DNA-based identification of forensically significant beetles from southern Africa

A thesis submitted in fulfilment of
the requirements for the degree of

Master of Science

in Zoology

at

Rhodes University

Isabel Judith Collett

Abstract

Necrophilous insects, if correctly identified, can provide useful forensic information. Research in this area has focussed on flies and beetles remain comparatively under-studied, partly because some adult carrion beetles are difficult to identify morphologically, as are their juvenile stages, often requiring specialist expertise in both cases. Molecular taxonomy has been proposed as a solution to these problems. DNA “barcodes" are short fragments of mitochondrial cytochrome oxidase I (COI) DNA that are anticipated to delineate species. This approach is becoming increasingly popular, but has been met with varying enthusiasm from taxonomists. This thesis examines their use in identifying forensically significant beetles.

The DNA barcodes of 234 specimens of 25 forensically significant southern African beetle species from seven families (Cleridae, Dermestidae, Silphidae, Staphylinidae, Scarabaeidae, Trogidae and Histeridae) were obtained. Thirty-three initial barcode amplification failures were overcome by using primers other than the standard Folmer pair, undermining the barcode concept’s hope of universal primers that would allow even non-specialists to produce barcodes. Another 150 specimens (64%) entirely failed to yield barcodes, including 18 fresh specimens of three species of Trogidae, implying another lack of universality of the barcoding protocol. The majority of the beetles clustered with confamilials on neighbour-joining and maximum likelihood trees, but 1.3% of the barcodes failed to cluster with their respective families, raising questions concerning the associating power of barcodes. The identification tools of the GenBank and BOLD on-line DNA sequence databases identified 21% of the specimens to the species level, 6% of them correctly. There was evidence of a paralogous sequence in the Cleridae that, while supporting identification now that it has been associated with a morphological identification, would hamper attempts at identification by clustering or phylogenetic analysis.

Distance and haplotype network analyses of the barcodes of six widespread species showed that they are not geographically structured. Barcodes are thus unlikely to be indicators of the region of origin of a species and will not determine whether a corpse has been relocated after death.

To assess whether a different mitochondrial DNA fragment might address (some of) these problems, a 2.2 kb fragment extending from the 5’ end of the COI gene to the 3’ end of the Cytochrome Oxidase II (COII) gene was analysed for nine species. It was found that, for Dermestidae, Scarabaeidae and Histeridae, higher degrees of diversity occurred downstream

of the barcode region, but the region of highest diversity in the Cleridae was in the barcode region. Thus, finding a more reliable fragment along the COI-COII region for each family may make robust and guaranteed DNA-based identification of these beetles more likely.

The possibility of a forensic specimen being incorrectly or not identified based on its barcode alone exists in about 40% of cases, even with the new barcodes reported here. Forensic science sets a very high bar in assessing the performance of its techniques, and it is concluded that barcodes currently have unsettling failure rates as court-worthy evidence.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFLP	Amplified Fragment Length Polymorphism
BLAST	Basic Local Alignment Research Tool
BOLD	Barcode of Life Data systems
bp	base pairs
CDS	Coding Sequence
COI	Cytochrome Oxidase I
COII	Cytochrome Oxidase II
CytB	Cytochrome oxidase B
dNTP	Deoxynucleoside Triphosphate
DNA	Deoxyribonucleic Acid
EF1α	Elongation Factor-1 Alpha
ITS	Internal Transcribed Spacers
ML	Maximum Likelihood
mtDNA	Mitochondrial DNA
nBLAST	Nucleotide Basic Local Alignment Research Tool
NCBI	National Centre for Biotechnology Information
nDNA	Nuclear DNA
NJ	Neighbour Joining
NUMTs	Nuclear Copies of Mitochondrial Genes
PCR	Polymerase Chain Reaction
PepCK	Phosphoenolpyruvate Carboxykinase
PMI	Post Mortem Interval
RAPD	Random Amplification Polymorphic DNA
RFLP	Restriction-Fragment-Length Polymorphisms
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
SNP	Single-nucleotide Polymorphism
Taq	*Thermus aquaticus* DNA (polymerase)
t-RNA	Transfer Ribonucleic Acid
UPGMA	Unweighted Pair Group Methods with Arithmetic Mean
UV	Ultraviolet

Preface

My thanks go to my supervisors, Martin Villet and Alicia Timm for being enthusiastic, selfless and always available for assistance when I needed them. I am enormously appreciative of all your help.

Many thanks to Jennifer and Giles Gush of Amakhala Game Reserve for allowing the collection of insects from carcasses on the reserve; Nicolette Bouwer, Alienor Brassine, Emma Bruce-Miller, John Midgley, Ryan Wasserman and Milena Wolmarans for donating specimens; and everyone else that assisted me with collecting specimens.

I thank Slavomir Mazur for identification of the Histeridae and Alfred Newton for identifying the Staphylinidae. I also thank John Midgley for his generous input and for identifying specimens. I am extremely grateful for all the consistent effort you have made on my behalf.

A special thanks to Karim Dhanani for all his assistance and support in the final stages of writing. Your consistent motivation and encouragement are greatly appreciated.

Deep thanks to my parents, for all the inspiration, patience and unconditional support that I have received from them during my tertiary education. This would not have been possible without them. I also thank them for cultivating my love of the natural world and for always encouraging me to learn.

Partial funding was provided by the National Research Foundation (NRF) of South Africa.

Disclaimer

Any opinion, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Research Foundation.

Declaration

The following thesis has not been submitted to any university other than Rhodes University, Grahamstown, South Africa. The work presented here is that of the author.

_____ Date:_____

1 General introduction: the role of taxonomy in forensic entomology

1.1 Identifying organisms

Systematics provides the filing system needed for the organisation and retrieval of the knowledge that biology produces. The core concept in this filing system is the species, and the disciplines of identification and classification are the means by which information about a specimen is accessed. Identification is extremely important in forensic entomology because one must access information about the correct species in legal cases (Villet & Amendt, 2011; Ridgeway *et al*., 2014). Classification is also pertinent when information about a particular species has not yet been gathered and one must extrapolate from the biology of its close relatives.

1.2 Beetles in Forensic Entomology

Beetles may infest stored food products and the damage that they cause is often a huge setback in the industry. Lawsuits surrounding insurance claims often seek to determine which party (sender, shipper or receiver) is liable for the damage. The stage of development of immature beetles associated with stored products can be used to indicate the duration and origin of infestation, which are major concerns of stored-product forensic entomology.

Medico-criminal forensic entomology uses insects, particularly flies and beetles, to provide evidence when investigating criminal cases, which are often concerned with the death of humans and animals (Hall, 1990; Amendt *et al*., 2007). Insects can help to estimate the time of death and establish the geographic location and nature of the death of a corpse or carcass (Catts & Goff, 1992; Anderson & Cervenka, 2002). Among other applications, insects can also be used to detect toxins in corpses (Introna *et al*., 2001).

In 1855 the first entomological case to estimate a post-mortem interval (PMI: the time since death) was concluded by Dr Marcel Bergeret using knowledge about clothes moth pupae and flesh fly larvae to determine the length of time the baby had been dead and thus who the possible parents were. In 1894, Jean Mégnin made observations on the stages of human decomposition and the associated waves of insect succession which allowed for estimates of the PMI to be made. This marked the establishment of forensic entomology as a science

(Benecke, 2001). Insects are now used in two primary contexts in forensic entomology: insect development and ecological succession. These will be described in some detail to give readers an appreciation of why identifying the insects involved are important.

1.2.1 Insect development

By observing the stage of development of specific insects, it is often possible to estimate their age, and therefore when they colonised their food source, whether that is a stored product or a dead body (Villet *et al*., 2010).

Eggs of carrion flies and beetles are usually laid in wounds and hatch into first instar larvae. They undergo complete metamorphosis and moult through a number of larva instars, and emerge as an adult from a pupa. For many species, the duration of each stage of development has been recorded. Once adults have emerged, it is very difficult to determine which generation is present, so this method is only useful when a corpse is discovered fairly recently after death. The generation time for most beetles is longer than that of the most important fly families. It is thus possible to determine the PMI until the last beetle family to arrive has completed one generation.

1.2.2 Ecological succession

Forensic entomologists are also able to estimate the PMI based on the predictable pattern of presence of certain indicator insects at carcasses (Smith, 1986; Kocárek, 2003; Schilthuizen *et al*., 2011; Villet, 2011). A carcass changes predictably in nature over the course of decomposition and so does its community of insects, a process commonly regarded as an example of ecological succession (cf. Braack, 1987; Villet, 2011). Using insect succession to determine PMI relies on knowledge of the local necrobionts and their sequence of colonisation. There have been clear associations drawn between beetle families present, habitat types and state of decomposition (Table 1.1).

The first insects to arrive at a dead body are generally flies (Calliphoridae and Sarcophagidae), usually within 24 h of its death. Other fly species, such as the Piophilidae, are attracted to the corpse during the later protein-breakdown phase of decomposition (Campobasso, 2001). The early colonisation of flies is followed by beetles, either preying on arthropods developing on the carcass or feeding on the carrion itself (Braack, 1981, 1987; Smith, 1986; Catts & Haskell, 1990; Kocárek 2001; Villet, 2011). Staphylinidae and Silphidae arrive within 24 h after death and may remain on the carcass until it has reached Advanced Decay; Histeridae, Dermestidae and Trogidae arrive soon afterwards and remain at least until

the tissue has started to dry (Table 1.1). Cleridae occupy a carcass sporadically and numbers fluctuate over the course of decomposition (Kulshrestha & Satpathy, 2007).

For this method of estimating a PMI to succeed, knowledge of the geographic, seasonal and habitat effects is needed. Some beetles are associated with buried bodies (e.g. Histeridae, Silphidae and Staphylinidae) (VanLaerhoven & Anderson, 1999; Bourel *et al*., 2004). Necrophagous beetles may also indicate movement of bodies (Benecke, 1998) and can be used as toxicological specimens for bodies containing toxic substances e.g. drugs, poisons and heavy metals (Bourel *et al*., 2001).

1.3 Southern African necrophagous beetles

Carrion beetles are present in every ecosystem besides the Polar Regions and play a major role in the long-term decomposition of animal carcasses. Of the two dozen carrion beetle families that occur worldwide, ten families occur in southern Africa, comprising about 90 species in total (Villet, 2011). In addition, members of the Scarabaeidae were also found feeding on carrion and may be important forensic indicators in South Africa (Midgley *et al*., 2012).Various morphological keys are employed to identify these beetles (Hinton, 1945; Smith, 1986; Gorham, 1987; Villet, 2011).

1.3.1 Silphidae

These beetles are large and robust with flattened bodies and, like staphylinids, have elytra that do not cover the abdomen completely. The duration of their presence overlaps with that of dipteran species, but their life cycle is longer (Midgley & Villet, 2008; Ridgeway *et al*., 2014), so that they are useful for estimating the PMI during the transition from the arrival of the first dipterans to the arrival of other beetle taxa (Villet, 2011).

Keys produced by Schawaller (1981, 1987) and Prins (1983) can be used to identify adults of the three southern African species (two *Thanatophilus* and one *Silpha* species). As there is no key to identify *Thanatophilus* larvae, these species were included in this study.

1.3.2 Staphylinidae

Commonly known as rove beetles, this family is extremely diverse and occupies nearly every terrestrial habitat on all continents except Antarctica (Grimaldi & Engel, 2005). Their distinguishing morphological feature is the reduced elytra that leave most of the abdomen exposed. Species that are associated with carrion are usually predatory and arrive at a carcass

in the bloat stage, attracted by the eggs and larvae of other carrion insects (Prins 1984a, 1984b; Braack, 1986).

Despite the ecological diversity of this family, the individual species show little morphological variation and are generally difficult to distinguish (Newton *et al*., 2000). There is presently no taxonomic key for the southern African species, but many staphylinid genera (e.g. *Aleochara*, *Philonthus*) are cosmopolitan and it is possible to identify adults to genus level using Newton *et al*. (2001); there is no larval identification system. There are almost 50000 described species in 31 subfamilies worldwide and it is believed that there are several times more undescribed species (Thayer, 2005), which makes them a good test case for identification methods.

1.3.3 Histeridae

These beetles are usually black, but some species have a blue sheen or red markings. Their bodies are ovoid and the elytra do not cover the entire abdomen, leaving the last two abdominal segments exposed. Adults and larvae of species that feed on carrion prey on other insects (Villet, 2011). They thus arrive at carrion during the Active Decay stage of decomposition, when potential prey are attracted to carrion.

This family does not have a comprehensive key to aid identification. The genera *Acritus*, *Atholus*, *Chaetabraeus*, *Chalcionellus*, *Hister*, *Hypocacculus*, *Pachylister*, *Paratropus* and *Saprinus* (Prins 1984a; Braack 1986) are found on carrion in southern Africa and general keys (Kovarik & Caterino, 2001; Caterino & Tishechkin, 2006) or those including regional taxa can be consulted for identification. Caterino & Vogler (2002) provided a key for larval identification.

1.3.4 Dermestidae

Dermestids are small to medium-sized beetles and have oval bodies. The forensically important species are brown to black and covered in varying amounts of hair. The larvae bear tufts of long setae. They are found on carcasses of varying age, but are more often present in later stages of decomposition (Braack, 1981). Dermestids are some of the most important pests of stored products and museum specimens (Bouchard *et al*., 2009).

The African species fall into three genera: *Dermestes*, *Attagenus* and *Anthrenus* (Prins, 1984b). Notes on the identification of larvae were written by Prins (1984b) and keys to adults and larvae were written by Peacock (1993) and Zhantiev & Volkova (1998).

1.3.5 Cleridae

These beetles are elongated with the first part of the thorax narrower than the rest of the body, giving them the appearance of having a neck. They have brightly coloured bodies, some with blue iridescence, and are covered by fine hairs. They are attracted to carrion in the later, drier stages of decomposition (Braack, 1981). Three *Necrobia* species are pests of stored products (Smith, 1986; Gorham, 1987; Rajendran & Hajira Parveen, 2005). *Necrobia rufipes* also eats maggots.

No taxonomic key exists for this family, but they are global pests (Smith, 1986; Gorham, 1987; Rajendran & Hajira Parveen, 2005).

1.3.6 Trogidae

Trogids are robust, brown to black beetles with arched, textured elytra. They are dominant at a carcass during the late stages of decomposition. Adults have been observed early in the decomposition process, but larvae eat older, dryer carcasses. African Trogidae species are in the genera *Trox* and *Omorgus*.

Adults can be identified using keys by Scholtz (1980, 1982, and 1983) and van der Merwe & Scholtz (2005). No key exists for larval identification.

1.3.7 Scarabaeidae

Scarabs are not commonly associated with carrion and African species tend to be incidental at a carcass or to feed on and remove material from the stomach (Braack 1981; Frolov & Scholtz, 2005; Tshikae *et al*., 2008). The occurrence of dung beetles at carrion in Africa has been documented (Braack 1981, 1986; Midgley *et al*., 2010, 2012; Villet 2011), but there has been little work done on their significance as forensic indicators. However, any species recorded on a carcass should be included in studies to determine their significance to forensic entomology in a particular area (Midgley *et al*., 2012).

1.4 Identifying Described Species

Identification commonly involved the examination of either gross physical morphology or molecular (amino acid/DNA) sequences for features that are unique to a particular species (Gullan & Cranston, 2010). Features that are not unique may still assist identification, but as is the case with all insects, it is not possible to always reliably identify an insect that is not physically complete. Thus, genetically-based identification is particularly useful (Floyd *et al*.,

2009). Pupal fragments or larval exuviae are commonly found on carrion and knowing the species to which they belong may be informative (Byrd & Castner, 2001; Catts & Goff, 1992).

Heritable morphological variation usually starts with changes at the molecular level through DNA mutation, although factors such as germline epigenetic changes may also play a role. At first, DNA mutation was largely empirically examined through experimental laboratory breeding of *Drosophila*, but after advances in biochemical and molecular techniques it was possible to measure variation more directly (deSalle & Templeton, 1988). In the 1970s, protein electrophoresis of variable allozymes in animals showed a more refined level of genetic diversity. This progressed to the examination of mitochondrial DNA in the 1980s, followed by nuclear DNA in the 1990s and the study of genomics from 2000 onwards (Allendorf, 2012). Morphology and a host of molecular methods (Table 1.2) are now being employed to discriminate one species from another.

1.4.1 Morphology

Identification based on physical traits has long been used as the standard method of identifying and distinguishing species, embodying the morphospecies concept (Mayr, 1942). Morphological differences are the easiest to compare to assign individuals to species and have been considered adequate for indicating species boundaries. This method has disadvantages in that morphologically cryptic species can be difficult to distinguish (Damm *et al*., 2010).

Higher taxa are relatively simple to distinguish, as the characteristics that define them are more obviously different than those of closely related species. For separating sister species, detailed taxonomic keys are used (McKelvey, 1982). For identification of closely related species, the use of keys can be extremely particular and may require extensive knowledge of a group. However, the problem is not always related to lack of expertise, but rather to the complete lack of species-level keys for certain groups, amongst other concerns.

For many groups that lack keys, the next best option is an identification guide. These may be parochial and incomplete, and fail to lead to species identification, but usually steer the user in the right direction (Freedman, 2005). Failing these methods, soliciting the help of an expert will suffice when there have been no published guides. However, some animal groups are extremely under-researched and expertise is limited, if not non-existent. In the absence of any living expert, museum collections can be extremely useful for species identification.

1.4.2 Molecular Methods

1.4.2.1 *Proteins*

Understanding genetic variation in natural populations started in the 1960s with protein electrophoresis and the use of allozymes (Powell, 1964). The term 'allozyme' was used to describe an enzyme coded by different alleles at a single locus that could be used as an indicator of genetic variation (Allendorf, 2012). Allozymes are first separated in a medium such as starch or polyacrylamide gel. Thereafter, the presence of specific proteins is confirmed by performing a proteolysis assay (Laemmli, 1970).

The use of proteins for species identification has some general weaknesses. Only soluble proteins can be assayed, limiting the proportion of the proteome available for study. Amino acid substitutions may not detectably change the molecular weight or charge properties of the protein, limiting the use of this method, as it relies on differences in these features to show variation.

1.4.2.2 *Nuclear DNA*

The nucleus of a cell contains the entire chromosomal complement of the organism to which it belongs. The entire set of genes and regulatory regions needed to produce and maintain an organism is contained in this DNA. Nuclear DNA contains regions of non-repetitive gene coding sequences (CDS) and repetitive non-coding DNA. The latter is more likely to vary between species and is thus more applicable for identification as coding DNA will remain more conserved between taxa (Zhang & Hewitt, 2003).

Phenotypic variation results from differences in DNA coding sequences or the regulation thereof through modulation of protein-DNA interactions or chemical changes to the DNA molecules themselves such as acetylation and methylation. Changes to the DNA sequence itself can be simple point mutations, where a single base change is incorporated into the sequence due to DNA polymerase errors. These occur as substitutions (one base substituted for another) or as indels (insertions or deletions), where an entire base is either lost or gained, but this may sometimes occur to a group of adjoining bases. The former may cause changes in the amino acid for which the DNA encodes and therefore the protein, and the latter may cause a frameshift mutation where the reading frame is altered for all bases downstream of the mutation and may render the subsequent protein useless unless the indel is three bases long (Loxdale & Lushai, 2009; Allendorf, 2012).

Single-copy protein-coding genes (such as alanyl-tRNA synthetase) occur once in the entire genome and code for specific proteins (Wiegmann *et al*., 2009). Another class of coding

DNA, regulatory elements (such as promoters and enhancers), control the rate of gene expression and timing, but do not produce a protein or rRNA. They are still considered coding because of their functional role in transcription and translation. Multigene families are often arranged tandemly on a chromosome, but may also be spread across the genome. The most prominent of these are the ribosomal RNA gene families, which occur in groups of three on the nuclear genome – the 18S, 5.8S and 28s ribosomal DNA (rDNA) families. These are separated by non-coding but transcribed DNA, the Internal Transcribed Spacers (ITS1 and ITS2) and are proceeded by an External Transcribed Spacer (ETS). This ribosomal cluster may be repeated up to thousands of times depending on the eukaryote. These genes are very important in molecular systematics where 18S and 28S give information at deep taxonomic levels and the ITS regions resolve shallower levels among closely related taxa (Weekers *et al*., 2001; Coleman & Vacquier, 2002; Avise, 2004; Young & Coleman, 2004).

A number of nuclear genes have proved useful for species identification. These include the noncoding ITS regions, 18S rRNA, Phosphoenolpyruvate Carboxykinase (PepCK) and Elongation Factor-1 (EF1a).

Minisatellites (ten to sixty base pairs) and microsatellites (two or three base pairs), are repeated many times in euchromatic regions of the genome. These types of nuclear DNA are highly variable and are popular as molecular markers as a result of their polymorphism and co-dominance (Avise, 2004).

1.4.2.3 *Mitochondrial DNA (mtDNA)*

The mitochondria found in plants and animals contain their own circular DNA (Galtier *et al*., 2009). Mitochondria originated from symbiotic bacteria that invaded or were captured by another bacterium, the outcome of which was beneficial to both organisms.

Mitochondrial DNA (mtDNA) was only discovered in 1964, but by the 1980s it was realised that it possesses a number of features that make it suitable for taxonomy. In animals, it is haploid and inherited through the female line, unlike nuclear DNA, so there is no recombination with paternal DNA at each generation, which makes it easier to follow an evolutionary pattern. It recombines, but within the same mitochondrion and with copies of itself (Allendorf, 2012). Mitochondrial DNA mutates on average about ten times faster than nuclear DNA, which makes it possible to determine recent changes in evolution and observe variation at species level (Martin, 1995; Avise, 2000; Zhang & Hewitt, 2003; Ho *et al*., 2005). Mitochondrial DNA has consistent gene content and is of much smaller length than nuclear DNA (consistently about 17 000 base pairs, whereas the size of nuclear DNA is more varied

between taxa, but is much longer). It is thus easy to isolate from nuclear DNA (Ingman *et al*., 2000; Shufran *et al*., 2000). Experimentally, mtDNA is easy to amplify because of the high number of copies per cell compared to nuclear DNA. Highly conserved regions typically flank its variable regions which allow for successful PCR primer design. Also, mtDNA is highly conserved across animals with little duplication, no introns and short intergenic regions (Gissi, *et al*., 2008; Ruiz, 2010; Simon *et al*., 1994, 2006).

