEAM, R. D. C. 2007. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22**,** 4673-4680.

TOJO, K. & MATSUKAWA, K. 2003. A Description of the Second species of the Family Dipteromimidae (Insecta, Ephemeroptera), and Genetic Relationship of Two Dipteromimid Mayflies Inferred from Mitochondrial 16S rRNA Gene Sequences. *Zoological Science,* 20**,** 1249-1259.

TSUTSUI, N. D., SUAREZ, A. V. & GROSBERG, R. K. 2003. Genetic diversity, asymmetrical aggression, and recognition in a widespread invasive species. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 1078-1083.

TSUTSUI, N. D., SUAREZ, A. V., HOLWAY, D. A. & CASE, T. J. 2000. Reduced Genetic Variation and the Success of an Invasive Species. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 5948-5953.

TSUTSUI, N. D., SUAREZ, A. V., HOLWAY, D. A. & CASE, T. J. 2001. Relationship among native and introduced populations of the Argentine ant (Linepithema humile) and the source of introduced populations. *Molecular Ecology,* 10**,** 2151-2161.

VIGNAL, A., MILAN, D., SANCRISTOBAL, M. & EGGEN, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution,* 34**,** 305-275.

VOGLER, A., WELSH, A. & HANCOCK, J. 1997. Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). *Molecular Biology and Evolution,* 14**,** 6-19.

VOGLER, A. P. & DESALLE, R. 1994. Evolution and Phylogenetic Information Content of the ITS-l Region in the Tiger Beetle Cicindela dorsalis. *Molecular Biology Evolution* 11**,** 393-405.
VORIS, H. K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography,* 27**,** 1153-1167.
WALSH, P. S., METZGER, D. A. & HIGUCHI, R. 1991. Chelex-100 as a medium for simple extraction of DNA for PCR based typing from forensic material. *BioTechniques,* 10**,** 506-513.
WALTER, G. H. 2003. *Insect Pest Management and Ecological Research,* New York, Cambridge University Press.
WANG, Y., YU, H., RAPHAEL, K. & GILCHRIST, A. S. 2003. Genetic delineation of sibling species of the pest fruit fly Bactocera (Diptera: Tephritidae) using microsatellites. *Bulletin of Entomological Research,* 93**,** 351-360.
WEBER, D. S., STEWART, B. S., GARZA, J. C. & LEHMAN, N. 2000. An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Current Biology,* 10**,** 1287-1290.
WEIDER, L. J., ELSER, J. J., CREASE, T. J., MATEOS, M., COTNER, J. B. & MARKOW, T. A. 2005. THE FUNCTIONAL SIGNIFICANCE OF RIBOSOMAL (r)DNA VARIATION: Impacts on the Evolutionary Ecology of Organisms. *Annual Review of Ecology, Evolution, and Systematics,* 36**,** 219-242.
WHEELER, Q. D. 2008. *The New Taxonomy,* Boca Raton, CRC Press.
WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In:* INNIS, M. A., GELFAND, D. H., SNINSKY, J. J. & WHITE, T. J. (eds.) *PCR Protocols: A Guide to Methods and Applications.* San Diego: Academic Press, Inc.
WHITMOYER, R. E., NAULT, L. R. & BRADFUTE, O. E. 1972. Fine Structure of Aceria tulipae (Acarina: Eriophyidae). *Annals of the Entomological Society of America,* 65**,** 201-215.
WIENS, J. J., CHIPPINDALE, P. T. & HILLIS, D. M. 2003. When Are Phylogenetic Analyses Misled by Convergence? A Case Study in Texas Cave Salamanders. *Systematic Biology,* 52**,** 501-514.
WILLIAM, C. B. I. & PIESMAN, J. 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 10034-10038.
WITTWER, C. T., HERRMANN, M. G., MOSS, A. A. & RASMUSSEN, R. P. 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques,* 22.
WRIGHT, E. J. 1994. Warehouse beetle, final report. *In:* CORPORATION, R. I. R. A. D. (ed.). CSIRO Entomology.
YANG, P. & ANDRÉ, C. 2007. Efficient PCR from whole blood using Finnzymes High
Performance PCR solution. *Nature Methods Application Notes*.
ZHANG, D.-X. & HEWITT, G. M. 1997. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology,* 25**,** 99-120.
ZHANTIEV, R. D. 2009. Ecology and Classification of Dermestid Beetles (Coleoptera,Dermestidae) of the Palaearctic Fauna. *Entomological Review,* 89**,** 157-174.