A shortcoming of using mtDNA as a species marker is the preferential amplification of nuclear pseudogenes by universal primers, which may yield a sequence with a different set of single nucleotide polymorphisms (SNPs) to those in the mitochondrial copy or the presence of (non-triplet) indels (Bensasson *et al*., 2001). Amplification occurs more readily in pseudogenes when changes have occurred in the primer binding sites of the mtDNA (Moulton *et al*., 2010). In addition, there are a few instances where mtDNA has been paternally inherited in species that usually exhibit maternal inheritance such as mice (Gyllensten *et al.*, 1991), honeybees (Meusel & Moritz, 1993) and some cicadas (Fontaine *et al.*, 2007). Furthermore, some species (e.g. mussels) show double inheritance of mtDNA where both maternal and paternal molecules are present (Hoeh *et al.*, 1991; Penman, 2002). Also, when identifying animals, data from mtDNA may show accurate family groupings but should not be relied on for information regarding animal evolution before 50-150 million years ago as mutation saturation may have been reached (Whitfield & Kjer, 2008). Also, mtDNA may not accurately reflect demographics or evolutionary history because there is evidence that it is affected by natural selection (Dowling, *et al*., 2008; Galtier *et al.*, 2009; Balloux, 2010). However, for an idea of recent phylogeny or for identification, mtDNA is useful and possibly more so than nuclear DNA.

A number of mitochondrial genes have proved useful for identification. These include Cytochrome Oxidase I (COI), Cytochrome Oxidase II (COII) and ribosomal RNA (e.g. 16S rRNA and 12S rRNA).

1.4.3 Methods for DNA Taxonomy

The DNA-based taxonomic methods considered for this study are discussed below and outlined in Table 1.2. Direct nucleotide sequencing was chosen for species identification as it provides the finest level of resolution possible and can be directly compared to other taxa.

The Polymerase Chain Reaction (PCR) is a method of cloning a fragment of extracted DNA that mimics the polymerisation of DNA in the cell using tiny amounts of reagents (10-100 µl). It is a simple reaction involving a few components and is highly sensitive in that it can make millions of copies from one molecule of DNA.

Restriction Fragment Length Polymorphism (RFLP) examines variation in DNA by gel electrophoresis, similar to allozyme techniques, after the DNA has been fragmented by restriction enzymes.

Random Amplified Polymorphic DNA (RAPD) analysis is the use of short primers (10 bp), amplifying a random sequence under conditions of low stringency, producing many polymorphic fragments of DNA. In theory, using an ample number of primers (in separate reactions), one should get a random sample of a large portion of the genome. These PCR products are separated by electrophoresis and the resulting banding patterns give profiles that provide information that can be used in population studies, genetic mapping, and possibly phylogenetics (Williams *et al*., 1993).

Amplified Fragment Length Polymorphism (AFLP) is designed to assay a large part of an organism's genome to detect species-level variation. AFLP combines the property of restriction enzymes to cut the genome into many fragments with the amplifying power of PCR to generate a large number of fragments distributed across the genome.

Microsatellite DNA is short, tandemly-repeated DNA sequences consisting of two to six base pairs. They are dispersed throughout the genome and form a class of non-coding DNA, so they are under no known selective constraint and thus have very high mutation rates and are highly polymorphic within populations as a result (Brohede *et al*., 2002).

Direct DNA sequencing is the process of determining the order of nucleotides within a DNA molecule. It includes any method that is used to determine the specific order of the four nucleotide bases in a strand of DNA, which can range in length from a small polynucleotide to an entire genome. A variety of search algorithms are available for identifying sequenced DNA, depending on the type of sequence. The Barcode of Life Database systems (BOLD) and GenBank are commonly-used platforms for searching for sequences.

Genes used for identification should possess different characteristics to those used to infer phylogeny. Slowly-evolving genes are best used in phylogenetic studies as they contain information pertaining to deeply rooted relationships (Sperling & Roe, 2009).

Wells and Stevens (2008) discuss the limitations of RAPD, AFLP etc., but the finest level of genomic resolution, the nucleotide level, is provided by direct DNA sequencing. Sequences can be confidently used to show variation and are more objective than the other methods. Consequently, they are favoured above other techniques for species identification. In spite of the effort put into using DNA for species identification, it has not yet been accepted as a method of identifying species in legal cases. It is used to confirm identification by other methods, but has not been trusted to be used on its own because mtDNA-based identification

does not always correspond to species limits determined by other methods (Wells *et al*., 2007; Whitworth *et al*., 2007; Wells & Stevens, 2008), particularly in the case of hybrids where two sequences could represent one species.

1.5 Identifying Undescribed Species

One of the bigger questions in taxonomy is whether two specimens that do not match exactly belong to the same species. Hebert *et al*. (2004a, 2004b) initially proposed a threshold for differentiating species, known as the 'barcoding gap' or the '10-fold rule', where the interspecific variation in a specific gene sequence is ten times that of the intraspecific variation. Analyses have shown the COI barcode sequence to discriminate 95% of species and not diverge by more than 2.5% within species (Hebert *et al*., 2003a, 2003b). However, assessing the existence of the barcoding gap begs the question of what exactly a species is. Currently, there are about 30 different species definitions (Mayden, 1997; de Queiroz, 1998; Harrison, 1998; Coyne & Orr, 2004).

Taxonomists have been describing species since the founding of Linnaean taxonomy in 1758. There are currently around 1 million described animal species out of a possible 7.7 million yet to be discovered (Mora *et al*., 2011). Thus, the majority of species are still to be described and many named taxa are actually species complexes (Bickford *et al*., 2007; Trontelj & Fišer, 2009). Nevertheless, there is still the need for a scientific consensus to be reached with regard to the working definition of a 'species' and which categories need to be incorporated when delimiting them (May & Harvey, 2009).

1.6 Motivation, scope and aims for thesis

1.6.1 Aim 1: Providing barcodes

This study aimed to sequence the barcoding region of the COI gene in a number of South African carrion beetles with the intention of adding them to the GenBank database, primarily to be of use to forensic entomologists. This will facilitate the identification of morphologically challenging specimens of beetles (e.g. juveniles and damaged adults) found on carcasses by comparing their COI sequences to those in the database. This would allow for identification of more specimens per carcass by non-specialists, especially if specimens are damaged beyond physical recognition, and reduces the time taken to identify juveniles with few or no distinctive physical characteristics because it is not necessary to wait for them to mature before identifying them. Thus, using DNA barcodes, forensic entomologists should be

able to determine the PMI of a carcass in a shorter time frame (Malgorn & Coquoz, 1999). Also, the barcodes obtained in this study will be helpful in identifying beetles that are particularly difficult to identify using external characteristics (e.g. staphylinids).

1.6.2 Aim 2: Interpreting variation

The second outcome of the study is to determine whether there is a significant amount of variation between individuals from different populations. Identification requires only that there are characters that identify each target taxon uniquely. An objective of this study is to determine whether DNA barcodes allow for unambiguous differentiation between closely related taxa of carrion beetles. Specimens from different areas within southern Africa will also be assessed for intraspecific variation.

1.6.3 Aim 3: Placing species not represented in barcode databases

Most described species are not represented in the barcode databases, and many more are not even described. If the specimen one is trying to identify represents one of these species, one needs a means to know this. Also, one would want to know what its closest barcoded relatives are to pursue other means of identification more readily and draw on biological information about those relatives that might have some forensic relevance. One way to do this is to submit the barcode in question to on-line identification tools that are available through the websites of Genbank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Barcode of Life Database (http://www.boldsystems.org/). A second approach is to include the specimen in a phylogenetic analysis with likely relatives and conspecifics. Both of these methods will be tested in this study with reference to forensic needs.

1.6.4 Aim 4: Comparing barcodes to other regions of COI and COII and mini-barcodes

It is debatable whether barcodes represent the best fragment of DNA for identifying species (Dayrat, 2005). To test this, the region downstream from COI, including tRNA leucine and COII will be sequenced and the variability along the entire region (including the barcoding fragment) will be analysed using sliding window analysis to determine how variable the barcoding region is for these beetles. Interfamily diversity will be calculated for the entire region as well as for the barcodes and a fragment length analysis will determine the minimum sequence length required to yield accurate identification, comparable to mini-barcodes.

Mini-barcodes are partial sequences of the barcode region that are hoped to yield the same quality of identification while bringing down the costs of sequencing by consuming less

of the reagents. Cost is a consideration in commercial and other forensics laboratories. The sliding window analysis will be used to seek effective mini-barcodes for the species studied here.

Table 1.1: Beetle families and their associated stages of decomposition. Adapted from Braack (1987) and Villet (2011).

	Fresh	Bloat	Active Decay	Advanced Decay	Skeletal
Silphidae	•	•	•	•	
Staphylinidae	•	•	•	•	•
Cleridae		•	•	•	
Scarabaeidae		•	•		
Histeridae			•	•	•
Dermestidae			•	•	•
Trogidae				•	•

Table 1.2: Summary of molecular identification methods.

Method	Application	Advantages	Limitations	Literature
PCR-RFLP	Population- and species-level	Cheap, straightforward (little expertise needed), can use extract, larger sample than DNA sequencing	Null alleles - where mutations in the restriction site result in loss of a fragment, only small part of the genome screened	Peng *et al*. (2003); Beebe *et al*. (2007); Li *et al*. (2007)
RAPD	Population- and species-level	Cheap, no need for previous knowledge of genome, produces many markers	Difficult to reproduce, requires good quality DNA, not comparable across studies, dominant marker	Benecke (1998); Al-Barrak *et al*. (2004)
AFLP	Population- and species-level	Reproducible, cheap, no need for previous knowledge of genome, produces many markers	Requires good quality DNA, not comparable across studies, dominant marker	Parsons and Shaw (2001)
Microsatellites	Kinship, populations	Cheap, screens large proportion of genome, reproducible, highly variable	Expertise needed, time consuming, high mutation rates (slippage), difficult to obtain from some insects	Tsutsui *et al*. (2001); Zakharov & Hellman (2008)
Direct sequencing	All taxonomic levels	Highest resolution of genetic variation and homology, comparable across taxa	Moderately expensive, only screens short targeted fragment	Zaidi *et al.* (2011); Kengne *et al.* (2007); Mardulyn and Whitfield (1999); McDonagh *et al.* (2009); Raupach *et al.* (2010)

2 COI DNA Barcoding

2.1 Introduction

The term 'DNA barcode' is employed to describe a standard region of DNA that is anticipated to distinguish species from one another on the basis of its nucleotide sequence. It has been proposed that a ~658 bp fragment of the cytochrome c oxidase subunit I (COI) gene protein-coding sequence (CDS), as flanked by the universal primers LCO1490 and HCO2198 (Folmer *et al*., 1994), has the requisite characteristics to act as a species-specific barcode that could be used to identify any animal quickly and accurately (Hebert *et al*., 2004a; Hajibabaei *et al*., 2007; Burns *et al*., 2008; Walker & Leedham-Green, 2010).

Having coverage of all species would allow barcodes to be applied universally, including undescribed species (Rubinoff *et al*., 2006; Valentini *et al*., 2009). Conserved flanking regions of DNA across taxa are thus imperative for design of universal primers and the presence of these regions are required for a successful barcode.

Barcodes have been used in the tracking of illegal bush meat (Baker *et al*., 1996) and the detection of illegal whaling (Eaton *et al*., 2009). It has forensic use as mtDNA can be extracted from very small amounts of tissue and used for identification. No taxonomic knowledge is needed for the sequencing and analysis of COI so it is reputedly available to laypersons (Hebert *et al*., 2003a).

2.2 Caveats for DNA Barcoding

At present, the laboratory equipment needed to obtain a sequence from extraction is expensive and stationary. The aim of the Barcodes of Life project is to develop portable, hand-held devices that are connected to databases via satellite which are capable of sequencing COI from an organism and identifying it (Walker & Leedham-Green, 2010; http://www.dnabarcoding.ca/barcode_initiative.php). However, problems associated with molecular systematics such as contamination, paralogy, incomplete lineage sorting, hybridisation and identification errors could be experienced. In the laboratory these problems are usually apparent and can be solved through analysis of the data. They become more important when there is no longer a person involved in the process of sequencing and communication is directly between the database and DNA sequencer (Will *et al*., 2005).

Paralogy. Sequences are said to be paralogous if a gene is duplicated to occupy two different positions on the same part of the genome. Recombination of bases in COI is rare and the haploid nature of mtDNA allows for the assumption that analysed sequences are homologous, but parts of the mitochondrial genome are often copied to the nuclear genome (Mourier *et al*., 2001). PCR products of COI will usually produce true mitochondrial sequences, but sometimes nuclear copies of mitochondrial genes (NUMTs) will be preferentially amplified and sequenced instead. This will result in the produced sequences being paralogous to those in the database and comparisons cannot be made to identify the specimen (Williams & Knowlton, 2001; Thalmann *et al*., 2004).

Incomplete Lineage Sorting. Incomplete lineage sorting is "the presence of ancestral lineages in two or more species that are more closely related to each other than to other lineages within a species" (Wahlberg *et al*., 2003). Much of the evolutionary history will be missed if a single gene sequence is used to identify species, and recently diverged taxa may be overlooked easily (Choat, 2006).

Horizontal Gene Transfer. In horizontal gene transfer, genetic material from one organism is incorporated into another via methods other than traditional reproduction. It can be relatively common in plants, but is considered rare in animals, except in cases of hybridisation (Kurland *et al*., 2003). Rot *et al*. (2006) show evidence for this to have happened in a sea sponge. This would give an incorrect COI sequence for that species and it would not correlate to the correct sequence on the barcode database.

The movement of a gene from one individual to the gene pool of another through repeated backcrossing of an interspecific hybrid with one of its parents is known as introgression. This is another source of paraphyletism of mtDNA and is also not uncommon in animals (Quesada *et al*., 1995). It has been documented in tuna (*Thunnus*) (Ardura *et al*., 2013) species and *Lucilia* blowflies (Sonet *et al*., 2012; Sonet *et al*., 2013; Williams & Villet, 2013) so mitochondrial markers would thus not be suitable for identifying these taxa on their own. As a result of this, mitochondrial gene trees will not be a correct representation of the species. Nuclear DNA is completely responsible for the phenotype of an organism. If an organism of one species were to have the mtDNA from another (as in horizontal gene transfer and introgression) it would still belong to the species of its nuclear DNA and identification based on COI would therefore be incorrect (Kurland *et al*., 2003).

GenBank has many errors in its database. These occur in sequences that are submitted under the wrong identification or even contain mistakes in the sequences themselves (Wells & Stevens, 2008). The Barcode of Life project addressed these errors by establishing quality

standards for uploading sequences into BOLD and requiring voucher specimens for each sequence. However, many sequences are submitted to GenBank (and therefore BOLD) and they will not all be verified after they have received their initial name. Once the error has been made it is difficult to pick it out and once the identification is used in one journal, a cascading set of errors could be started (Will *et al*., 2005). This highlights the need for accurately identified specimens prior to their inclusion in a barcoding database, especially when there are no other sequences for that species (Meier *et al*., 2006).

2.2.1 Detecting Known and Unknown Species

Another aim of the Barcode of Life project is to identify and get COI sequences for as many species as possible in the shortest time. For one organism to be considered a separate species from another, the COI sequences need to have a divergence of more than 2.5% (Hebert *et al*., 2003b). In other words, if the sequence of one organism is less than 2.5% divergent from a sequence in the database, it will be identified as that species. The fundamental idea behind this is that intraspecific divergence is lower than interspecific distance and the standard divergence threshold should be ten times the intraspecific distance (10-fold rule) (Hebert *et al*., 2004b). This value is usually low (less than 2.5%) across a wide range of taxa (Hebert *et al*., 2003a; Ball *et al*., 2005; Ward *et al*., 2005; Hajibabaei *et al*., 2006). Several studies have investigated this rule since it does not have strong biological reasoning (Lefebure *et al*., 2006; Kartavtsev & Lee, 2006). It has not been shown to have a universal application in the delineation of species (Pamilo *et al*., 2007; Song *et al*., 2008) and should thus not be explicitly relied on when choosing a threshold to discriminate species.

Another method estimates species boundaries using a general mixed Yule-coalescent (GMYC) model to analyse mtDNA branching times and to show the lineages from coalescence to speciation on a phylogenetic tree (Pons *et al*., 2006). This approach has been tested and has shown potential (Monaghan *et al*., 2009; Papadopolou *et al*., 2009), but population bottlenecks in the past may interfere with this method of constructing phylogeny. Also, possible introgression and incomplete lineage sorting could also impact the result, especially when analysing large datasets (Raupach *et al*., 2010).

A cutoff distance has also been considered as an indicator for species discrimination (Floyd *et al*., 2002; Hebert *et al*., 2003a; Blaxter, 2003). Any specimen that falls below this threshold is considered to be a conspecific and those that lie above it are considered to be heterospecifics (Meier *et al*., 2006; Whitworth *et al*., 2007). It has been suggested that this is effective in distinguishing species and that a measure of variability between intra- and

interspecific variability (mean interspecific distance being ten times that of the mean intraspecific distance [Hebert *et al*., 2004a]). Ideally, there should be a range between the intra- and interspecific pairwise distances that is not occupied, an idea called the 'barcoding gap'. Hence, non-overlapping ranges of intra- and interspecific sequence divergences are often used as evidence for species delineation. Barcoding promoters seem to accept that a species is differentiated from others if there is such a clear division. However, the barcoding gap is defined misleadingly in many papers as they compare the mean values of all intra- and interspecific distances (one value compared with another) which would generally be different, but it is important to compare the frequencies of each value for intra- and interspecific distance (Meier *et al*., 2006). The distance threshold between sequences is therefore a point of contention (Puillandre *et al*., 2012; Virgilio *et al*., 2012; Zhang *et al*., 2012).

A function of distance thresholds is to prevent misidentification of specimens without conspecifics in a reference library (Virgilio *et al*., 2012). Before barcoding there was no reason for a set threshold across all species as they have varying coalescent depths (the point on the phylogeny where two specimens share the same ancestor) as a result of differences in population size, rate of mutation and time since speciation (Monaghan *et al*., 2009; Fujita *et al*., 2012). An interspecific divergence threshold of around 2.5% was suggested by Hebert *et al*. (2003a, 2003b), but would not necessarily be applicable in all organisms e.g. plants. A threshold of 1%, for example, could be adequate in some cases, but would yield varying rates of false positives (saying two specimens are different species when they are not) and false negatives (saying two specimens are not different species when they are), depending on the sequence set. Relying on any set threshold may result in these errors, so it would be better to optimise species cut-offs directly from the data (Meyer & Paulay, 2005; Virgilio *et al*., 2012). Therefore, the threshold approach may not rely on having a universal value as higher taxa would be more diverse than closely related ones.

There have been cases that show the barcoding gap to not exist (Meyer & Paulay, 2005; Wiemers & Fiedler, 2007; Yassin *et al*., 2010), and even where it does exist, it is inconsistent and not fully reliable in many groups (Meyer & Paulay, 2005; Whitworth *et al*., 2007; Trewick, 2007). Burns *et al.* (2010) pointed out that having a consistent but arbitrary percentage or degree for species differentiation is unrealistic.

In a study by Hebert *et al*. (2004a) seeking a relationship between species defined by taxonomy and those defined by barcoding, DNA barcodes of 260 of the 667 bird species that breed in North America were sequenced. They found that every single one of the 260 species had a different COI sequence. 130 species were represented by two or more specimens and in

all of these species, COI sequences were either identical or were most similar to sequences of the same species. COI variations between species had an average of 7.9%, whereas variation within species averaged 0.4%. In four cases there were deep intraspecific divergences, indicating possible new species (Hebert *et al*., 2004a). Three of these four species are already split into two by some taxonomists. However, Ardura *et al*. (2013) tested various mitochondrial markers for identification of marine and freshwater fish and found that COI and Cytochrome oxidase B (CytB) gave ambiguous identifications. The non-coding D-loop region was comparatively more variable for the species used in the study. Ideal markers may thus vary between animal groups and it could be necessary to develop taxon-specific barcodes, rather than assume one short region to sufficiently discriminate all species.

2.2.2 Geographical Variation

Most insects are free to migrate within their distribution and their populations should be subject to a degree of gene flow (Pamilo *et al*., 1997). The presence of shared or transitional haplotypes at a locality is partly indicative of a species' ability to disperse, although factors such as mutation and genetic drift may also play a role. Genetic diversification within and between populations of flightless beetles is higher than in those that can fly and the speciation rate in flightless lineages is twice that of the flying species (Ikeda *et al.,* 2012). While at odds with the goals of barcoding, such population differentiation could aid forensic entomology if it indicates the origin of insect specimens. Bergsten *et al.* (2012) showed barcodes to change slightly within species between regions of origin, which is encouraging for forensic applications looking at whether bodies and their insects have been moved. However, their study as pitched on a global scale, while bodies are usually moved on much smaller scales, more like the local scale of this thesis. There is thus an opportunity here to test the extent to which forensic entomology can detect the transport of bodies.

2.2.3 Analytical Methods

Identification begins with by aligning sequences to allow homologous characters to be identified. The result is then analysed with a grouping procedure that may rely on a phylogenetic algorithm. Analyses can be performed on aligned sequences that include deletions at various points in the region used. Analysis coding gaps as a fifth state character or as separate presence/absence characters outperformed those treating gaps as missing data (Ogden & Rosenberg, 2006).

The neighbour-joining (NJ) algorithm has been applied to the data of almost every barcoding study as part of the standard analysing practice to produce a visual representation in the form of a tree (Casiraghi *et al*., 2010). It was first added to the barcoding protocol by Hebert *et al.* (2003a). They produce a hierarchical clustering phenogram based on a distance matrix of similarity of the sequences (Little & Stevenson, 2007). The benefit of using them is that they are quick and easy to create for large datasets. However, the use of NJ in barcoding has been debated for its appropriateness (Will & Rubinoff, 2004; Meier *et al*., 2008; Goldstein & De Salle, 2010). Neighbour-joining analysis has not been particularly useful for identification purposes (Meier *et al*., 2006; Virgilio *et al*., 2010; Zhang *et al*., 2012) since trees may be incorrectly interpreted when constructed with reference to a library with an inadequate number of reference samples. The tree cannot be exclusively relied on to determine whether an unknown specimen belongs to a species unless it falls in the middle of a species-cluster and it is not possible to gauge its relatedness to the other specimens on the tree. If a sequence does not have an exact match, the tree cannot give a "no identification" result as a sequence will always inaccurately group with the most similar available sequences (Will *et al*., 2005; Collins *et al*., 2012). This problem would not be prevented using any other tree inference method as datasets with incomplete lineage sorting would lead to incorrect or ambiguous identifications (Lowenstein *et al*., 2010). However, their function in barcoding is to represent the data in a way that is visually interpretable and to allow for pinpointing of quirky results in datasets with known identifications. When trying to identify an unknown specimen, comparison to other species in a NJ tree would be reckless if done without finding the best close match for it on barcoding databases.

2.3 Conclusion

DNA barcoding has changed the systematics landscape in under a decade, despite its potential pitfalls. Work continues in determining the limits of its applicability and in the verification of the accuracy of the methodology. However, it can be stated with certainty that DNA barcoding is an accepted method of species identification routinely used by the majority of biodiversity related biological scientists today.

The premise of barcoding assumes that identifications made based on a COI sequence are augmented with conventional methods of identification (Hebert *et al.*, 2003a). A taxonomic system assisted by barcoding will significantly reduce the amount of work for taxonomists. Time spent distinguishing species from each other will be shortened and additional study will

further support identifications made through COI barcodes (Hurst & Jiggins, 2005). In the future, barcoding could become vitally important in biodiversity research and it would be unreasonable to disregard the advantages that it could offer despite its potential pitfalls. Scientists should capitalize on the potential that DNA barcoding could have as well as the financial benefits it would offer.

In our work which follows, the ability of COI barcodes to identify specimens of southern African carrion beetles at various life stages was tested. The aim of this study was to determine if the sequences obtained uphold the accepted barcoding criteria and assess how suitable they are for distinguishing species in a forensic entomological analysis.

3 Barcoding of southern African carrion beetles (Insecta: Coleoptera)

3.1 Introduction

In Chapter 1 it was argued that forensic entomology requires sound identification of the insects used as evidence and that molecular taxonomy might provide a tool to do that. In Chapter 2 it was explained that "DNA barcoding" has been proposed to have a significant application in this endeavour (Cameron *et al*., 2006; Hanner *et al*., 2011). In this chapter, barcodes are presented for a variety of southern African carrion beetles. The results are examined to see how they bear out the promise of the barcoding concept. In particular, this chapter aims to establish the potential for DNA barcodes to accurately distinguish both adults and juveniles from other species and to successfully identify unknown taxa. Analytical methods are compared and the effect of geographical variation is analysed. The effect of paralogy and hybrids on barcoding is also assessed.

DNA barcoding has been proposed not only as a means of identification, but also to define species' boundaries and allow the discovery of species (Hebert *et al*., 2003a). It has been used to address a number of problems including resolving adult and larval identifications (Gossner & Hausmann, 2009), controlling the species of fish sold in supermarkets (Rasmussen *et al*., 2009) and in food quality control (Jones *et al*., 2013). It has been used successfully in the identification of fishes (Ward *et al*., 2005), crustaceans (Lefebure *et al*., 2006), North American birds (Hebert *et al*., 2004b; Aliabadian *et al*., 2013), tropical lepidopterans (Hajibabaei *et al*., 2006) and cave-dwelling spiders (Paquin and Hedin, 2004), to name a few.

Although neighbour-joining analysis is favoured as an identification tool in barcoding studies (Hebert *et al*., 2003a; Casiraghi *et al*., 2010), it is known to produce artefacts under various conditions. It has also been documented to perform erratically for specimen identification (Meier *et al*., 2006; Whitworth *et al*., 2007; Virgilio, *et al*., 2010; Little, 2011; van Nieukerken *et al*., 2012; Zhang *et al*., 2012). When compared against with an incompletely sampled reference library, there may be no way to determine whether an inexactly matched specimen belongs to the closest topological species or a missing taxon because it is unable to give a 'no identification' in the absence of an exact match (Will *et al*., 2005; Collins *et al*.,

2012). For this reason, a variety of phylogenetic methods were used to analyse the entire data set.

3.2 Methods and Materials

3.2.1 Sampling

Three hundred and eighty-four beetles from the families Dermestidae, Silphidae, Histeridae, Staphylinidae, Scarabaeidae, Trogidae and Cleridae were collected from mammal carcasses from various locations in southern Africa (Table 3.1, Figure 3.1) and preserved in 100% ethanol. Voucher specimens were mounted and donated to the Albany Museum (Table 3.2).

All three legs from one side of each specimen were prepared for DNA analysis to allow for morphological identification of the remainder of the specimen. This made it possible to determine whether their relationships based on a neighbour-joining tree reflect expected taxonomic identifications.

Larval specimens of *Dermestes maculatus*, *D. haemorroidalis, Necrobia rufipes* and *Thanatophilus micans* were also processed and analysed to confirm that barcodes can be used to identify any life stage of an insect. These measures determine whether forensic entomologists could potentially use this method for identifying juvenile carrion beetles that cannot be readily identified morphologically.

3.2.2 DNA Extraction

The Qiagen DNeasy procedure for purification of DNA from animal tissues (Qiagen, Valencia, CA) was used to extract DNA according to manufacturer's specifications with slight modifications. Some of the samples were ground using liquid nitrogen before being placed in lysis buffer, as this was more effective in digesting tissues. The protocol called for elution with 200 µl H_2O but due to small amounts of tissue used in extraction, 20-60 µl H_2O was used to increase DNA concentration. DNA concentration was measured using a NanoDrop 2000 (ThermoScientific, Boston, Massachusetts).

3.2.3 Amplification

A 708 bp region of the COI gene was amplified using the primers LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer *et al.*, 1994) for most of the

specimens used in this study. For *S. bicolor*, *S. splendens*, *N. ruficollis*, *D. haemorrhoidalis* and all of the staphylinids, amplification with LCO1490 was unsuccessful and the primer TY-J-1460 (5'-TACAATTTATCGCCTAAACTTCAGCC-3') (Simon *et al*., 1994) was used as an alternative.

Additional primer sets were used to attempt amplification of the trogids following the failure of the barcoding primers. All combinations of the forward primers C1-J-1751 (5'-GGATCACCTGATATAGCATTCCC-3') and C1-J-2183 (5'-CAACATTTATTTTGATTTTTTGG-3') with reverse primers C1-N-2191 (5'-CCCGGTAAAATTAAAATATAAACTTC-3'), TL2-N-3014 (5'TCCAATGCACTAATCTGCCATATTA-3') and TL2-J-3033 (5'-AATATGGCAGATTAGTGC-3') (Simon, *et al*., 1994) were used.

The Polymerase Chain Reaction (PCR) reaction mixture was composed of 12.5 µl of the PCR master mix (Promega, Wisconsin, USA) which includes 3mM $MgCl_2$, *Taq* polymerase and 0.2 mM of each dNTP (deoxynucleotide triphosphate), between 1 and 1.5 µl (10 µM concentration) of both primers, an additional 2-3 µl $MgCl_2$ (50mM concentration) and 4.5-5.5 µl nuclease-free water, added to 2-3 µl of template DNA (of 55-100 ng/µl), bringing the total volume of each reaction to 25 µl. Cycling was carried out in a Thermo Hybaid PX2 (ThermoScientific, Boston, Massachusetts) thermocycler. The initial denaturing step was carried out at 94°C for between 30 seconds and 5 minutes, followed by 38-45 cycles of 94°C for 30 seconds, an annealing temperature of 38-48°C for 1 minute and extension of 72°C for 1.5 minutes. A final extension period of 5 minutes at 72°C was used, followed by holding at 4 or 15°C. The same range of cycling conditions was used for the additional trogid reactions, besides the annealing temperature which ranged between 38-50°C. Negative controls were used in all PCRs.

The DNA was visualised by electrophoresis on a 1% agarose gel, stained with ethidium bromide, SYBR Green or SYBR safe (Life Technologies, California, U.S.A.), at 80-100 V for 15-30 minutes, and viewed with UV transillumination. Product band size was confirmed against a 100 bp ladder (KAPA Biosystems, Cape Town, South Africa).

3.2.4 COI Sequencing

The COI barcoding region of the products of 26 samples was sequenced using an ABI PRISM Big Dye Terminator 3.1 Sequencing Kit (Perkin Elmer) at the Central Analytical Facility at Stellenbosch University. The remaining 208 were cycle-sequenced and sequenced

in both directions at Macrogen (Seoul, Korea) on an ABI3730 XL automatic DNA sequencer, using the same primers as used in PCR.

3.2.5 Alignment

Forward and reverse contigs were edited and aligned with BioEdit 7.1.3.0 (Hall, 1999) using the ClustalW 1.8 (Thompson *et al.*, 1997) algorithm. Sequences were checked based on signal strength and read length, and further analysed using MEGA 5.05 (Tamura *et al*., 2011) to check for evidence of stop codons and reading frame shifts that might be evidence of paralogy (Michu, 2007; Wild & Maddison, 2008).

Twelve additional sequences of beetles closely related to members of the six families sequenced were obtained from the GenBank nucleotide database (www.ncbi.nlm.nih.gov) and added to the data set for comparison to the southern African specimens to test their potential to cluster with, but be distinguished from, other closely related specimens (Table 3.3). The numbering of base pairs is based on the *Drosophila yakuba* genome which starts at Dy#1 and ends at Dy#16019 (Clary & Wolstenholme, 1985). The sequences were aligned with the *D. yakuba* sequence to show their position on the mitochondrial genome.

3.2.6 Specimen Identification

Each specimen was submitted to Nucleotide Basic Local Alignment Research Tool (nBLAST) and BOLD for identification against their databases. The best matches were recorded (Table 3.2).

Sequences were partitioned into codon positions using the protein translation tool in MEGA 5.05. To cross-validate the identification of each specimen, a dendrogram was created using neighbour-joining analysis using MEGA 5.05 (Tamura *et al*., 2011) to see whether samples clustered with conspecific samples from southern Africa and from GenBank or BOLD. Branch lengths were checked for evidence of aberrant lengths that might be evidence of paralogy (Michu, 2007).

A number of tree-based methods were performed to validate the outcome of each. A Maximum Likelihood (ML) tree was created using the entire data set with 1000 bootstrap replications using MEGA 5.05 (Tamura *et al*., 2011) based on the Kimura 2-parameter (K2P) model (heuristic search with 1000 random step-wise additions). K2P was used as it is the preferred model for the barcoding protocol (Hebert *et al*., 2003a). Bayesian Inference (BI) analysis was performed using the MrBayes v3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) software available through the CIPRES Scientific Gateway (Miller *et*

al., 2010). Four MCM chains were run simultaneously, and each was run for 10^6 generations with sampling every 1000^{th} cycle. One thousand trees were discarded as burn-in samples. Bayesian posterior probabilities were transferred to the ML tree to compare the outcome of the different algorithms with regard to consensus of sequence delineation. The neighbour joining tree was constructed in MEGA 5.05 (Tamura *et al*., 2011) using the K2P model and node support was assessed with 1000 bootstrap replicates, following the standard barcoding protocol (Hebert *et al*., 2003a).

3.2.7 Taxonomic and Geographical Variation

For assessment of sequence variation within species among geographic regions, intraspecific diversity and divergence was calculated using MEGA 5.05 (Tamura *et al*., 2011). Sequence variation was displayed using distance and haplotype networks for the barcode region. Calculations of genetic distances between species with two or more sequences were performed in MEGA 5.05 (Tamura *et al*., 2011), based on Kimura's (1980) 2-parameter (K2P) distance.

To visualise any phylogenetic uncertainty and determine whether distance-based methods resulted in similar phylogenetic relationships, we created a network diagram using SPLITSTREE 4.10 (www.splitstree.org: Huson & Bryant, 2006), which may provide a more appropriate representation of relationships at the intraspecific level. Unrooted distance trees were created with the uncorrected P distance in NeighborNet (Husen & Bryant, 2006) using species with more than ten sequences and varied locations. Trees were created to display nucleotide variation between locations of collection. Trees were created for *D. maculatus*, *F. forcipatus*, *S. bicolor*, *S. splendens*, and *T. micans* to display nucleotide variation between collection locations.

Parsimony haplotype networks (95% connection limit) were created in TCS 1.21 (Clement *et al*., 2000) to show the diversity and phylogenetic relationships among the different haplotypes and to qualitatively assess the distributions of six species' haplogroups (*D. maculatus*, *F. forcipatus*, *N. rufipes S. bicolor*, *S. splendens* and *T. micans*) from southern Africa.

3.3 Results

3.3.1 Amplification and Alignment of Sequences

Of the 384 DNA extracts, only 234 were successfully amplified and bidirectionally sequenced using the universal barcoding primers or the primer TY-J-1460 (Table 3.2). Barcodes were therefore obtained for only 61% of the specimens collected.

One family in particular, the Trogidae, was problematic in terms of PCR success even when additional primers were used. Six PCR products yielded bands for the trogids, but in four of these the peaks in the chromatograms were too small to be considered reliable.

A 577 bp fragment for a total of 30 species in five families were sequenced for the DNA barcoding region of COI. The alignment revealed 267 variable sites. The fragments contained no indels, except that both forward and reverse contigs of all 15 specimens of *N. rufipes* tested contained a 67 bp deletion (Dy #1851-1918).

3.3.2 Specimen Identification with BLAST and BOLD

The sequences matched those of conspecific beetle species using the BLAST function on the NCBI website and the BOLD Identification System (Table 3.2). Database searches showed 35.9% of nBLAST and 58.1% of BOLD searches to have >80% identity matches and >90% query cover, with the first result to satisfy both conditions.

3.3.3 Identifying Juveniles

Ten larvae (*D. haemorrhoidalis* (2), *D. maculatus* (6) and *T. micans* (2)) amplified successfully. Two *N. rufipes* larvae and three eggs were not amplifiable. All ten juvenile specimens clustered with conspecific adults.

3.3.4 Geographical Variation

The average intraspecific distance was plotted against the number of specimens to determine whether a relationship between the two variables exists (Figure 3.2). Since the graph does not show a strong relationship ($R^2 = 0.0279$) it can be assumed that a larger sample of sequences will not yield a lower mean intraspecific distance.

The haplotype networks illustrate the variation within the COI barcoding region for southern African necrophagous beetles. For the determination of ancestral haplotypes, the ancestral state is the state that is present in the highest number of populations, and defined by the number of lineages that arise from them, shown in the networks as a square or rectangle.

The location of haplotype pie charts are shown in Figure 3.1 and the chart in Figure 3.3 shows colours corresponding to their respective haplotypes.

Necrobia rufipes (Figure 3.4). Generally speaking, there is no definite geographical pattern in the networks. However, the specimens from Griekwastad and Cradock group together and are separate from the other localities. Specimens from SGR and AEP appear to be closely related and three of the BCC sequences fall on the opposite side of the distance network to the others. Those from GHT fall with BCC and AEP, and GWD respectively.

Saprinus splendens (Figure 3.5). Three of the Williston (WSN A54, D53 and D65) specimens on the *S. splendens* tree show a tendency to cluster together more than those from others sites, but there is no definite geographical organisation for any of the localities.

Saprinus bicolor (Figure 3.6). Four haplotypes were found in the samples sequenced from GHT and SGR. The SGR haplotype was the only one found in this location, but was shared by the GHT samples.

Thanatophilus micans (Figure 3.7). Similarly, in the *T. micans* distance network, a haplotype was shared between beetles from BCC and MCK, and the SGR and GHT specimens cluster separately.

Dermestes maculatus (Figure 3.8). The predominant root haplotype of *D. maculatus* (A) is distributed across the country (present in the Limpopo, Northern Cape and Eastern Cape) and found exclusively in Hondeklipbaai, Griekwastad and Grahamstown. This is one mutational step away from the second-most-widely distributed one (B) present in the Kuruman and Saldanha Bay specimens. The G haplotype is found only in the Northern Cape specimens suggesting a more localised distribution of that particular sequence.

Frankenbergius forcipatus (Figure 3.9). These specimens were all collected from the same carcass in the Grahamstown area, but there are seven haplotypes present.

3.3.5 Genetic Variation

The average intraspecific variation was 0.3% and the average interspecific variation was 1.9%, corresponding to 6.3 times that of the intraspecific variation. There is, however, overlap between the two measures (Figure 3.10). Interspecific and intraspecific variations share the values of 0.4% (two and three specimens respectively), 0.5% (two and one), 1.1% (ten and one) and 1.5% (eight and one) (Table 3.4).

Intraspecific distances were calculated in MEGA 5.05 (Tamura *et al*., 2011) for all species with two or more sequences. The highest intraspecific K2P distances belonged to *Necrobia rufipes* (0.48) and *T. micans* (0.36) (Table 3.5).

3.3.6 Influence of the Analytical Method

The maximum likelihood (ML) bootstrap values were largely in agreement with the support metrics of the neighbour joining (NJ) and Bayesian inference (BI) calculations (Figure 3.11). A circular topography of the tree in Figure 3.11 is shown in Figure 3.12. All families besides the Staphylinidae were differentiated from one another. The dermestids and staphylinids were monophyletic, but their placements in relation to other families did not reflected those in the molecular phylogeny of Hunt *et al.* (2007). In the ML tree, the histerids are paraphyletic with the staphylinids and dermestids. The clerids are monophyletic with this clade, which diverges from the scarabs. The placement of specimens on the neighbour-joining tree was generally in agreement with the maximum likelihood tree (Figure 3.13). The topological order of families was different to that on the ML tree and species represented by few sequences were not as clearly discriminated.

Three specimens do not cluster with their respective families: Histeridae D80 lay between the dermestids and silphids, Histeridae 4a lay within the staphylinids and *Aleochara* sp 11ai lay within the clerids. These cases give the dataset a 1.3% failure rate (3 out of 234 specimens).

3.4 Discussion

3.4.1 Ease of Amplification and Alignment

This study assessed the value of DNA barcoding and how it can be applied as a method of identifying beetles of forensic importance.

The overall success rate of amplification was 66%. Amplification success was limited with the universal primers. The forward primer (LCO1490) had more of a problem annealing to the target site than the reverse primer and was replaced with TY-J-1460 in 93 (39.7%) specimens to sequence the barcode region. The applicability of universal primers is a main criterion for successful DNA barcoding, but fell short for a number of specimens in this study. The annealing temperature(s) used in this study (38-48°C) were consistently lower than the recommended annealing temperature for the barcoding primers (50°C [Hebert *et al*., 2003a]). This suggests that the primers are not as specific as intended. Also, the PCRs may have been obstructed by inhibitors in the DNA extract elution buffer. For non-experts optimising a PCR protocol may require proficiency beyond their level of expertise and is time-consuming. This does not bear out the barcoding principles of a standardised protocol with the possibility of

being used by laypersons. The species of one family in particular, the Trogidae, proved particularly difficult to amplify. A range of PCR protocols were used on their DNA extracts, but out of the 38 extracts, only two could be amplified. Degradation was ruled out as an explanation for amplification failures as fresh specimens were used in optimal-yield extractions that were still unsuccessful. Only one barcode for *Trox unistriatus* (AY165707) is found on GenBank and only 47 more (unreleased/private) are listed on the BOLD database, so it may be inherently difficult to amplify the barcoding region for this family. A possible explanation would be mutations in the primer binding sites of the family. The sister species to *T. micans*, *T. mutilatus*, could also not be amplified using a wide set of PCR conditions, so it leaves the question of whether the primers are indeed universal resolved for this family (Silphidae), but unresolved for the Trogidae. Similarly, Cox *et al*. (2013) had problems with barcode amplification of stag beetles (Lucanidae) and thus used a fragment downstream of the standard region for discrimination purposes.

Folmer *et al*., (1994) based the design of LCO1490 and HCO2198 on the sequences of a number of species across a range of taxa (insects, molluscs, crustaceans, amphibians and mammals). The primers worked consistently for a broad range of taxa in a single PCR so their failure for these beetles is uncharacteristic. The problem may lie in the reagents used in the PCR recipe. Derycke *et al*. (2010) had greater PCR success using TopTaq DNA polymerase (Qiagen) over Taq DNA polymerase (Qiagen) and DyNAzyme DNA polymerase (New England Biolabs) so the Master Mix (Promega) used in this study may not be as high-yielding as another one could be. DreamTaq (Fermentas) was used in four additional trogid PCRs but was also unsuccessful. However, the DNA barcoding protocol does not specify the need for a particular brand or type of DNA polymerase so it may be worth investigating a range to either determine an optimal type or rule out the need for one. A significant amount of protocol optimising was required to yield successful PCRs, despite the nominal standardisation of the technique.

At present, barcode libraries are constructed of sequences from recently collected specimens or those that have had their DNA preserved by freezing, stored in ethanol or other appropriate forms of preservation (Hajibabaei *et al*., 2006). In terms of forensic application, the use of freshly killed and correctly preserved tissue appeared to be important for PCR success. This is not always possible at a crime scene and is another counter-argument for using DNA as an identifying technique with the chance of a misidentification, albeit small. From experience gained here, specimens need to be killed in >95% ethanol and preferably amplified within about six months of collection. Situations may arise where the person responsible for

collecting samples is untrained in the required conditions and freezes the specimens or allows them to rot or dry out. The latter is unlikely to occur in forensic casework, but may need to be taken into consideration. The availability of good quality specimens for initial DNA extraction seems to be a limiting factor in this kind of study. For research purposes, this may be a problem as specimens may not have been preserved in such a way as to prevent DNA from degrading. Although PCR allows for some degradation, it still has limitations to the extent of damage for it to still work. Practically speaking, forensic entomologists would have difficulty getting usable sequences from old or unsuitably preserved specimens, which may often be the case when required to give an expert opinion based on samples with degraded DNA.

The barcoding technique bills itself as being quick and precise. However, achieving the ideal conditions for successful amplification of DNA may be extremely time-consuming and expensive. In these cases, identifications done based on morphology could be quicker and cheaper.

3.4.2 Juvenile Stages

Juvenile specimens of *D. maculatus, D. haemorrhoidalis* and *T. micans* were correctly identified as they clustered with their respective adults and their identifications were later confirmed by a taxonomist with relevant expertise. It is possible that the DNA of the *N. rufipes* larvae was too degraded to amplify, since the adults of this family (Cleridae) would amplify readily. Fresh eggs contain one cell's worth of genetic material, so the extractions may have been too diluted for amplification to occur. To circumvent this problem as a forensic entomologist, one could use several fresh eggs (if available and obviously from one parent) in one extraction to increase the chance of amplification. Alternatively, direct PCR could also be investigated.