Appendix I

DNA Sequence Information

ID	Ch	Genbank Accession Numbers		
		COI	*CYT b*	*18S*
A.verbasci	3	HM243383	HM243269	HM243239
AriahP2	3		HM243270	
AriahP3	3	HM243384		
AriahP4	3	HM243385	HM243271	
AriahP5	3		HM243272	
AriahP6	3	HM243386	HM243273	
Balak1	3		HM243274	
Balak2	3	HM243387		HM243240
Balak3	3		HM243275	
Balak4	3		HM243276	HM243241
Balak5	3		HM243277	HM243242
Balak6	3		HM243278	HM243243
Booroowa1	3	HM243388	HM243279	HM243244
Booroowa4	3	HM243389		
Caracabal2	3	HM243390	HM243280	
Caracabal4	3		HM243281	
Caracabal5	3		HM243282	
Caracabal6	3		HM243283	
Caracabal8	3		HM243284	
Carnanmah1	3	HM243391	HM243285	
Carnanmah2	3	HM243392	HM243286	
Carnanmah3	3	HM243393	HM243287	
CB1	3	HM243394		HM243245
CB3	3	HM243395	HM243288	HM243246
CB4	3	HM243396	HM243289	HM243247
CB6	3		HM243290	HM243248
Colleam1	3	HM243397	HM243291	HM243249
Colleam3	3	HM243398	HM243292	HM243250
Colleam6	3	HM243399	HM243293	
Colleam8	3	HM243400		
Coolaman1	3		HM243294	
Coolaman2	3	HM243401		
Coolaman3	3	HM243402		
Coolaman4	3		HM243295	
Coolaman5	3		HM243296	
Coolaman6	3		HM243297	
Coolaman7	3		HM243298	
Coolaman8	3		HM243299	
Cowra2	3		HM243300	
Cowra3	3		HM243301	
Cowra4	3		HM243302	
Cowra5	3		HM243303	
Cowra7	3	HM243403		
Cowra8	3	HM243404	HM243304	

ID	Ch	Genbank Accession Numbers		
		COI	*CYT b*	*18S*
Dimboola1	3	HM243405		
Dimboola2	3	HM243406	HM243305	
Dimboola3	3	HM243407	HM243306	
Dimboola4	3		HM243307	
Dimboola5	3	HM243408	HM243308	
Dimboola6	3	HM243409	HM243309	
Dimboola7	3	HM243410	HM243310	
Dimboola8	3	HM243411	HM243311	
Garrah1	3	HM243412		
Garrah2	3	HM243413	HM243312	
Geelong1	3	HM243414	HM243313	
Geelong2	3	HM243415	HM243314	HM243251
Geelong3	3	HM243416	HM243315	
Geraldton1	3	HM243417	HM243316	
Griffith10	3	HM243418	HM243317	HM243252
Griffith11	3		HM243318	
Griffith2	3		HM243319	
Griffith4	3		HM243320	
Griffith3	3	HM243419		
Griffith6	3		HM243321	
Griffith9	3	HM243420	HM243322	HM243253
Katanning1	3	HM243421	HM243323	
LDNP1	3	HM243422	HM243324	
LDNP2	3		HM243325	
LDNP3	3	HM243423		
LDNP4	3		HM243326	
LDNP5	3	HM243424	HM243327	HM243254
LDNP6	3		HM243328	
LDNP7	3	HM243425	HM243329	
LDNP8	3	HM243426	HM243330	
Lockhart1	3		HM243331	
Lockhart4	3		HM243332	
Lockhart6	3		HM243333	
Lockhart3	3	HM243427		
Lockhart5	3	HM243428		
Lockhart7	3	HM243429		
McClevie1	3	HM243430	HM243334	
Mtong1	3		HM243335	
Mtong2	3	HM243431		
Mtong3	3		HM243336	
Mtong4	3		HM243337	
Mtong5	3		HM243338	
Mtong6	3		HM243339	
Mtong7	3	HM243432	HM243340	
Mtong8	3		HM243341	
NarrabinW1	3	HM243433	HM243342	
NarrabinW2	3	HM243434	HM243343	
NarrabinW3	3	HM243435	HM243344	
Nhill1	3	HM243436		
Nhill2	3	HM243437	HM243345	
Nhill6	3	HM243438		
Nhill7	3	HM243439		