3.4.3 Paralogy

The 67 bp gap in the *N. rufipes* sequences is characteristic of an unexpressed nuclear copy of COI being amplified instead of the mitochondrial copy since COI is a coding region and should have no deletions, let alone a frame-shift deletion. A possible explanation for the gap is a mutation-containing primer binding site for one or both of the universal primers that occurs in the mitochondrial but not the nuclear sequence. This would cause amplification of only the NUMT. Considering the abundance of mtDNA, chances are high that the mitochondrial sequence would be amplified, but these sequences are unlikely to have a mitochondrial origin because of the deletion. The barcoding primers (LCO1490 and HCO2198)

may bind preferentially to the nuclear pseudogene in this species, so a different primer combination may resolve this problem. Moulton *et al*. (2010) suggest increasing primer specificity (i.e. species-specific primers) to reduce this problem, but they found that not all NUMTs have deletions within their sequences (as was the case with the sequences of *N. rufipes*) so more stringent quality control measure will be have to be implemented before using the NUMT sequences as barcodes.

In terms of barcoding, *N. rufipes* is not successful as a result of the repeated amplification of the pseudogene with the barcoding primers. However, as it still clusters with its sister species, *N. ruficollis*, on the NJ and ML trees it appears that losing 10% of its barcode does not necessarily detract from its taxonomic utility. It is a factor to consider when using barcodes to identify this species, especially in a forensic setting. A gap of this size and position may be diagnostic of this species, especially since the specimens were readily amplifiable under a broad range of PCR conditions.

The first metazoan pseudogene was found in a grasshopper (*Locusta migratoria*) and Lopez *et al*. (1997) subsequently found that nearly half of the domestic cat's mitochondrial genome was duplicated in the nuclear genome. Since then over 82 eukaryotes have been found to have pseudogenes, or nuclear copies of mtDNA (NUMTs) (Bensasson *et al*., 2001; Song *et al*., 2008).

3.4.4 Misidentifications

Three specimens were assigned to other families in the phylogenetic analyses, giving this dataset a 1.3% failure rate. This raises concerns for its application in forensics if there is no match for these sequences on BOLD or Genbank since there is a chance that the barcode may not be an accurate representation of the species.

Discrepancies may arise due to contamination from other samples at any stage of the laboratory process, which is easy to avoid but is a hazard to be aware of. It is also possible that these specimens were mislabelled in the laboratory. When these errors occur, it is not possible to determine the exact cause without re-extracting and reanalysing the DNA. The use of COI barcodes in forensic casework is reliant on the existence of a sound reference database. The phylogenetic analyses performed here are used to assess the similarity of sequences belonging to the same taxa since the objective of this study is to assess the potential of barcoding with these species. Since many of these species are not present in the barcode database, even the forensically common ones, it is difficult to assess the extent of their success. Sonet *et al.* (2013) examined the capacity for the BLAST and BOLD identification systems to accurately identify

forensically important Belgian and French fly species. As with the outcome of the beetle specimens used in the present study, some of the species were identifiable, but others are inadequately represented in the database, posing a problem for genetic identification.

The only species with a 100% match in a public database was *D. haemorrhoidalis*. The best close match to many was "Coleoptera sp" (Table 3.2) or species in the same family, which would be useful only if one was trying to determine higher taxonomic levels of a specimen (fragment or juvenile).

Long branch attraction, an intermittent artefact of most phylogenetic reconstruction methods, may explain *A. rarepunctatus* (Histeridae) clustering with the two *Philonthus* (Staphylinidae) species. However, it would be more likely for this to be the case if the *A. rarepunctatus* was more closely related to the *Philonthus* specimens as individuals are more likely to hybridise with sister genera or at least within their family. It is possible for the anomalies in the trees to be a result of cross-sample contamination, mislabelling or misidentification in spite of every care taken to prevent these. Because of its maternal inheritance and haploid nature, mtDNA has an effective population size that is four times smaller than that of nDNA. This leads to faster lineage sorting (Raupach *et al*., 2010). It is only possible to detect the discrepancies of mtDNA if compared to a male lineage, for example the Y chromosome.

Bacterial symbionts were not a factor in this study as each sequence was similar to members of the Coleoptera and extractions were performed only on legs to prevent this problem and to save the rest of the insect as a voucher specimen.

3.4.5 Success of Species Identification

Individuals belonging to the same family are generally closely grouped on the ML tree (Figure 3.11) showing that there is less variation between their sequences than there is between them and other families. A species divergence threshold approach assumes that intraspecific variation does not exceed a certain value (usually 2.5% different [Hebert *et al*., 2003b]). This has been relied on to discriminate species because it does not require knowledge of the species population structure or phylogenetic relationships and is faster in terms of making identifications. However, it discredits character-based information which is needed if identifications are to take an integrative approach (DeSalle, 2006). The frequency of each variation value is shown to determine whether a significant barcoding gap exists (Figure 3.10). The 10-fold rule does not apply here as the average interspecific variation is only 6.3 times that of the intraspecific variation. These results do not support the species divergence threshold of

2.5% nor do they conform to the 10-fold rule illustrated by the lack of a gap between intra-and interspecific frequencies (Figure 3.10).

Scarabs and staphylinids are thought to be relatively closely related (Caterino *et al.*, 2006; Hunt *et al*., 2007), but are not shown as monophyletic on the tree. This may be due to the inability of COI to resolve ancient phylogenetic relationships. Although the majority of the specimens on the trees fell within their families, there were a few anomalies with the dataset that did not support the principles of DNA barcoding and the possible explanations should be addressed. For instance, the staphylinids do not remain in a cluster together but alternate with the histerids. A probable explanation for this is that mutation saturation could have occurred when all the possible mutations at a single base (being limited to A, C, G and T) have taken place over time. The effect on a phylogenetic tree would be the divergence of lines before fusing again (Henn *et al.*, 2009). The sequences would return to those of previous generations and would provide inaccurate representation.

3.4.6 Geographical Variation

It seems that these beetles share a proportion of their DNA with those from other locations, so their dispersal is obviously not as limited as that of sedentary organisms. The distance networks are largely uniform in the haplotype distribution among localities. No distinct clustering occurs in any of the trees showing that these beetle species' sequences should match with those of any other area in southern Africa. Specimens of *D. maculatus* and *D. haemorrhoidalis s*how a slightly more discrete segregation than those of *S. splendens* or *T. micans*. From this outcome, it should be possible to correctly identify a specimen from anywhere in the country based on its barcode. This reflects well on the barcoding concept since little variation is shown within species (in comparison with intraspecific variation), but is not a useful characteristic for forensic entomologists wanting to find the region of origin of a beetle found on a corpse. These beetles may operate on a temporal scale where the haplotype present on one collecting occasion will be different to one collected at another time. Relatively few beetles are attracted to a corpse and if they continue to breed there would be a relatively small founding set of alleles in the community. Intermediate haplotypes may be a result of the recurrence of single nucleotide polymorphisms in the same positions on the genome throughout many generations. Incomplete lineage sorting may be the case with the *A. erichsoni* and *S. cupreus* specimens. The samples were collected from the same location, Kuruman and Grahamstown respectively, but are divided into two monophyletic groups (Figure 3.11). In the same way, the more numerously barcoded species (*S. splendens*, *T. micans*, *D. maculatus*)

show a degree of dissimilarity. Incomplete lineage sorting could be at play here, but it could also be due to subtle differences in the sequences. Heteroplasmy may also explain the presence of more than one haplotype in an individual, but the same extracts would have to be further tested by repeating the PCR several times to detect another haplotype in a single specimen (Jinbo *et al*., 2011).

The impression given by the haplotype networks is that COI haplotypes are largely not location-specific, although it would be short-sighted to completely rule out the ability of barcodes to determine the region of origin of a specimen. For instance, the G haplotype of *D. maculatus* is only present in the Northern Cape.

The relationships shown in the maximum likelihood tree (Figure 3.11) generally reflect those shown in the neighbour-joining tree (Figure 3.13) so the anomalies present in both are not due to the methods used to construct the phylogenies. For instance, the specimens that do not fall in the expected locations on the trees could have been mislabelled, misidentified or contaminated with the genetic material of another specimen. However, the Staphylinidae are split in the NJ tree, but fall in a single cluster in the ML tree. In both trees the scarabs are separated by the trogids. These families are closely related (Hunt *et al*., 2007), so it is not an entirely odd outcome. The close resemblance of the NJ and ML trees also affirms the NJ method as a method of showing similarity. The ML method has been considered to be a particularly thorough way of estimating a phylogenetic tree and is a likely representation of how the sequences fall on the tree in relation to one another. The neighbour-joining method gives one tree and is the quickest and more haphazard of the methods used to infer phylogeny (Collins & Cruickshank, 2013). For these specimens, either of the two methods could have been used to show relationships and the maximum likelihood method would not need to be relied on as it takes much longer and needs more computational power to get the same outcome. Barcoding protocol requires sequences to be sufficiently divergent to be distinguished using the NJ method.

3.5 Conclusions

Finding a relatively short region of sufficient taxon-wide variability with the ability to discriminate species and the added requirement of conserved flanking regions for universal primers is a demanding task. Using a single marker is optimistic anyway (Will & Rubinoff, 2004; Will *et al*., 2005). As it stands, an ideal marker has not been found and may not even exist (Valentini *et al*., 2009). Further research needing time and resources will have to be

carried out if there is to be a definitive, unequivocal region to use. It is in the best interests of taxonomists to continue implementing an integrative approach when it comes to identifying animals. The future application of COI as a forensic marker, integrating existing gene sequences, appears to be inevitable (Jinbo *et al*., 2011; Nelson *et al*., 2012; Taylor & Harris, 2012). The use of a supplementary nuclear gene for the identification of species has been suggested and it may be the only method of unequivocally identifying forensically significant species (Schilthuizen *et al*., 2011; Boehme *et al*., 2011; Meiklejohn *et al*., 2013). Van Der Bank *et al*. (2013) suggest integrating other molecular techniques altogether (e.g. AFLP, microsatellites) along with other genes.

It is in the best interests of forensic entomologists to have a time-efficient method of identification that does not require high levels of expertise so locating a complementary nuclear gene sequence to validate the use of barcodes would be extremely valuable. However, laboratory costs should decrease dramatically and technology will become faster, which will allow for cheaper sequencing and application of the technique to a broader range of animals (Taylor & Harris, 2012)

It is in the best interests of forensic entomology to develop a completely trustworthy amplification method and develop an integrated identification method protocol. Future work should be to further optimise this protocol to test a range of Taq polymerases outside of Promega MasterMix and DreamTaq to determine whether there is a 'best' kind to use or to analyse the primer binding sites over a number of beetle families to see whether mutations exist that will prevent amplification. Having reliable COI sequences will make barcoding a powerful tool in forensic entomology once there are sufficient reference sequences for comparison.

Table 3.1: Abbreviations of the locations of collection and their corresponding coordinates.

			Coordinates	
		Abbreviation	**Latitude (S)**	**Longitude (E)**
South Africa	Addo Elephant Park	AEP	33° 28' 18" S	25° 44' 48" E
	Blue Canyon Conservancy	BCC	24° 21' 17" S	31° 03' 01" E
	Cradock	CDK	31° 55' 23" S	25° 24' 01" E
	Grahamstown	GHT	33° 16' 17" S	26º 27' 25" E
	Griekwastad	GWD	28° 53' 35" S	23 00' 11" E
	Hondeklipbaai	HKB	30° 18' 37" S	17° 16' 31" E
	Kokstad	KSD	30° 27' 49" S	29° 27' 07" E
	Kuruman	KMN	26° 57' 40" S	21° 50' 59" E
	Mafikeng	MFK	25° 51' 31" S	25° 43' 57" E
	Northern Cape roadside	NCR	28° 53' 35" S	23° 00' 11" E
	Rietvlei Nature Reserve	RNV	25° 53' 02" S	28° 17' 22" E
	Saldanha Bay	SDB	32° 48' 37" S	18° 10' 02" E
	Shamwari Game Reserve	SGR	33° 26' 11" S	26° 04' 27" E
	Williston	WSN	31° 15' 41" S	21° 07' 15" E
Namibia	Mashatu Game Reserve	MGR	22° 14' 44" S	29° 08' 10" E
Zambia	Muckleneuk Farm	MCK	16° 38' 50" S	27° 0' 15" E

Table 3.2: Specimens used in this study, with locations, GenBank accession numbers, Albany Museum catalogue numbers, and best matches found using nBLAST and BOLD. Juvenile specimens are marked with an 'L'.

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Cleridae										
Gyponyx sp.	D29	Table Farm	33°12'58"S	26°16'13"E	87%	KC524711.1	91%	MACOL1756-12.COI-5P	KF956174	AMGS:80766
Necrobia ruficollis	C82	Grahamstown	33°16'17"S	26°27'25"E	88%	JQ344786.1	89%	ETKH542-12.COI-5P	KF956175	AMGS:80958
Necrobia ruficollis	I80	Shamwari	33°16'08"S	26°07'18"E	88%	JQ344786.1	89%	NEONU451-11.COI-5P	KJ140510	AMGS:80785
Necrobia rufipes	I76	Addo	33°28'18"S	25°44'48"E	88%	JQ344786.1	87%	GBDP1984-06.COI-5P	KF956153	AMGS:80769
Necrobia rufipes	I77	Addo	33°28'18"S	25°44'48"E	88%	JQ344786.1	88%	GBCL0862-06.COI-5P	KF956154	AMGS:80770
Necrobia rufipes	D13	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	88%	JQ344786.1	89%	GBCL0862-06.COI-5P	KF956176	AMGS:80771
Necrobia rufipes	D15	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	89%	JQ344786.1	88%	GBCL0862-06.COI-5P	KF956155	AMGS:80772
Necrobia rufipes	D27	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	89%	JQ344786.1	88%	GBCL0862-06.COI-5P	KF956156	AMGS:80773
Necrobia rufipes	D28	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	89%	JQ344786.1	87%	CNBPN570-13.COI-5P	KF956157	AMGS:80774
Necrobia rufipes	C96	Fish River	31°55'23"S	25°24'01"E	90%	JQ344786.1	86%	CNBPN570-13.COI-5P	KF956158	AMGS:80775
Necrobia rufipes	C97	Fish River	31°55'23"S	25°24'01"E	90%	JQ344786.1	86%	CNBPN570-13.COI-5P	KF956159	AMGS:80776
Necrobia rufipes	D1	Fish River	31°55'23"S	25°24'01"E	89%	JQ344786.1	85%	GBDP1984-06.COI-5P	KF956160	AMGS:80777
Necrobia rufipes	C36	Grahamstown	33°16'17"S	26°27'25"E	86%	JQ344786.1	91%	GBCL0862-06.COI-5P	KF956161	AMGS:80956
Necrobia rufipes	D83	Grahamstown	33°16'17"S	26°27'25"E	87%	JQ344786.1	98%	GBMIN40358-13.COI-5P	KF956162	AMGS:80957
Necrobia rufipes	D81	Griekwastad	28°53'35"S	23°00'11"E	87%	JQ344786.1	90%	GBCL0862-06.COI-5P	KF956163	AMGS:80780
Necrobia rufipes	D82	Griekwastad	28°53'35"S	23°00'11"E	87%	JQ344786.1	85%	GBCL0862-06.COI-5P	KF956164	AMGS:80781
Necrobia rufipes	C78	Shamwari	33°16'08"S	26°07'18"E	89%	JQ344786.1	96%	GBCL0862-06.COI-5P	KF956165	AMGS:80786
Necrobia rufipes	I79	Shamwari	33°16'08"S	26°07'18"E	88%	JQ344786.1	98%	GBMIN26225-13.COI-5P	KF956166	AMGS:80787
Necrobia rufipes	I81	Shamwari	33°16'08"S	26°07'18"E	87%	JQ344786.1	86%	GBDP4815-08.COI-5P	KF956167	AMGS:80788
Necrobia rufipes	I83	Shamwari	33°16'08"S	26°07'18"E	87%	JQ344786.1	86%	GBDP1984-06.COI-5P	KF956168	AMGS:80789
Necrobia rufipes	I84	Shamwari	33°16'08"S	26°07'18"E	85%	JQ344786.1	88%	GBDP1984-06.COI-5P	KJ140511	AMGS:80790
Necrobia rufipes	I85	Shamwari	33°16'08"S	26°07'18"E	87%	JQ344786.1	97%	GBDP1984-06.COI-5P	KF956169	AMGS:80791
Necrobia rufipes	I86	Shamwari	33°16'08"S	26°07'18"E	86%	JQ344786.1	86%	GBDP1984-06.COI-5P	KJ140512	AMGS:80792
Necrobia rufipes	I87	Shamwari	33°16'08"S	26°07'18"E	86%	JQ344786.1	89%	GBCL0862-06.COI-5P	KF956170	AMGS:80835

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Necrobia rufipes	I88	Shamwari	33°16'08"S	26°07'18"E	88%	JQ344786.1	91%	GBDP1984-06.COI-5P	KF956171	AMGS:80836
Necrobia rufipes	I90	Shamwari	33°16'08"S	26°07'18"E	86%	JQ344786.1	98%	GBCL0862-06.COI-5P	KF956172	AMGS:80837
Necrobia rufipes	I91	Shamwari	33°16'08"S	26°07'18"E	88%	JQ344786.1	92%	GBCL0862-06.COI-5P	KF956173	AMGS:80879
Dermestidae										
Dermestes haemorrhoidalis	B27	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	94%	GBMIN12719-13.COI-5P	KF956177	AMGS:80793
Dermestes haemorrhoidalis	B28	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	97%	GBMIN12719-13.COI-5P	KF956178	AMGS:80794
Dermestes haemorrhoidalis	B29	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	90%	COLEO067-10.COI-5P	KF956179	AMGS:80795
Dermestes haemorrhoidalis	B30	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	96%	GBMIN12719-13.COI-5P	KF956180	AMGS:80796
Dermestes haemorrhoidalis	B31	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	97%	COLEO067-10.COI-5P	KF956181	AMGS:80797
Dermestes haemorrhoidalis	B32	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	91%	COLEO067-10.COI-5P	KF956182	AMGS:80798
Dermestes haemorrhoidalis	B33	Culture	33°18'34"S	26°31'11"E	100%	KC407716.1	92%	COLEO067-10.COI-5P	KF956183	AMGS:80799
Dermestes haemorrhoidalis	D31	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	98%	COLEO067-10.COI-5P	KF956184	AMGS:80800
Dermestes haemorrhoidalis	D32	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	99%	COLEO067-10.COI-5P	KF956185	AMGS:80801
Dermestes haemorrhoidalis	D33	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	84%	COLEO067-10.COI-5P	KF956186	AMGS:80802
Dermestes haemorrhoidalis	D34	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	96%	COLEO067-10.COI-5P	KF956187	AMGS:80803
Dermestes haemorrhoidalis	D35	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	97%	GBMIN12719-13.COI-5P	KF956188	AMGS:80804
Dermestes haemorrhoidalis L	D84	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	95%	COLEO067-10.COI-5P	KF956189	AMGS:80805
Dermestes haemorrhoidalis L	D85	Grahamstown	33°16'17"S	26°27'25"E	100%	KC407716.1	85%	GBMIN12719-13.COI-5P	KF956190	AMGS:80806
Dermestes maculatus L	Larva 2	Grahamstown	33°16'17"S	26°27'25"E	88%	FJ819672.1	95%	USCOL048-09.COI-5P	KF956191	AMGS:80856
Dermestes maculatus L	Larva 3	Grahamstown	33°16'17"S	26°27'25"E	88%	FJ819672.1	94%	USCOL048-09.COI-5P	KF956192	AMGS:80857
Dermestes maculatus	C95	Fish River	31°55'23"S	25°24'01"E	88%	FJ819672.1	95%	USCOL048-09.COI-5P	KF956193	AMGS:80807
Dermestes maculatus	D79	Griekwastad	28°53'35"S	23°00'11"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956194	AMGS:80810
Dermestes maculatus	D5	Mashatu Game Reserve	22°14'44"S	29°08'10"E	88%	FJ819672.1	96%	USCOL048-09.COI-5P	KF956195	AMGS:80811
Dermestes maculatus	C59	Hondeklipbaai	30°18'37"S	17°16'31"E	88%	FJ819672.1	90%	USCOL048-09.COI-5P	KF956196	AMGS:80812
Dermestes maculatus	D52	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	96%	USCOL048-09.COI-5P	KF956197	AMGS:80813
Dermestes maculatus	D53	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	85%	USCOL048-09.COI-5P	KF956198	AMGS:80814
Dermestes maculatus	D54	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	99%	USCOL048-09.COI-5P	KF956199	AMGS:80815
Dermestes maculatus	D55	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	93%	USCOL048-09.COI-5P	KF956200	AMGS:80816

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Dermestes maculatus	D56	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	86%	USCOL048-09.COI-5P	KF956201	AMGS:80817
Dermestes maculatus	D57	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	96%	USCOL048-09.COI-5P	KF956202	AMGS:80818
Dermestes maculatus	D58	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	98%	USCOL048-09.COI-5P	KF956203	AMGS:80819
Dermestes maculatus	D59	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956204	AMGS:80820
Dermestes maculatus	D60	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956205	AMGS:80821
Dermestes maculatus	D61	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	91%	USCOL048-09.COI-5P	KF956206	AMGS:80822
Dermestes maculatus	D69	Kuruman	26°57'40"S	21°50'59"E	88%	FJ819672.1	88%	USCOL048-09.COI-5P	KF956207	AMGS:80823
Dermestes maculatus	D2	Mashatu Game Reserve	21°28'37"S	28°15'36"E	88%	FJ819672.1	98%	USCOL048-09.COI-5P	KF956208	AMGS:80824
Dermestes maculatus	D3	Mashatu Game Reserve	21°28'37"S	28°15'36"E	88%	FJ819672.1	91%	USCOL048-09.COI-5P	KF956209	AMGS:80825
Dermestes maculatus	1ai	Rietvlei Nature Reserve	25°53'02"S	28°17'22"E	88%	FJ819672.1	92%	USCOL048-09.COI-5P	KF956210	AMGS:80826
Dermestes maculatus	Larva 1	Rietvlei Nature Reserve	25°53'02"S	28°17'22"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956211	AMGS:80827
Dermestes maculatus	Larva 4	Rietvlei Nature Reserve	25°53'02"S	28°17'22"E	88%	FJ819672.1	99%	USCOL048-09.COI-5P	KF956212	AMGS:80828
Dermestes maculatus	C60	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956213	AMGS:80829
Dermestes maculatus	C62	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	96%	USCOL048-09.COI-5P	KF956214	AMGS:80830
Dermestes maculatus	C65	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	91%	USCOL048-09.COI-5P	KF956215	AMGS:80831
Dermestes maculatus	C66	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	91%	USCOL048-09.COI-5P	KF956216	AMGS:80832
Dermestes maculatus	C68	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	88%	USCOL048-09.COI-5P	KF956217	AMGS:80833
Dermestes maculatus	C69	Saldanha Bay	32°48'37"S	18°10'02"E	88%	FJ819672.1	99%	USCOL048-09.COI-5P	KF956218	AMGS:80834
Dermestes maculatus	1aii	Shamwari	33°16'08"S	26°07'18"E	88%	FJ819672.1	84%	USCOL048-09.COI-5P	KF956219	AMGS:80768
Dermestes maculatus	Larva 5	Shamwari	33°16'08"S	26°07'18"E	88%	FJ819672.1	88%	USCOL048-09.COI-5P	KF956220	AMGS:80782
Dermestes maculatus	Larva 6	Shamwari	33°16'08"S	26°07'18"E	88%	FJ819672.1	95%	USCOL048-09.COI-5P	KF956221	AMGS:80783
Dermestes maculatus	D70	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	91%	USCOL048-09.COI-5P	KF956222	AMGS:80838
Dermestes maculatus	D71	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	98%	USCOL048-09.COI-5P	KF956223	AMGS:80839
Dermestes maculatus	D72	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	95%	USCOL048-09.COI-5P	KF956224	AMGS:80840
Dermestes maculatus	D73	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	93%	USCOL048-09.COI-5P	KF956225	AMGS:80841
Dermestes maculatus	D74	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	89%	USCOL048-09.COI-5P	KF956226	AMGS:80842
Dermestes maculatus	D75	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	87%	USCOL048-09.COI-5P	KF956227	AMGS:80843
Dermestes maculatus	D76	Williston	31°15'41"S	21°07'15"E	88%	FJ819672.1	86%	USCOL048-09.COI-5P	KF956228	AMGS:80844

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Silphidae										
Thanatophilus micans	D11	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	92%	GBCL15306-13.COI-5P	KF956229	AMGS:80845
Thanatophilus micans	D21	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	95%	GBCL15306-13.COI-5P	KF956230	AMGS:80846
Thanatophilus micans	D22	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	93%	GBCL15306-13.COI-5P	KF956231	AMGS:80847
Thanatophilus micans	D23	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	89%	GBCL15306-13.COI-5P	KF956232	AMGS:80848
Thanatophilus micans	D24	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	87%	GBCL15306-13.COI-5P	KF956233	AMGS:80849
Thanatophilus micans	D8	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	96%	GBCL15306-13.COI-5P	KF956234	AMGS:80850
Thanatophilus micans	D9	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	90%	KC510121.1	86%	GBCL15306-13.COI-5P	KF956235	AMGS:80851
Thanatophilus micans	A1	Grahamstown	33°16'17"S	26°27'25"E	90%	KC510121.1	91%	GBCL15306-13.COI-5P	KF956236	AMGS:80962
Thanatophilus micans	A2	Grahamstown	33°16'17"S	26°27'25"E	90%	KC510121.1	88%	GBCL15306-13.COI-5P	KF956237	AMGS:80963
Thanatophilus micans	D36	Grahamstown	33°16'17"S	26°27'25"E	89%	KC510121.1	84%	GBCL15306-13.COI-5P	KF956238	AMGS:80964
Thanatophilus micans	D37	Grahamstown	33°16'17"S	26°27'25"E	90%	KC510121.1	86%	GBCL15306-13.COI-5P	KF956239	AMGS:80965
Thanatophilus micans	I96	Grahamstown	33°16'17"S	26°27'25"E	90%	KC510121.1	97%	GBCL15306-13.COI-5P	KF956240	AMGS:80966
Thanatophilus micans	I97	Grahamstown	33°16'17"S	26°27'25"E	90%	KC510121.1	85%	GBCL15306-13.COI-5P	KF956241	AMGS:80808
Thanatophilus micans	I93	Grahamstown	33°18'48"S	26°31'07"E	90%	KC510121.1	88%	AUSBC2371-12.COI-5P	KF956242	AMGS:80929
Thanatophilus micans	I94	Grahamstown	33°18'48"S	26°31'07"E	90%	KC510121.1	86%	AUSBC2371-12.COI-5P	KF956243	AMGS:80930
Thanatophilus micans	I95	Grahamstown	33°18'48"S	26°31'07"E	90%	KC510121.1	96%	AUSBC2371-12.COI-5P	KF956244	AMGS:80933
Thanatophilus micans	I99	Grahamstown	33°18'48"S	26°31'07"E	90%	KC510121.1	86%	GBCL15306-13.COI-5P	KF956245	AMGS:80934
Thanatophilus micans	I98	Kokstad	30°27'49"S	29°27'07"E	90%	KC510121.1	97%	GBCL15306-13.COI-5P	KF956246	AMGS:80862
Thanatophilus micans	C1	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	89%	GBCL15306-13.COI-5P	KF956247	AMGS:80863
Thanatophilus micans	C2	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	98%	GBCL15306-13.COI-5P	KF956248	AMGS:80864
Thanatophilus micans	C3	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	95%	GBCL15306-13.COI-5P	KF956249	AMGS:80865
Thanatophilus micans	C4	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	85%	GBCL15306-13.COI-5P	KF956250	AMGS:80866
Thanatophilus micans	C5	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	88%	GBCL15306-13.COI-5P	KF956251	AMGS:80867
Thanatophilus micans	C6	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	96%	AUSBC2371-12.COI-5P	KF956252	AMGS:80868
Thanatophilus micans	C7	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	90%	GBCL15306-13.COI-5P	KF956253	AMGS:80869

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Thanatophilus micans	C52	Muckleneuk	16°38'50"S	27°00'15"E	90%	KC510121.1	95%	GBCL15306-13.COI-5P	KF956254	AMGS:80870
Thanatophilus micans	3a	Shamwari	33°16'08"S	26°07'18"E	91%	KC510121.1	88%	GBCL15306-13.COI-5P	KF956255	AMGS:80871
Thanatophilus micans	3ai	Shamwari	33°16'08"S	26°07'18"E	90%	KC510121.1	87%	GBCL15306-13.COI-5P	KF956256	AMGS:80872
Thanatophilus micans	3aii	Shamwari	33°16'08"S	26°07'18"E	90%	KC510121.1	97%	AUSBC2371-12.COI-5P	KF956257	AMGS:80873
Staphylinidae										
Aleochara sp.	11ai	Shamwari	33°26'11"S	26°04'27"E	87%	GQ980958.1	99%	GBCL0867-06.COI-5P	KF956258	AMGS:80874
Philonthus sp.1	B48	Mafikeng	25°51'31"S	25°43'57"E	87%	GQ980924.1	84%	CNSLU032-13.COI-5P	KF956259	AMGS:80875
Philonthus sp.1	B57	Shamwari	33°26'11"S	26°04'27"E	88%	EU162437.1	99%	GBCL0867-06.COI-5P	KF956260	AMGS:80876
Philonthus sp.1	B60	Shamwari	33°26'11"S	26°04'27"E	86%	KC132739.1	87%	SSWLD2769-13.COI-5P	KF956261	AMGS:80877
Philonthus sp.2	4a	Shamwari	33°26'11"S	26°04'27"E	86%	KC132739.1	86%	SSWLD2769-13.COI-5P	KF956262	AMGS:80878
Philonthus sp.2	11a	Shamwari	33°16'08"S	26°07'18"E	87%	GQ980925.1	85%	SSWLD2769-13.COI-5P	KF956263	AMGS:80784
Philonthus sp.2	B56	Shamwari	33°26'11"S	26°04'27"E	87%	GQ980925.1	92%	BBCCA3402-12.COI-5P	KF956264	AMGS:80880
Platydracus hottentotus	10a	Grahamstown	33°19'09"S	26°30'43"E	86%	GU347040.1	89%	BBCCA3402-12.COI-5P	KF956265	AMGS:80881
Platydracus hottentotus	B46	Shamwari	33°26'11"S	26°04'27"E	87%	GU347040.1	91%	SSWLD2769-13.COI-5P	KF956266	AMGS:80882
Scarabaeidae										
Frankenbergerius forcipatus	I59	Grahamstown	33°19'49"S	26°26'16"E	90%	EU162441.1	87%	GBCLS005-13.COI-5P	JN817504	AMGT:59483
Frankenbergerius forcipatus	I60	Grahamstown	33°19'49"S	26°26'16"E	89%	EU162444.1	98%	GBCLS005-13.COI-5P	JN817512	COLS12360
Frankenbergerius forcipatus	I61	Grahamstown	33°19'49"S	26°26'16"E	89%	EU162444.1	84%	GBCLS005-13.COI-5P	KJ140513	AMGS:80885
Frankenbergerius forcipatus	I62	Grahamstown	33°19'49"S	26°26'16"E	89%	EU162444.1	95%	GBCLS005-13.COI-5P	JN817505	AMGT:59484
Frankenbergerius forcipatus	I63	Grahamstown	33°18'48"S	26°31'07"E	90%	EU162441.1	84%	GBCLS005-13.COI-5P	JN817506	AMGT:59487
Frankenbergerius forcipatus	I64	Grahamstown	33°18'48"S	26°31'07"E	90%	EU162441.1	95%	GBCLS005-13.COI-5P	KJ140514	AMGS:80883
Frankenbergerius forcipatus	I65	Grahamstown	33°19'49"S	26°26'16"E	88%	EU162444.1	97%	GBCLS005-13.COI-5P	JN817513	COLS12360
Frankenbergerius forcipatus	I66	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162444.1	93%	GBCLS005-13.COI-5P	JN817507	COLS12361
Frankenbergerius forcipatus	I67	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162441.1	96%	GBCLS005-13.COI-5P	JN817508	AMGT:59485
Frankenbergerius forcipatus	I68	Grahamstown	33°18'48"S	26°31'07"E	90%	EU162441.1	87%	GBCLS005-13.COI-5P	JN817509	AMGT:59486
Frankenbergerius forcipatus	I69	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162441.1	92%	GBCLS005-13.COI-5P	JN817510	COLS12360
Frankenbergerius forcipatus	I71	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162444.1	89%	GBCLS005-13.COI-5P	KJ140515	AMGS:80884
Frankenbergerius forcipatus	I73	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162441.1	91%	GBCLS005-13.COI-5P	JN817514	COLS12360

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Frankenbergerius forcipatus	I74	Grahamstown	33°19'49"S	26°26'16"E	89%	EU162444.1	99%	GBCLS005-13.COI-5P	JN817515	COLS12361
Frankenbergerius forcipatus	I75	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162444.1	93%	GBCLS005-13.COI-5P	JN817511	COLS12360
Liatongus sp.	D38	Kuruman	26°57'40"S	21°50'59"E	90%	EU162463.1	95%	MACOL2258-12.COI-5P	KF956267	AMGS:80886
Liatongus sp.	D39	Kuruman	26°57'40"S	21°50'59"E	90%	EU162463.1	95%	MACOL2258-12.COI-5P	KF956268	AMGS:80887
Liatongus sp.	D40	Kuruman	26°57'40"S	21°50'59"E	89%	EU162463.1	84%	MACOL357-10.COI-5P	KF956269	AMGS:80888
Onthophagus sp.1	A55	Shamwari	33°26'11"S	26°04'27"E	90%	EU162471.1	87%	MACOL358-10.COI-5P	KF956270	AMGS:80889
Onthophagus sp.1	A57	Shamwari	33°26'11"S	26°04'27"E	90%	EU162471.1	98%	MACOL358-10.COI-5P	KF956271	AMGS:80890
Onthophagus sp.1	A58	Shamwari	33°26'11"S	26°04'27"E	89%	EU162471.1	96%	MACOL358-10.COI-5P	KF956272	AMGS:80891
Onthophagus sp.1	A59	Grahamstown	33°20'13"S	26°40'49"E	90%	EU162471.1	93%	MACOL358-10.COI-5P	KF956273	AMGS:80896
Onthophagus sp.1	B14	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162471.1	89%	GBCL3743-08.COI-5P	KJ140516	AMGS:80900
Onthophagus sp.2	A74	Grahamstown	33°20'13"S	26°40'49"E	89%	EU162460.1	94%	GBCL3743-08.COI-5P	KF956274	AMGS:80897
Onthophagus sp.2	B18	Grahamstown	33°18'48"S	26°31'07"E	89%	EU162460.1	92%	COAS1483-12.COI-5P	KF956275	AMGS:80926
Onthophagus sp.3	A60	Grahamstown	33°20'13"S	26°40'49"E	89%	EU162471.1	86%	COAS1483-12.COI-5P	KF956276	AMGS:80931
Onthophagus sp.3	A61	Grahamstown	33°20'13"S	26°40'49"E	89%	EU162458.1	90%	GBDP13875-13.COI-5P	KF956277	AMGS:80932
Sarophorus tuberculatus	C9	Grahamstown	33°18'48"S	26°31'07"E	84%	KC132820.1	85%	GBCL4041-09.COI-5P	KF956278	AMGS:80901
Sarophorus tuberculatus	C10	Grahamstown	33°18'48"S	26°31'07"E	85%	JX913737.1	96%	GBCL0859-06.COI-5P	KF956279	AMGS:80893
Sarophorus tuberculatus	C11	Grahamstown	33°18'48"S	26°31'07"E	88%	JN817504.1	99%	SAFIN317-12.COI-5P	KF956280	AMGS:80895
Trogidae										
Trox sulcatus	A95	Grahamstown	33°18'48"S	26°31'07"E	86%	EU162436.1	90%	MHFLI432-07.COI-5P	KF956281	AMGS:80899
Trox fasicularis	B1	Grahamstown	33°18'48"S	26°31'07"E	86%	KC499890.1	94%	MHFLI432-07.COI-5P	KF956282	AMGS:80898
Histeridae										
Atholus erichsoni	D42	Kuruman	26°57'40"S	21°50'59"E	83%	JN581900.1	99%	COLEO067-10.COI-5P	KF956283	AMGS:80903
Atholus erichsoni	D43	Kuruman	26°57'40"S	21°50'59"E	83%	JN581900.1	95%	COLEO067-10.COI-5P	KF956284	AMGS:80904
Atholus erichsoni	D44	Kuruman	26°57'40"S	21°50'59"E	83%	JN581900.1	93%	COLEO067-10.COI-5P	KF956285	AMGS:80905
Atholus erichsoni	D45	Kuruman	26°57'40"S	21°50'59"E	82%	GU013589.1	90%	SSWLD3438-13.COI-5P	KF956286	AMGS:80906
Atribalus rarepunctatus	I57	Grahamstown	33°18'48"S	26°31'07"E	85%	AY847540.1	97%	MACOL2461-12.COI-5P	KF956287	AMGS:80939
Hister nomas	C77	Grahamstown	33°19'09"S	26°30'43"E	82%	FJ819855.1	97%	GBCL0902-06.COI-5P	KF956288	AMGS:80908
Hister nomas	I33	Grahamstown	33°18'48"S	26°31'07"E	81%	FJ819855.1	98%	MACOL2461-12.COI-5P	KF956289	AMGS:80942

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Hister nomas	I49	Grahamstown	33°18'48"S	26°31'07"E	82%	FJ819855.1	97%	BBCCA4238-12.COI-5P	KF956290	AMGS:80943
Histeridae sp.	D77	Northern Cape roadside	28°53'35"S	23°00'11"E	83%	KC440158.1	98%	GBCL15230-13.COI-5P	KF956291	AMGS:80911
Histeridae sp.	D80	Northern Cape roadside	28°53'35"S	23°00'11"E	85%	FN263049.1	95%	COLEO067-10.COI-5P	KF956292	AMGS:80912
Pelorurus sp.	A46	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	84%	GU347192.1	99%	COLEO067-10.COI-5P	KF956293	AMGS:80914
Pelorurus sp.	C72	Hondeklipbaai	30°18'37"S	17°16'31"E	83%	GU347192.1	95%	COLEO067-10.COI-5P	KF956294	AMGS:80915
Pelorurus sp.	D47	Kuruman	26°57'40"S	21°50'59"E	84%	GU347192.1	85%	COLEO067-10.COI-5P	KF956295	AMGS:80916
Pelorurus sp.	D4	Mashatu Game Reserve	22°14'44"S	29°08'10"E	84%	GU347192.1	96%	COLEO067-10.COI-5P	KF956296	AMGS:80917
Pelorurus sp.	A50	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	85%	GQ980910.1	88%	COLEO067-10.COI-5P	KF956297	AMGS:80918
Pelorurus sp.	A53	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	84%	GQ980892.1	99%	GBCL15888-13.COI-5P	KF956298	AMGS:80919
Pelorurus sp.	A54	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	84%	GU347192.1	90%	GBCL15888-13.COI-5P	KF956299	AMGS:80920
Pelorurus sp.	9a	Shamwari	33°26'11"S	26°04'27"E	84%	GU347192.1	95%	COLEO067-10.COI-5P	KF956300	AMGS:80921
Pelorurus sp.	D66	Williston	31°15'41"S	21°07'15"E	84%	HM803547.1	99%	GBCL15888-13.COI-5P	KF956301	AMGS:80922
Saprinus sp.	I46	Shamwari	33°26'11"S	26°04'27"E	83%	KC132818.1	94%	GBCL15888-13.COI-5P	KF956302	AMGS:80923
Saprinus sp.	I47	Shamwari	33°26'11"S	26°04'27"E	83%	KC132818.1	89%	GBCL15888-13.COI-5P	KJ140517	AMGS:80924
Saprinus sp.	A38	Grahamstown	33°19'09"S	26°30'43"E	83%	KC132818.1	91%	ASNOR3315-12.COI-5P	KJ140518	AMGS:80925
Saprinus bicolor	6a	Grahamstown	33°18'48"S	26°31'07"E	84%	HM803626.1	85%	COLEO067-10.COI-5P	KJ140519	AMGS:80961
Saprinus bicolor	6ai	Grahamstown	33°18'48"S	26°31'07"E	84%	HQ978627.1	84%	GBCL0862-06.COI-5P	KF956303	AMGS:80967
Saprinus bicolor	I15	Grahamstown	33°18'48"S	26°31'07"E	84%	HM803626.1	99%	GBCL0862-06.COI-5P	KF956304	AMGS:80858
Saprinus bicolor	I24	Grahamstown	33°18'48"S	26°31'07"E	84%	HM803626.1	93%	GBCL0862-06.COI-5P	KF956305	AMGS:80859
Saprinus bicolor	I27	Grahamstown	33°18'48"S	26°31'07"E	84%	HM803626.1	92%	COLEO067-10.COI-5P	KF956306	AMGS:80860
Saprinus bicolor	A47	Grahamstown	33°20'13"S	26°40'49"E	84%	HQ978627.1	89%	GBCL0862-06.COI-5P	KF956307	AMGS:80892
Saprinus bicolor	A48	Grahamstown	33°20'13"S	26°40'49"E	84%	HQ978627.1	89%	GBCL0862-06.COI-5P	KF956308	AMGS:80894
Saprinus bicolor	I50	Grahamstown	33°18'48"S	26°31'07"E	84%	HQ978627.1	97%	GBCL0862-06.COI-5P	KF956309	AMGS:80861
Saprinus bicolor	I58	Grahamstown	33°18'48"S	26°31'07"E	84%	HQ978627.1	85%	COLEO067-10.COI-5P	KF956310	AMGS:80907
Saprinus bicolor	8a	Shamwari	33°26'11"S	26°04'27"E	84%	HM803626.1	89%	GBCL0862-06.COI-5P	KF956311	AMGS:80935
Saprinus bicolor	8ai	Shamwari	33°26'11"S	26°04'27"E	84%	HM803626.1	92%	GBCL0862-06.COI-5P	KF956312	AMGS:80936
Saprinus bicolor	A16	Shamwari	33°26'11"S	26°04'27"E	84%	HM803626.1	88%	COLEO067-10.COI-5P	KF956313	AMGS:80937
Saprinus bicolor	A42	Shamwari	33°26'11"S	26°04'27"E	84%	HM803626.1	90%	COLEO067-10.COI-5P	KF956314	AMGS:80938