ID	Ch	Genbank Accession Numbers		
		COI	*CYT b*	*18S*
Nhill8	3	HM243440		
PA10	3	HM243441		
PA11	3		HM243346	
PA12	3		HM243347	HM243255
PA14	3	HM243442	HM243348	HM243256
PA3	3	HM243443	HM243349	HM243257
PA5	3	HM243444		
PA6	3	HM243445		
PA8	3	HM243446	HM243350	
PA9	3		HM243351	
Pithara1	3	HM243447	HM243352	
Quirindi1	3	HM243448	HM243353	
Rainbow8	3	HM243449		
Rock3	3		HM243354	
Rock4	3		HM243355	HM243258
Rock6	3		HM243356	
Rock7	3		HM243357	
SwanH1	3		HM243358	
SwanH2	3		HM243359	
SwanH3	3	HM243450		
SwanH4	3		HM243360	
SwanH5	3	HM243451		
SwanH7	3	HM243452		HM243259
SwanH8	3	HM243453	HM243361	
SwanH9	3		HM243362	
T.granarium	3	HM243454	HM243363	HM243260
Tamworth2	3	HM243455	HM243364	
Tamworth3	3	HM243456		HM243261
Tamworth4	3		HM243365	
Tamworth6	3	HM243457		
Tamworth7	3	HM243458	HM243366	
Tamworth8	3		HM243367	
Three Springs1	3	HM243459		HM243262
Varley1	3		HM243368	HM243263
Varley2	3	HM243460	HM243369	
Varley3	3	HM243461	HM243370	
Varley5	3		HM243371	
Varley6	3	HM243462	HM243372	HM243264
Wagin1	3	HM243463	HM243373	
WestWylong1	3	HM243464	HM243374	HM243265
WestWylong2	3	HM243465	HM243375	HM243266
WestWylong3	3		HM243376	
WestWylong4	3	HM243466	HM243377	HM243267
WestWylong5	3		HM243378	HM243268
WestWylong6	3	HM243467	HM243379	
WestWylong8	3	HM243468	HM243380	
Willowtree1	3		HM243381	
Willowtree2	3	HM243469		
Wyalkatchem1	3	HM243470	HM243382	
0	5	HQ419076	HQ419112	HQ419149
71	5		HQ419141	HQ419171
73	5	HQ419106	HQ419143	HQ419172

ID	Ch	Genbank Accession Numbers		
		COI	***CYT b***	***18S***
80	5	HQ419107		HQ419173
96	5	HQ419108	HQ419145	HQ419174
97	5	HQ419109	HQ419146	HQ419175
98	5	HQ419110	HQ419147	
105	5	HQ419081		
106	5	HQ419082		
107	5	HQ419083	HQ419115	HQ419152
115	5		HQ419117	
116	5	HQ419085	HQ419118	HQ419153
119	5	HQ419086	HQ419119	HQ419154
120	5	HQ419087	HQ419120	HQ419155
121	5	HQ419088		HQ419156
124	5	HQ419090	HQ419123	HQ419157
130	5		HQ419124	HQ419158
132	5	HQ419091	HQ419125	HQ419159
134	5	HQ419092	HQ419126	HQ419160
135	5		HQ419127	
137	5		HQ419128	HQ419161
138	5	HQ419093	HQ419129	HQ419162
142	5	HQ419094	HQ419130	HQ419163
146	5	HQ419095	HQ419131	HQ419164
147	5	HQ419096		
153	5		HQ419132	HQ419165
157	5	HQ419098	HQ419133	HQ419166
160	5	HQ419099	HQ419134	
165	5		HQ419135	
168	5		HQ419136	HQ419167
170	5	HQ419100	HQ419137	HQ419168
176	5	HQ419101		
179	5	HQ419102	HQ419138	
184	5	HQ419103	HQ419139	HQ419169
4 *T. variabile* Reference	6	HQ419111		
5 *T. variabile*	6	HQ419079		
6 *T. variabile*	6	HQ419080		
Orphinus Outgroup	6	Same as ID 119 from Chapter 5		
7 *T. granarium*	6	HQ419089		
8 *T. granarium*	6	HQ419090		
9 *T. granarium*	6	HQ419097		
10 *T. variabile*	6	HQ419103		
11 *T. variabile*	6	HQ419104		
12 T. granarium Reference	6	Same as *T. granarium* from Chapter 3		

Appendix II

Multiple incursions and putative species revealed using a mitochondrial and nuclear phylogenetic approach to the *Trogoderma variabile* (Coleoptera: Dermestidae) trapping program in Australia

Appendix III

A rapid non-destructive DNA extraction method for insects and other arthropods

Appendix IV

Molecular phylogeny supports the paraphyletic nature of the genus *Trogoderma* (Coleoptera: Dermestidae) collected in the Australasian ecozone