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Saprinus cruciatus	6aii	Grahamstown	33°18'48"S	26°31'07"E	83%	GQ980928.1	84%	GBCL0862-06.COI-5P	KF956315	AMGS:80909
Saprinus cruciatus	7a	Shamwari	33°22'59"S	26°02'32"E	83%	GQ980928.1	90%	GBCL4075-09.COI-5P	KF956316	AMGS:80940
Saprinus cruciatus	D41	Kuruman	21°51'31"S	26°57'27"E	83%	GQ980928.1	89%	GBCL4075-09.COI-5P	KF956317	AMGS:80941
Saprinus cupreus	I38	Grahamstown	3318'48"S	26°31'07"E	84%	JX278176.1	96%	GBCL4075-09.COI-5P	KF956318	AMGS:80910
Saprinus cupreus	I39	Grahamstown	3318'48"S	26°31'07"E	84%	JX278176.1	96%	GBCL4075-09.COI-5P	KF956319	AMGS:80902
Saprinus cupreus	I53	Grahamstown	33°19'43"S	26°26'16"E	84%	JX278176.1	91%	GBCL4075-09.COI-5P	KF956320	AMGS:80944
Saprinus cupreus	I56	Grahamstown	33°19'43"S	26°26'16"E	84%	JX278176.1	87%	COLEO067-10.COI-5P	KF956321	AMGS:80945
Saprinus cupreus	D68	Williston	31°15'41"S	21°07'15"E	85%	EU156680.1	96%	BBCCA3979-12.COI-5P	KF956322	AMGS:80913
Saprinus secchii	C70	Hondeklipbaai	30°18'37"S	17°16'31"E	83%	GU176344.1	92%	BBCCA2322-12.COI-5P	KF956323	AMGS:80946
Saprinus secchii	D48	Kuruman	26°57'40"S	21°50'59"E	83%	AB608757.1	97%	BBCCA2322-12.COI-5P	KF956324	AMGS:80947
Saprinus secchii	D49	Kuruman	26°57'40"S	21°50'59"E	83%	AB608757.1	90%	BBCCA2322-12.COI-5P	KF956325	AMGS:80948
Saprinus secchii	D50	Kuruman	26°57'40"S	21°50'59"E	83%	AB608757.1	86%	BBCCA2322-12.COI-5P	KF956326	AMGS:80949
Saprinus secchii	D51	Kuruman	26°57'40"S	21°50'59"E	83%	AB608757.1	99%	BBCCA2322-12.COI-5P	KF956327	AMGS:80950
Saprinus splendens	B50	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	94%	GQ980892.1	95%	COLEO067-10.COI-5P	KF956328	AMGS:80953
Saprinus splendens	B53	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	94%	GQ980892.1	93%	COLEO067-10.COI-5P	KF956329	AMGS:80951
Saprinus splendens	B54	Blue Canyon Conservancy	24°21'17"S	31°03'01"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956330	AMGS:80952
Saprinus splendens	C27	Grahamstown	33°19'09"S	26°30'43"E	94%	GQ980892.1	97%	COLEO067-10.COI-5P	KF956331	AMGS:80954
Saprinus splendens	I12	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	96%	COLEO067-10.COI-5P	KF956332	AMGS:80959
Saprinus splendens	I14	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	99%	COLEO067-10.COI-5P	KF956333	AMGS:80960
Saprinus splendens	I21	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	88%	COLEO067-10.COI-5P	KF956334	AMGS:80852
Saprinus splendens	I22	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	89%	COLEO067-10.COI-5P	KF956335	AMGS:80853
Saprinus splendens	I23	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	95%	COLEO067-10.COI-5P	KF956336	AMGS:80854
Saprinus splendens	I25	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	91%	COLEO067-10.COI-5P	KF956337	AMGS:80855
Saprinus splendens	I26	Grahamstown	33°18'48"S	26°31'07"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956338	AMGS:80927
Saprinus splendens	I29	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956339	AMGS:80809
Saprinus splendens	I34	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	88%	COLEO067-10.COI-5P	KF956340	AMGS:80778
Saprinus splendens	I36	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	93%	COLEO067-10.COI-5P	KF956341	AMGS:80779
Saprinus splendens	I8	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956342	AMGS:80767

Taxon	Lab number	Location	Latitude	Longitude	nBLAST	Accession	BOLD	Sequence ID	Accession number	Albany Museum Voucher
Saprinus splendens	I9	Grahamstown	33°16'17"S	26°27'25"E	94%	GQ980892.1	95%	COLEO067-10.COI-5P	KF956343	AMGS:80955
Saprinus splendens	I54	Grahamstown	33°18'48"S	26°31'07"E	94%	GQ980892.1	96%	COLEO067-10.COI-5P	KF956344	AMGS:80928
Saprinus splendens	5a	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	94%	COLEO067-10.COI-5P	KF956345	AMGS:80968
Saprinus splendens	5ai	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	92%	COLEO067-10.COI-5P	KF956346	AMGS:80969
Saprinus splendens	5aii	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956347	AMGS:80970
Saprinus splendens	9ai	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	91%	COLEO067-10.COI-5P	KF956348	AMGS:80971
Saprinus splendens	9aii	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	89%	COLEO067-10.COI-5P	KF956349	AMGS:80972
Saprinus splendens	A18	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	94%	COLEO067-10.COI-5P	KF956350	AMGS:80973
Saprinus splendens	A40	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	91%	COLEO067-10.COI-5P	KF956351	AMGS:80974
Saprinus splendens	A41	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	97%	COLEO067-10.COI-5P	KF956352	AMGS:80975
Saprinus splendens	I40	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	89%	COLEO067-10.COI-5P	KF956353	AMGS:80976
Saprinus splendens	I41	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	90%	COLEO067-10.COI-5P	KF956354	AMGS:80977
Saprinus splendens	I42	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	91%	COLEO067-10.COI-5P	KF956355	AMGS:80978
Saprinus splendens	I43	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	92%	COLEO067-10.COI-5P	KF956356	AMGS:80979
Saprinus splendens	I44	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	86%	COLEO067-10.COI-5P	KF956357	AMGS:80980
Saprinus splendens	I45	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	88%	COLEO067-10.COI-5P	KF956358	AMGS:80981
Saprinus splendens	B47	Shamwari	33°26'11"S	26°04'27"E	94%	GQ980892.1	89%	COLEO067-10.COI-5P	KF956359	AMGS:80982
Saprinus splendens	D62	Williston	31°15'41"S	21°07'15"E	94%	GQ980892.1	88%	COLEO067-10.COI-5P	KF956360	AMGS:80983
Saprinus splendens	D63	Williston	31°15'41"S	21°07'15"E	94%	GQ980892.1	89%	COLEO067-10.COI-5P	KF956361	AMGS:80984
Saprinus splendens	D64	Williston	31°15'41"S	21°07'15"E	94%	GQ980892.1	86%	COLEO067-10.COI-5P	KF956362	AMGS:80985
Saprinus splendens	D65	Williston	31°15'41"S	21°07'15"E	94%	GQ980892.1	86%	COLEO067-10.COI-5P	KF956363	AMGS:80986
Saprinus splendens	D67	Williston	31°15'41"S	21°07'15"E	94%	GQ980892.1	84%	COLEO067-10.COI-5P	KF956364	AMGS:80987

Table 3.3: Additional sequences of beetles closely related to members of the six families sequenced obtained from GenBank and used in analyses.

Family	Species	Accession number
Histeridae	*Iliotona dorcoides*	GU982707
Histeridae	*Iliotona beyeri*	GU982701
Silphidae	*Nicrophorus orbicollis*	EU271658
Silphidae	*Necrophila americana*	AY165669
Dermestidae	*Dermestes lardarius*	AY165734
Staphylinidae	*Nehemitropia lividipennis*	GQ980892
Staphylinidae	*Atheta celata*	GQ980911
Staphylinidae	*Pella humeralis*	GQ980880
Staphylinidae	*Liogluta microptera*	GQ980937
Staphylinidae	*Drusilla canaliculata*	GQ980874
Staphylinidae	*Thamiarea brittoni*	GQ980962
Trogidae	*Trox unistriatus*	AY165707

Table 3.4: Interspecific pairwise distances for all species in the study calculated using the Kimura 2-parameter model. Pairwise distances between taxa in analysis expressed as a percentage of nucleotide differences (p-distances). [1] *D. haemorrhoidalis*; [2] *D. maculatus*; [3] *Philonthus* sp. 1; [4] *Philonthus* sp. 2; [5] *P. hottentotus*; [6] *T. micans*; [7] *F. forcipatus*; [8] *Liatongus* sp.; [9] *Onthophagus* sp. 1; [10] *Onthophagus* sp. 2; [11] *Onthophagus* sp. 3; [12] *S. tuberculatus*; [13] *A. erichsoni*; [14] *H. nomas*; [15] *Pelorurus* sp.; [16] *Saprinus* sp.; [17] *S. bicolor*; [18] *S. cruciatus*; [19] *S. cupreus*; [20] *S. secchii*; [21] *S. splendens*; [22] *N. ruficollis*; [23] *N. rufipes*.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
[1]																							
[2]	1.35																						
[3]	2.12	1.24																					
[4]	1.64	1.22	1.67																				
[5]	1.69	1.60	2.07	2.20																			
[6]	1.64	1.57	1.63	1.80	1.73																		
[7]	1.95	1.53	1.79	2.12	2.03	1.32																	
[8]	1.58	1.19	1.69	1.76	2.01	1.29	1.25																
[9]	2.21	1.83	2.19	2.57	2.37	1.74	1.40	1.35															
[10]	1.21	1.06	1.48	1.42	1.51	0.81	0.92	0.36	1.15														
[11]	1.38	1.09	1.69	1.59	1.98	1.22	1.23	1.15	1.56	0.81													
[12]	1.54	1.11	1.62	1.65	1.98	1.70	2.09	1.83	2.08	1.43	1.60												
[13]	2.31	1.91	2.46	2.32	2.37	2.30	2.42	2.40	2.42	2.05	1.91	1.94											
[14]	2.11	1.43	1.81	2.03	2.64	2.24	2.19	2.19	2.26	1.95	1.83	1.39	2.01										
[15]	2.36	2.01	2.49	2.47	2.31	2.17	2.56	2.35	2.49	2.15	1.96	1.78	1.52	2.44									
[16]	1.88	1.66	1.85	2.09	2.39	1.90	2.30	2.06	2.16	1.71	1.71	1.31	2.08	1.21	2.41								
[17]	2.25	1.85	1.99	1.99	2.26	2.42	2.71	2.06	2.69	1.86	1.91	1.44	2.50	1.78	2.47	1.93							
[18]	1.87	1.60	2.20	2.34	2.22	1.89	2.20	2.15	2.43	1.64	2.06	1.19	1.97	1.65	2.19	1.70	2.27						
[19]	1.98	1.61	2.11	2.11	2.65	2.23	2.64	2.19	2.58	1.87	1.99	1.10	2.33	1.72	2.25	1.69	1.91	1.80					
[20]	1.99	1.39	1.97	2.13	2.44	2.29	2.61	2.03	2.53	1.95	2.09	1.15	2.47	1.55	2.44	1.67	1.59	1.74	1.44				
[21]	1.79	1.92	2.44	2.49	2.37	2.02	1.81	1.96	2.46	1.54	1.79	2.26	2.99	2.54	2.90	2.83	2.78	2.62	2.78	2.79			
[22]	1.92	1.39	1.94	1.88	2.02	1.82	1.52	1.37	2.10	1.08	1.58	1.93	2.51	2.27	2.62	2.44	2.22	2.44	2.33	2.56	1.67		
[23]	0.88	0.71	1.05	0.95	1.32	0.83	1.20	0.91	1.14	0.38	0.72	0.48	1.39	0.91	1.24	0.87	1.16	0.60	0.80	1.00	1.67	1.11	

Table 3.5: Intraspecific Kimura 2-parameter distances for species with two or more sequences.

Family	Species	Distance (K2P)	No. sequences
Dermestidae	*D. maculatus*	0.092	34
Dermestidae	*D. haemorrhoidalis*	0.204	14
Staphylinidae	*Philonthus* sp. 1	0.312	2
Staphylinidae	*Philonthus* sp. 2	0.275	2
Staphylinidae	*P. hottentotus*	0.121	2
Staphylinidae	*T. micans*	0.369	28
Scarabaeidae	*F. forcipatus*	0.112	15
Scarabaeidae	*S. tuberculatus*	1.492	3
Scarabaeidae	*Liatongus* sp.	0.217	3
Scarabaeidae	*Onthophagus* sp. 1	0.545	5
Scarabaeidae	*Onthophagus* sp. 2	0.318	2
Scarabaeidae	*Onthophagus* sp. 3	0.018	2
Histeridae	*Pelorurus* sp.	0.386	9
Histeridae	*Saprinus* sp.	0.265	3
Histeridae	*S. secchii*	0.276	5
Histeridae	*A. erichsoni*	1.104	4
Histeridae	*H. nomas*	0.087	3
Histeridae	*S. bicolor*	0.022	13
Histeridae	*S. cupreus*	0.397	4
Histeridae	*S. cruciatus*	0.136	3
Histeridae	*S. splendens*	0.137	37
Cleridae	*N. ruficollis*	0.119	2
Cleridae	*N. rufipes*	0.489	24

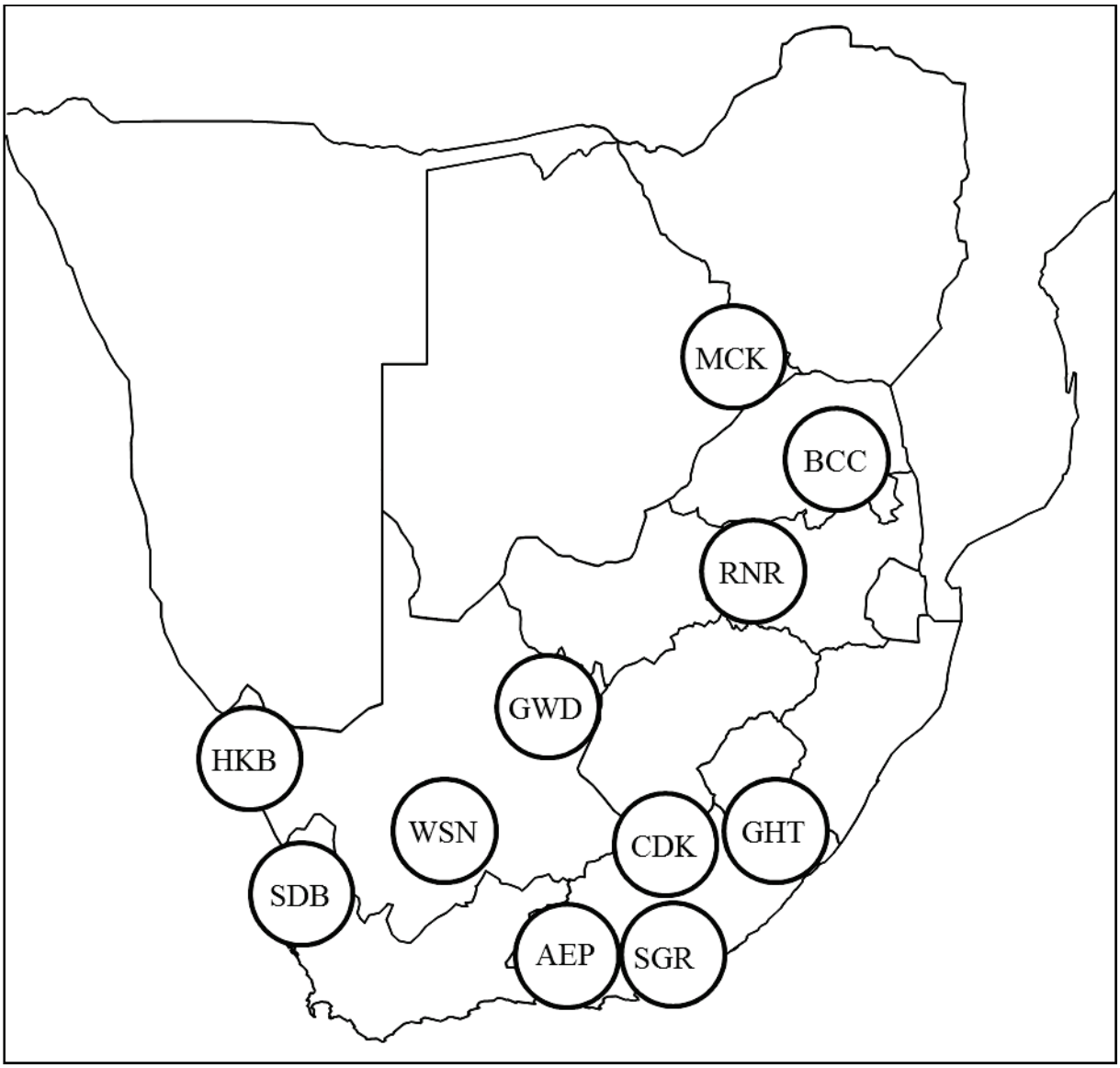

Figure 3.1: Map of southern Africa with locations of collection sites indicated. The locality abbreviations are explained in Table 3.1. Species with more than ten sequences were used in Figures 3.4 to 3.9 analyses.

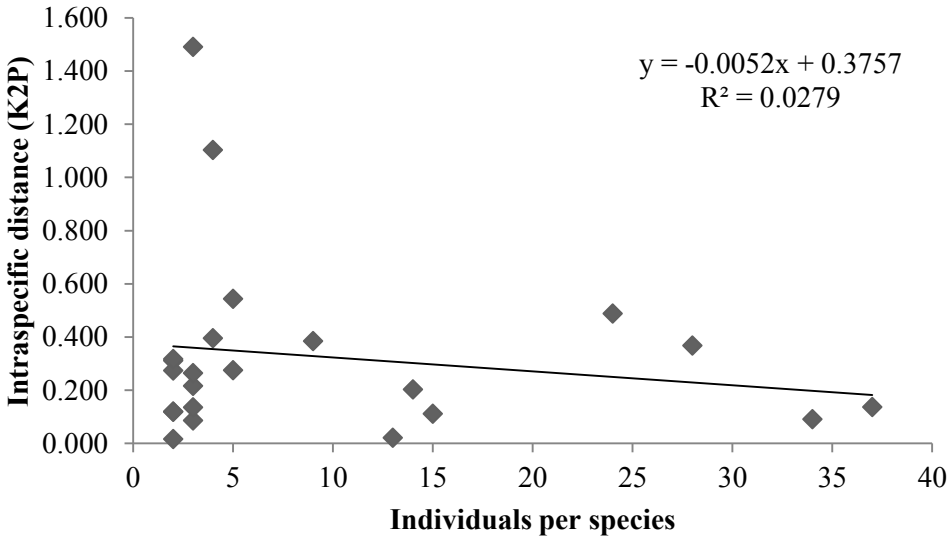

Figure 3.2: Effect of sample size on intraspecific genetic distance, based on analysis of ~570 bp fragments of the COI gene for 23 beetle species of forensic significance.

Haplotype	Colour
A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	
M	
N	
O	
P	
Q	
R	
S	
T	
U	
V	

Figure 3.3: Haplotypes are indicated by colour on pie charts. Colours correspond to the various haplotypes present in individual species on distribution maps. Species with more than ten sequences were used in Figures 3.4 to 3.9 analyses.

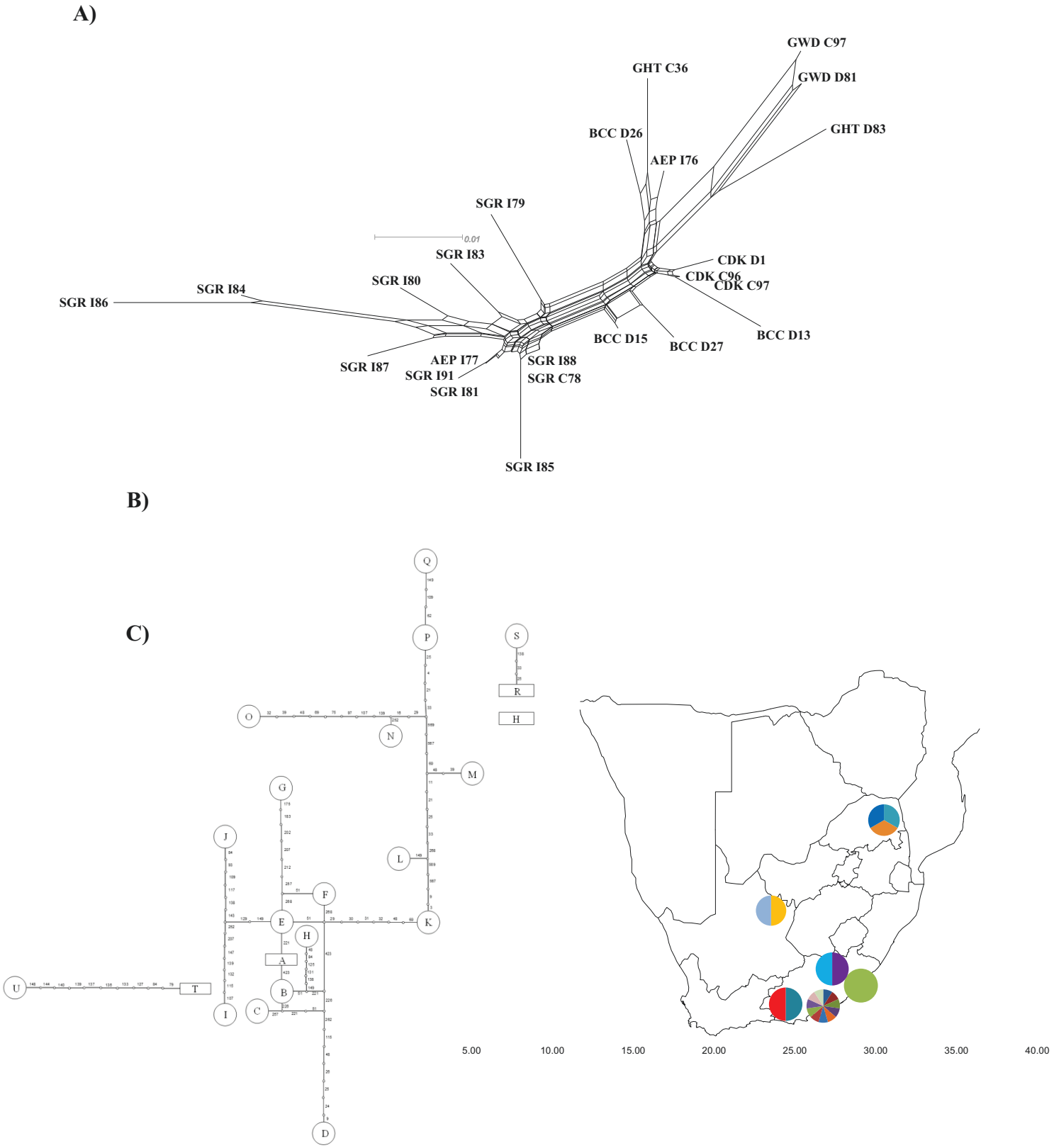

Figure 3.4: A) NeighbourNet tree, B) haplotype network and C) distribution map for *N. rufipes*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

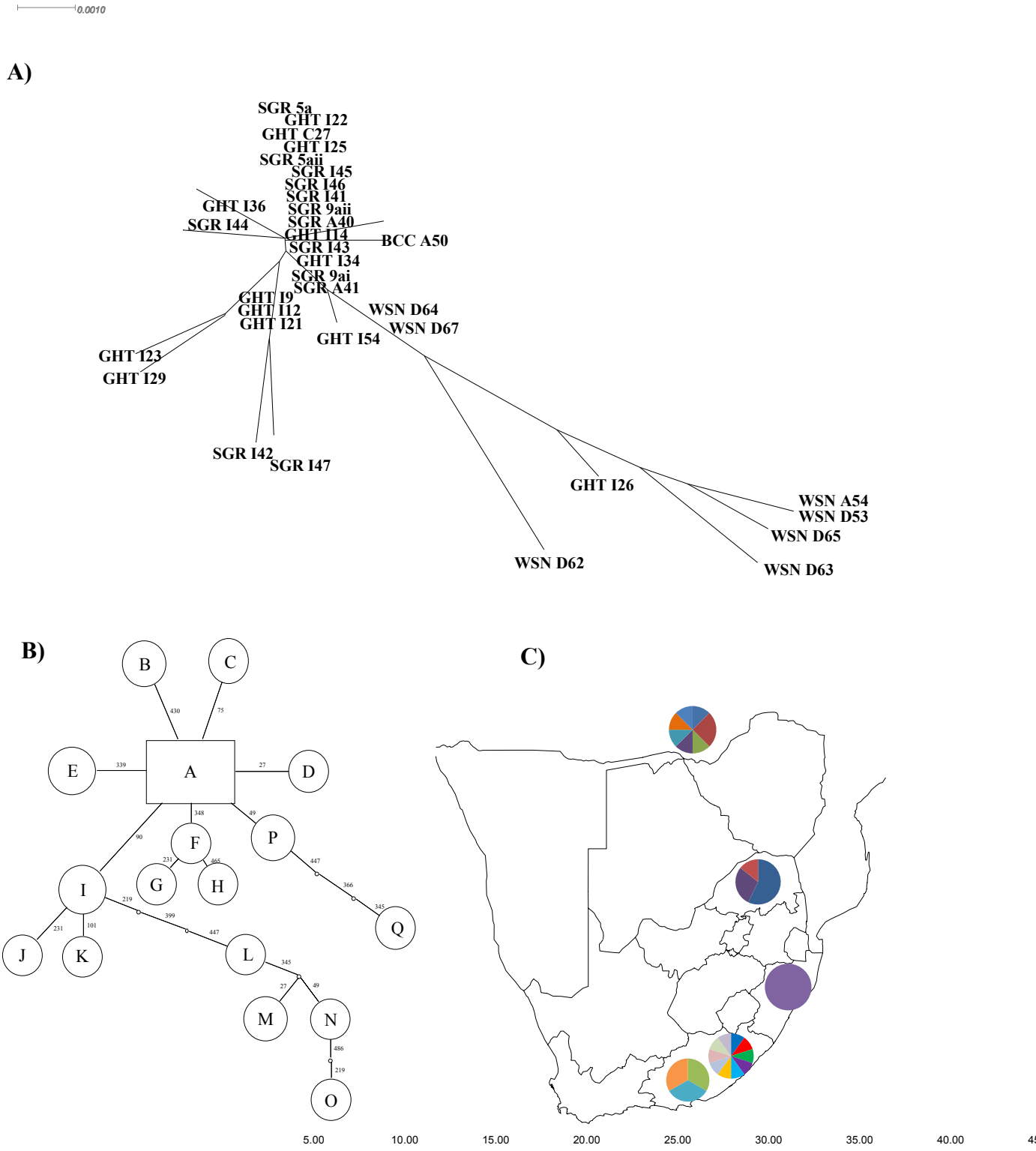

Figure 3.5: A) NeighbourNet tree, B) haplotype network and C) distribution map for *S. splendens*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

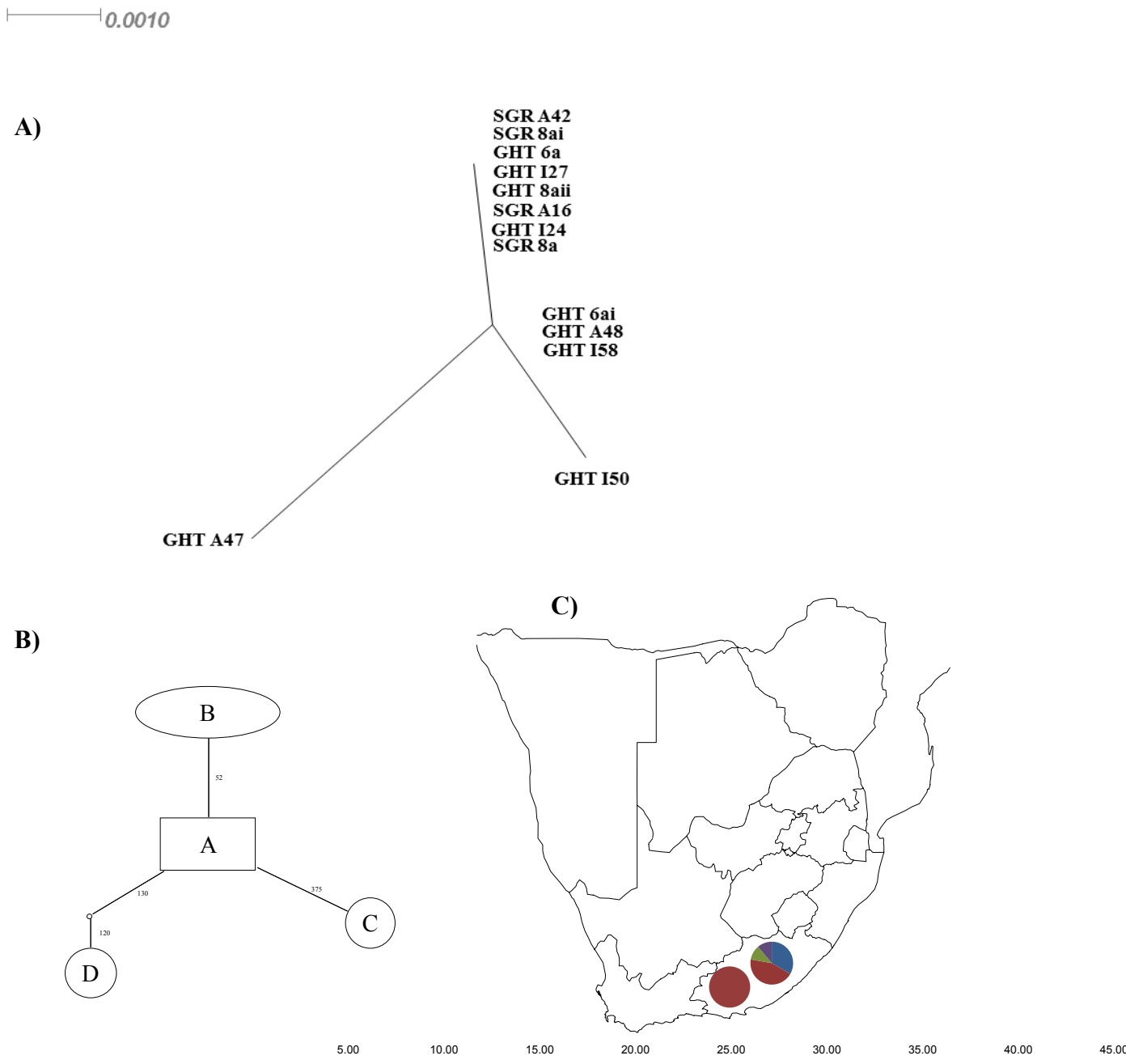

Figure 3.6: A) NeighbourNet tree, B) haplotype network and C) distribution map for *S. bicolor*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

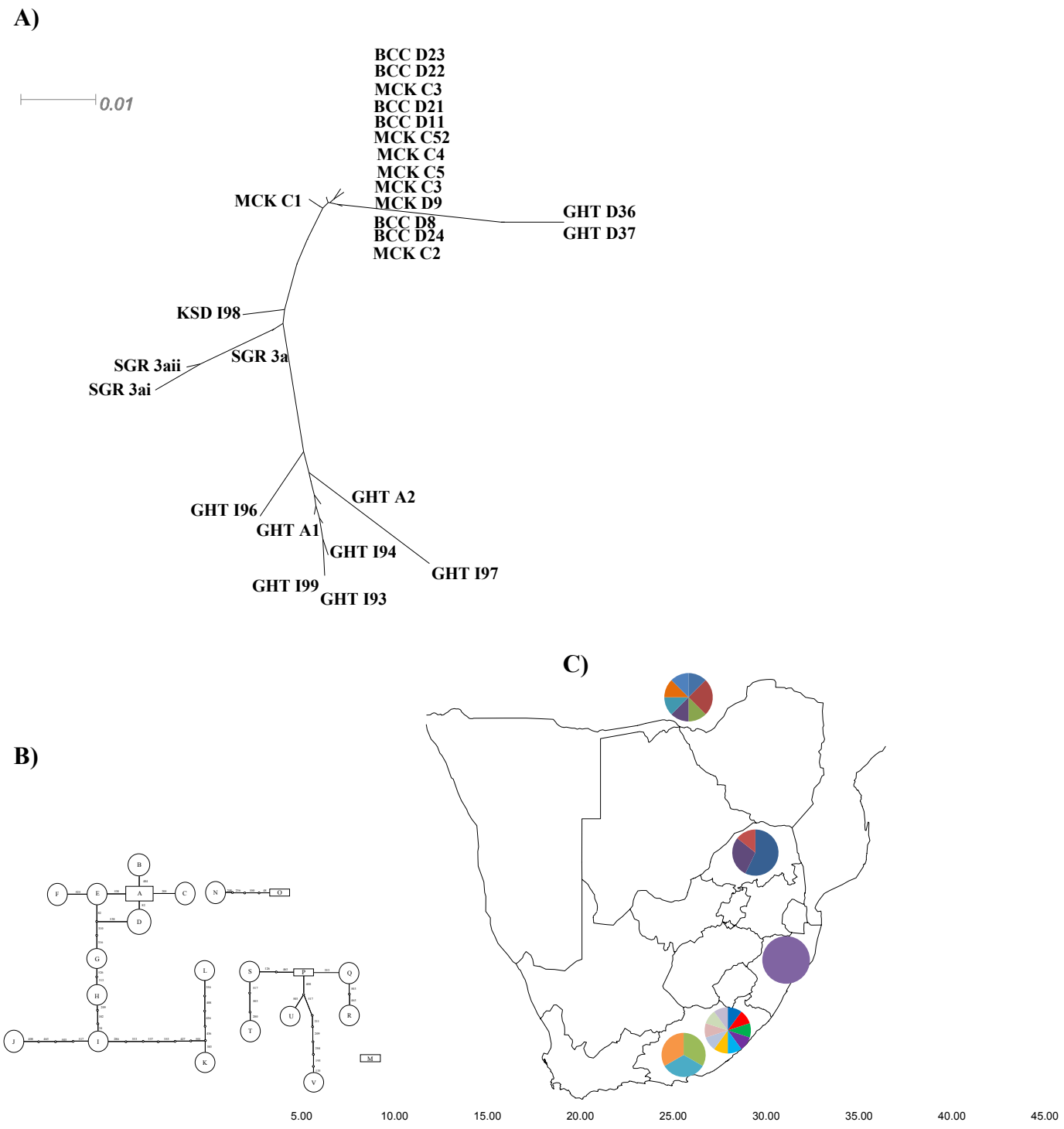

Figure 3.7: A) NeighbourNet tree, B) haplotype network and C) distribution map for *T. micans*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

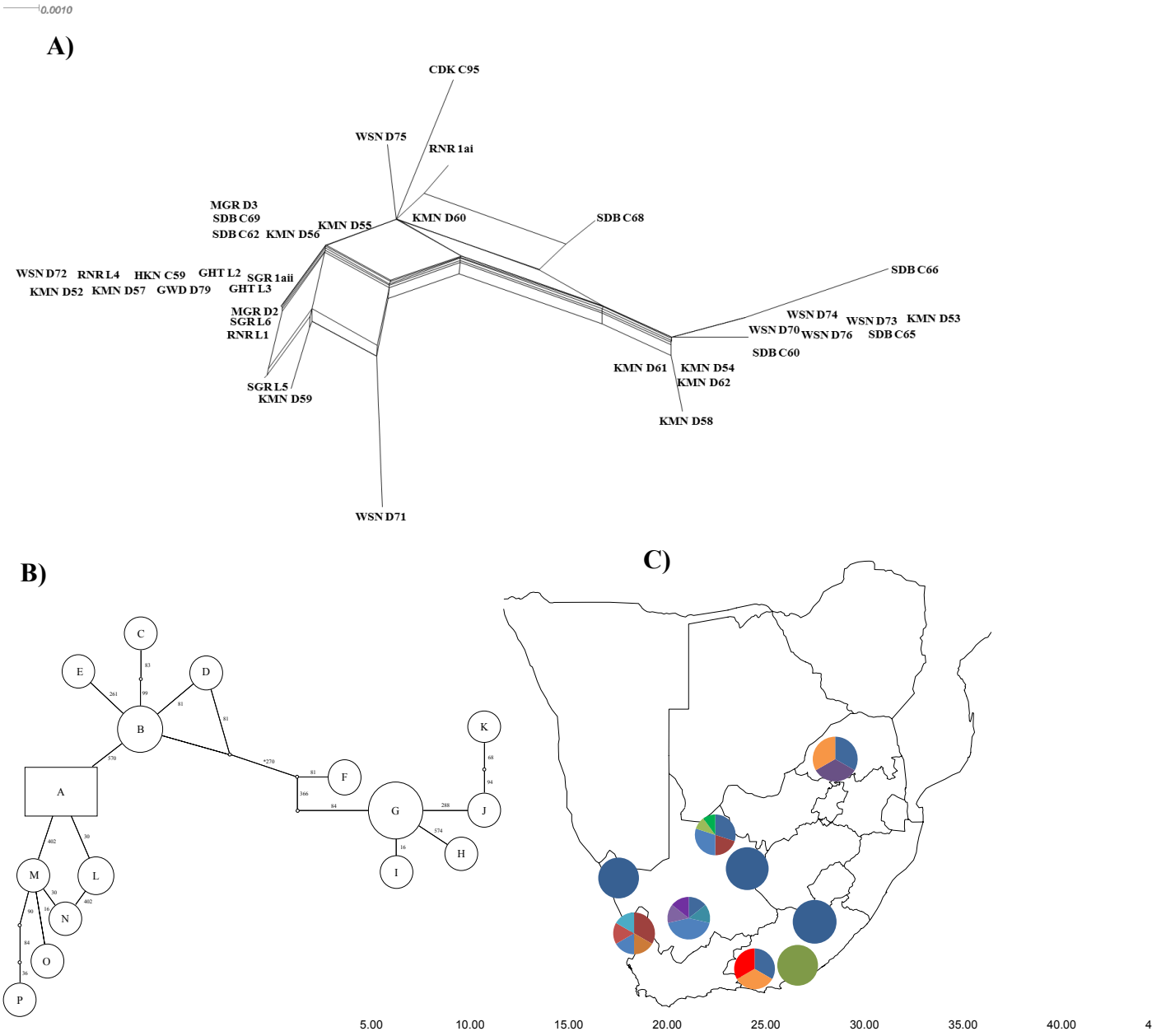

Figure 3.8: A) NeighbourNet tree, B) haplotype network and C) distribution map for *D. maculatus*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

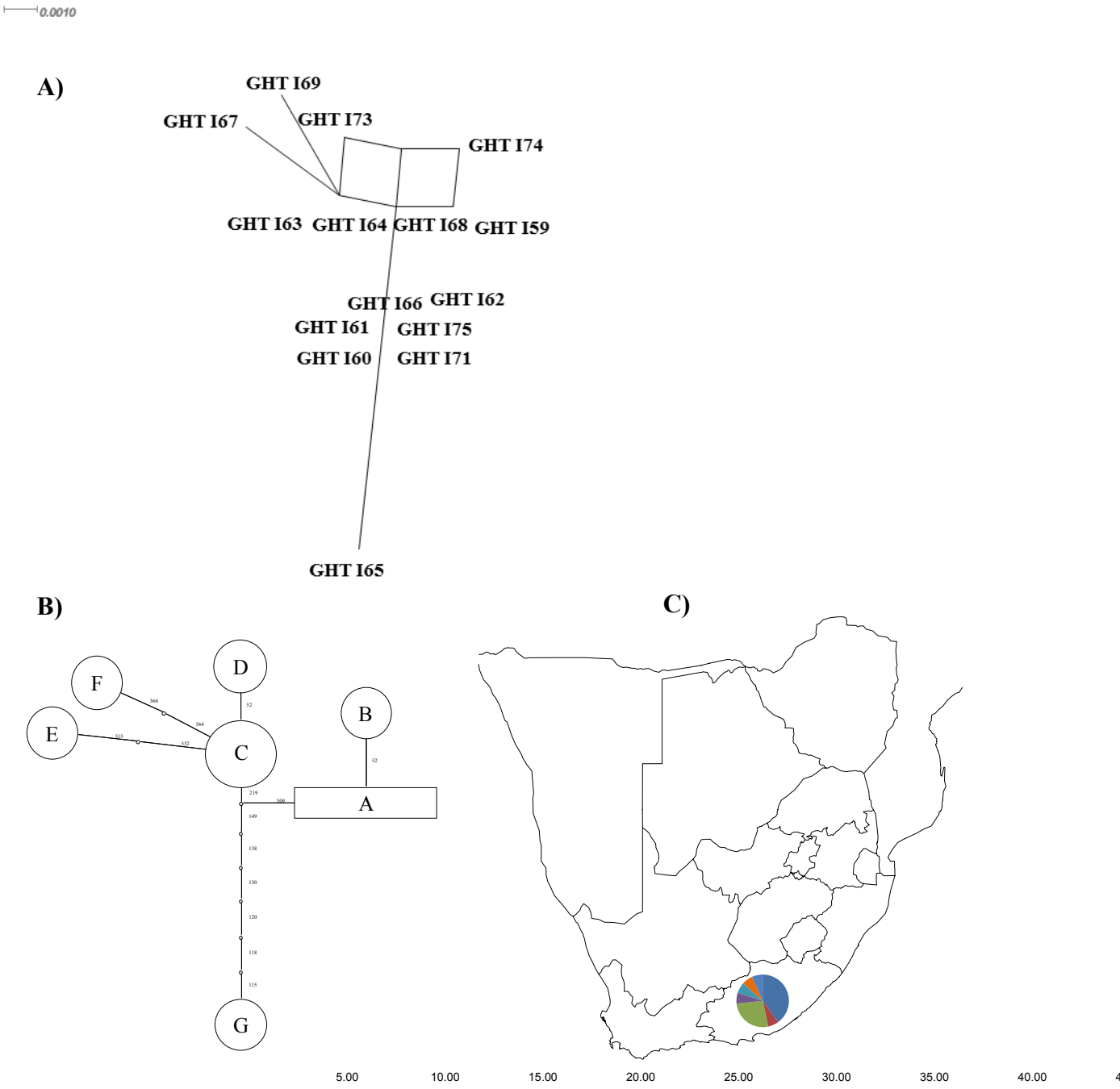

Figure 3.9: A) NeighbourNet tree, B) haplotype network and C) distribution map for *F. forcipatus*. Names of locations are abbreviated on the distance tree. The size of each circle in the haplotype network indicates the frequency of the haplotype. Small circles represent mutational steps separating haplotypes. Pie charts depict haplotypes present in that area and each colour of the pie charts represents a single haplotype.

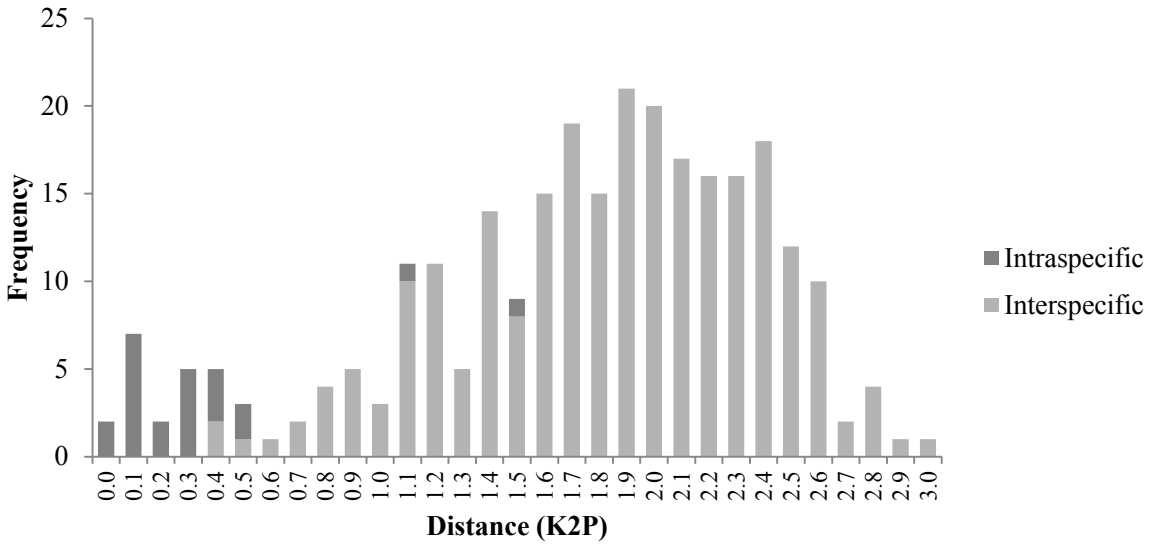

Figure 3.10: Frequency of genetic distance estimates (Kimura-2 parameter) within and between species, revealing the lack of a distinct "barcoding gap".

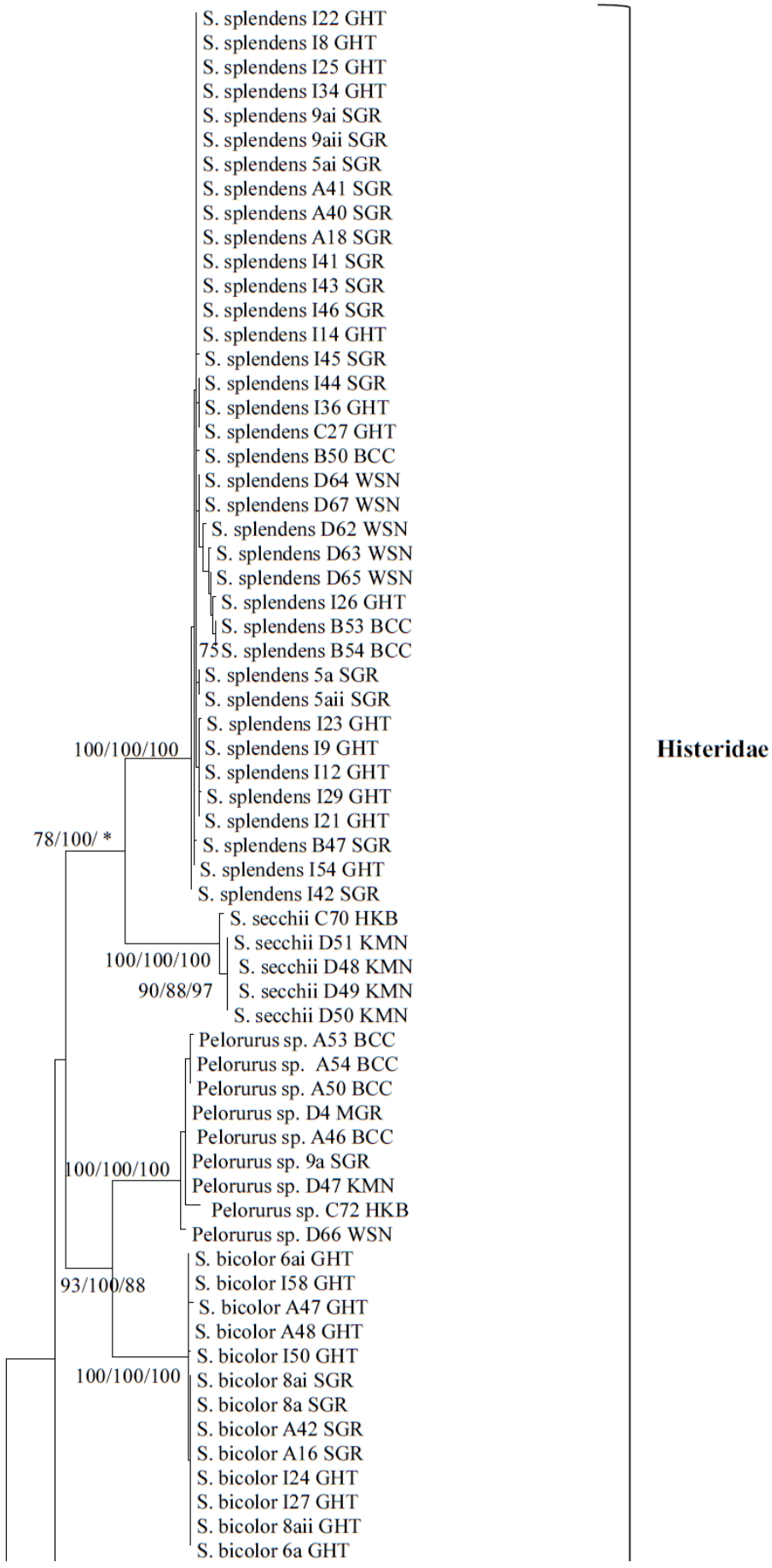

Figure 3.11: Maximum likelihood (ML) tree with 1000 bootstrap replicates. Bootstrap and branch support values are shown at the ends of nodes. Neighbour joining (NJ) values and posterior probabilities are shown in the format ML/NJ/BI. The tree is rooted with *Tetraphalerus bruchi* (Ommatidae).

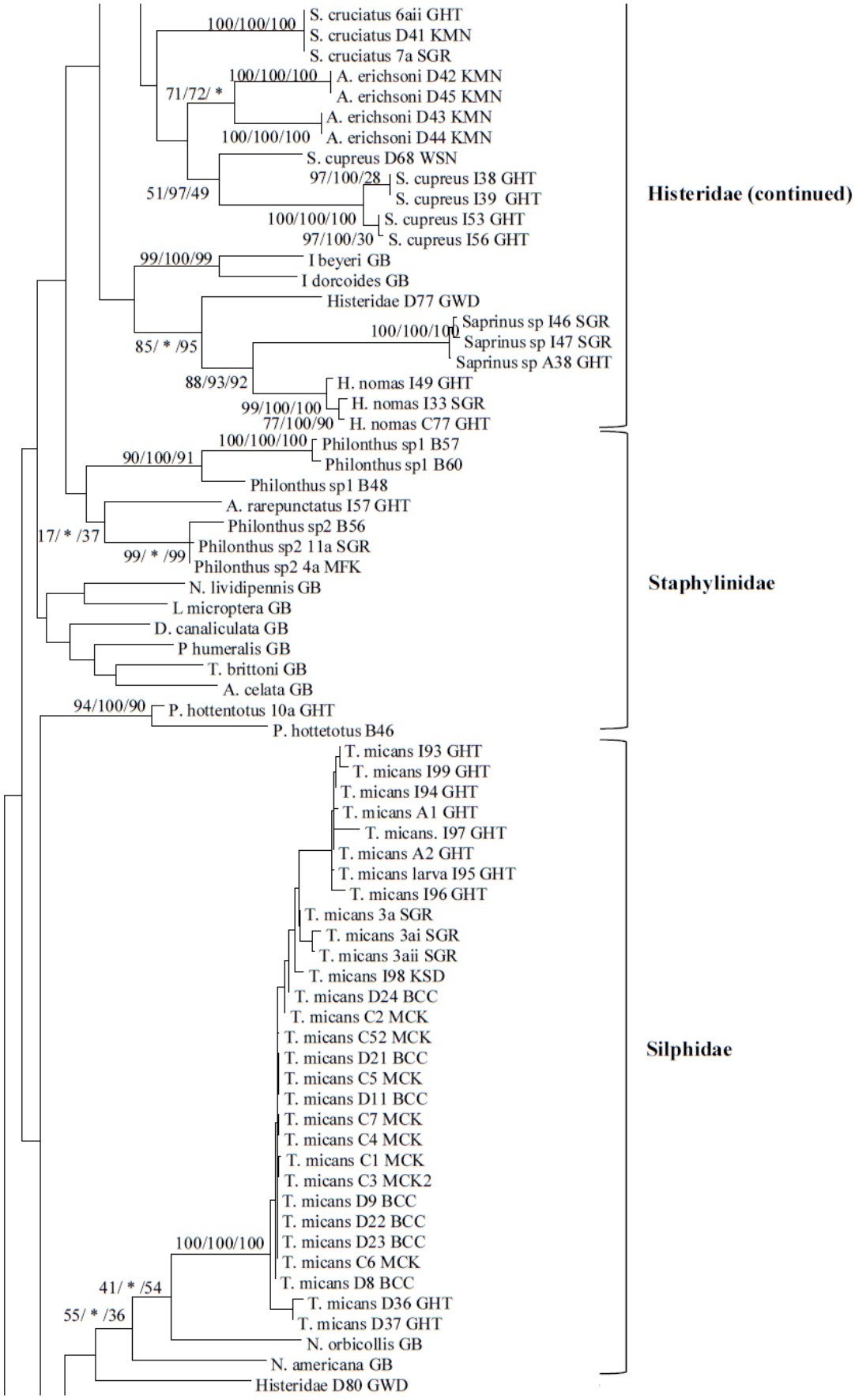

Figure 3.11: continued.

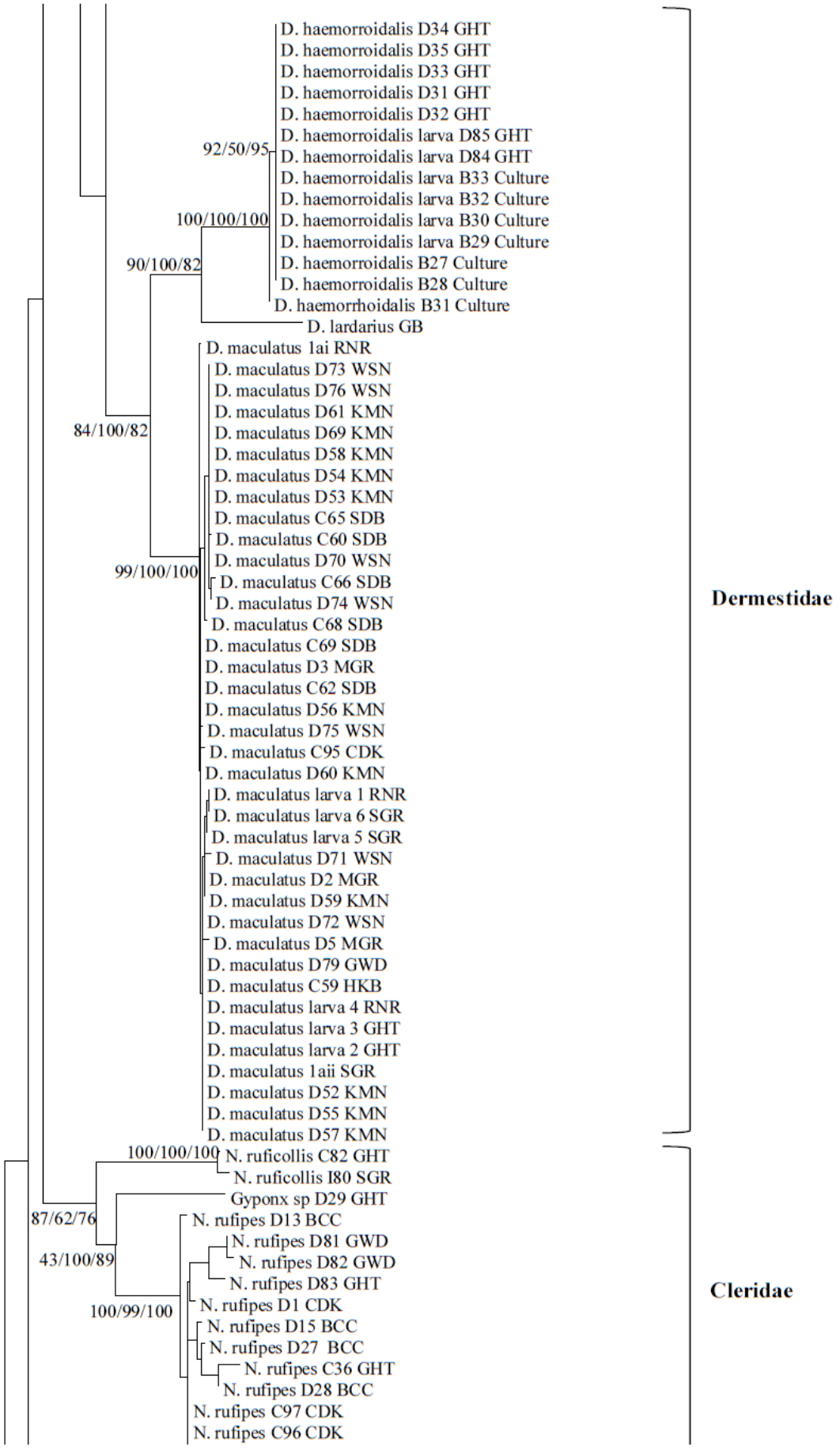

Figure 3.11: continued.

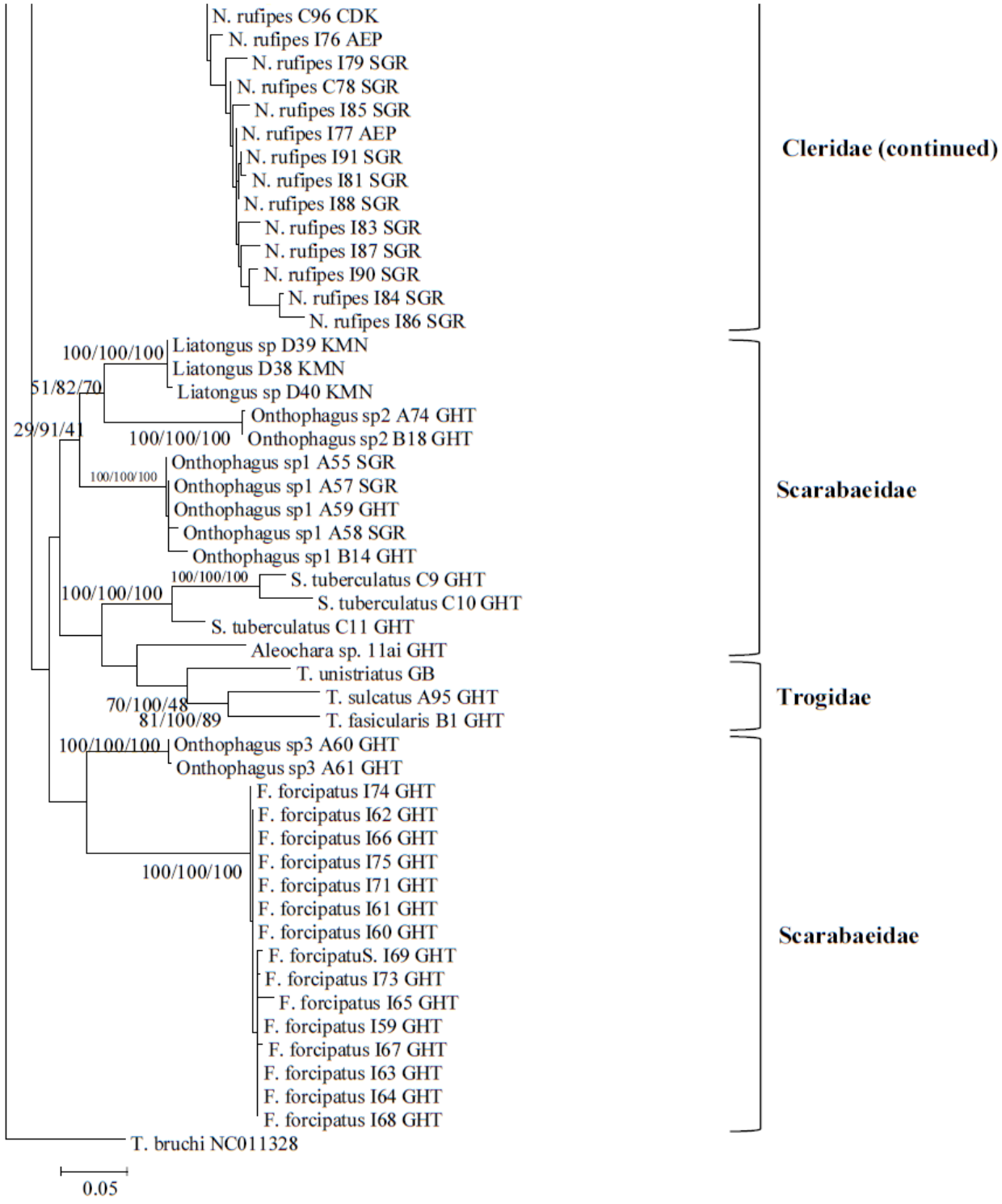

Figure 3.11: continued.

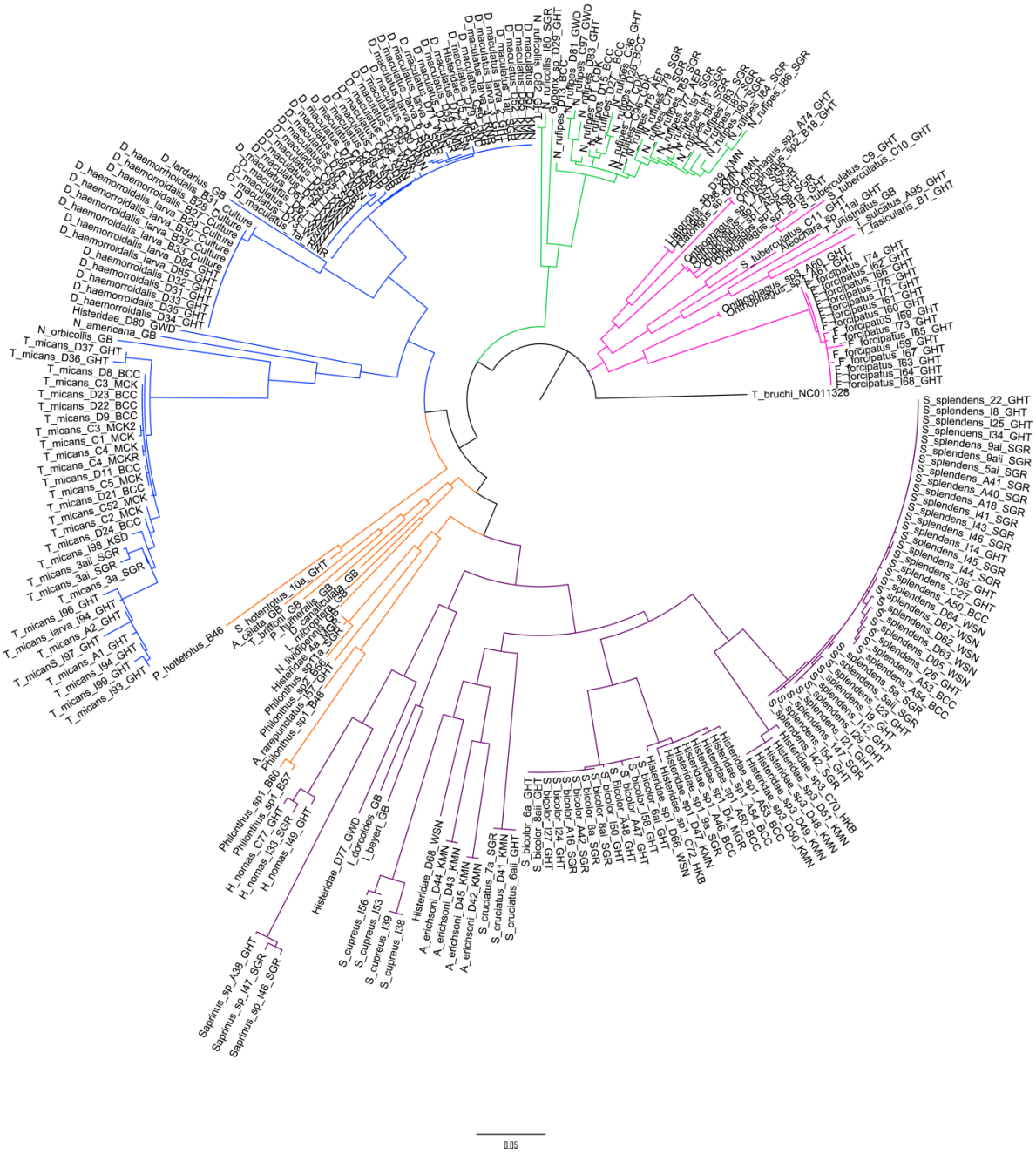

Figure 3.12: Circular topology of the maximum likelihood tree in figure 3.11.

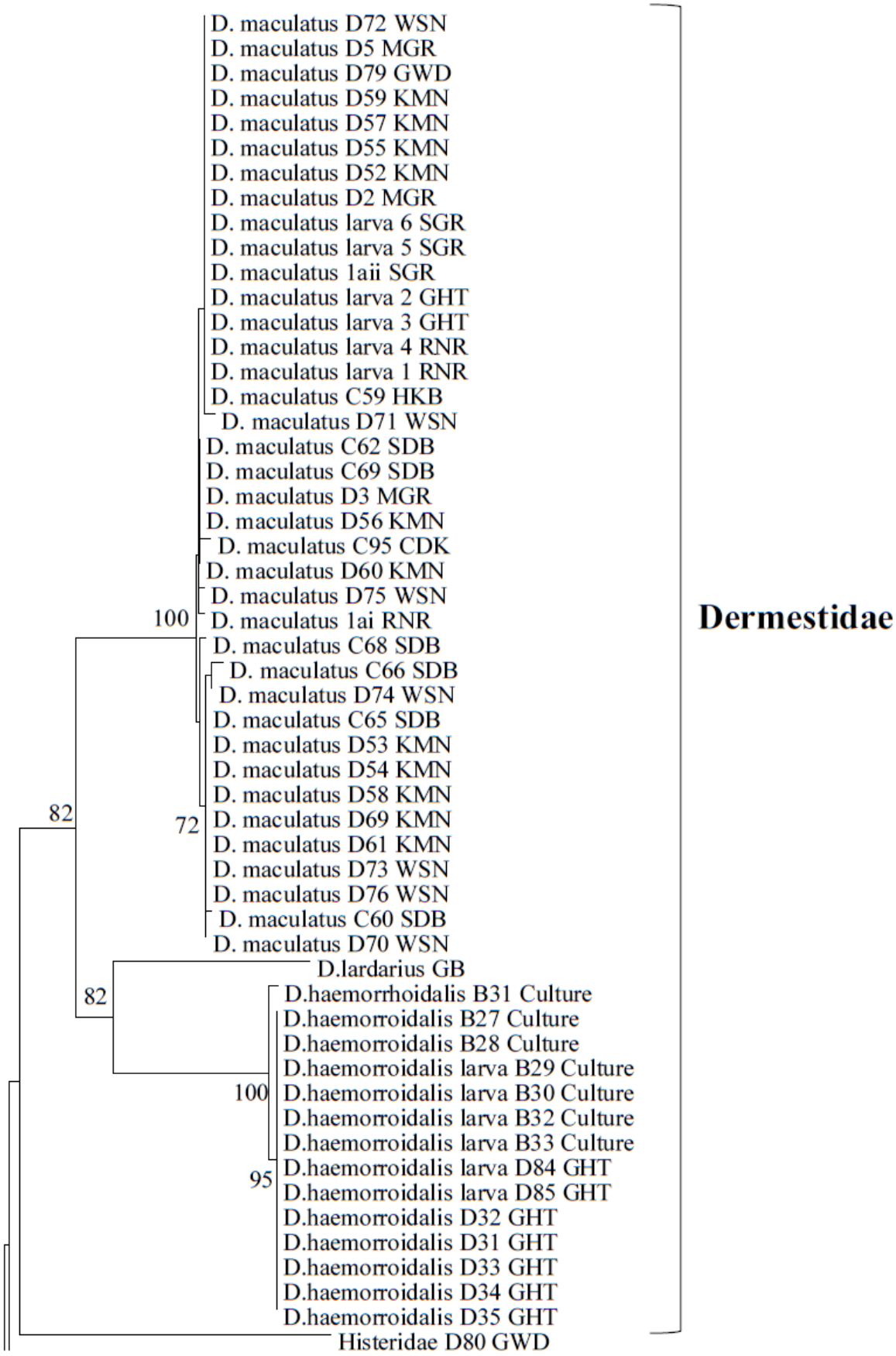

Figure 3.13: Kimura 2-parameter neighbour-joining tree (1000 bootstrap replicates) of some South African carrion beetles with additional sequences from GenBank. Numbers indicate the level of bootstrap support for the branch. Only values above 70% are shown. The tree is rooted with *T. bruchi* (Ommatidae).

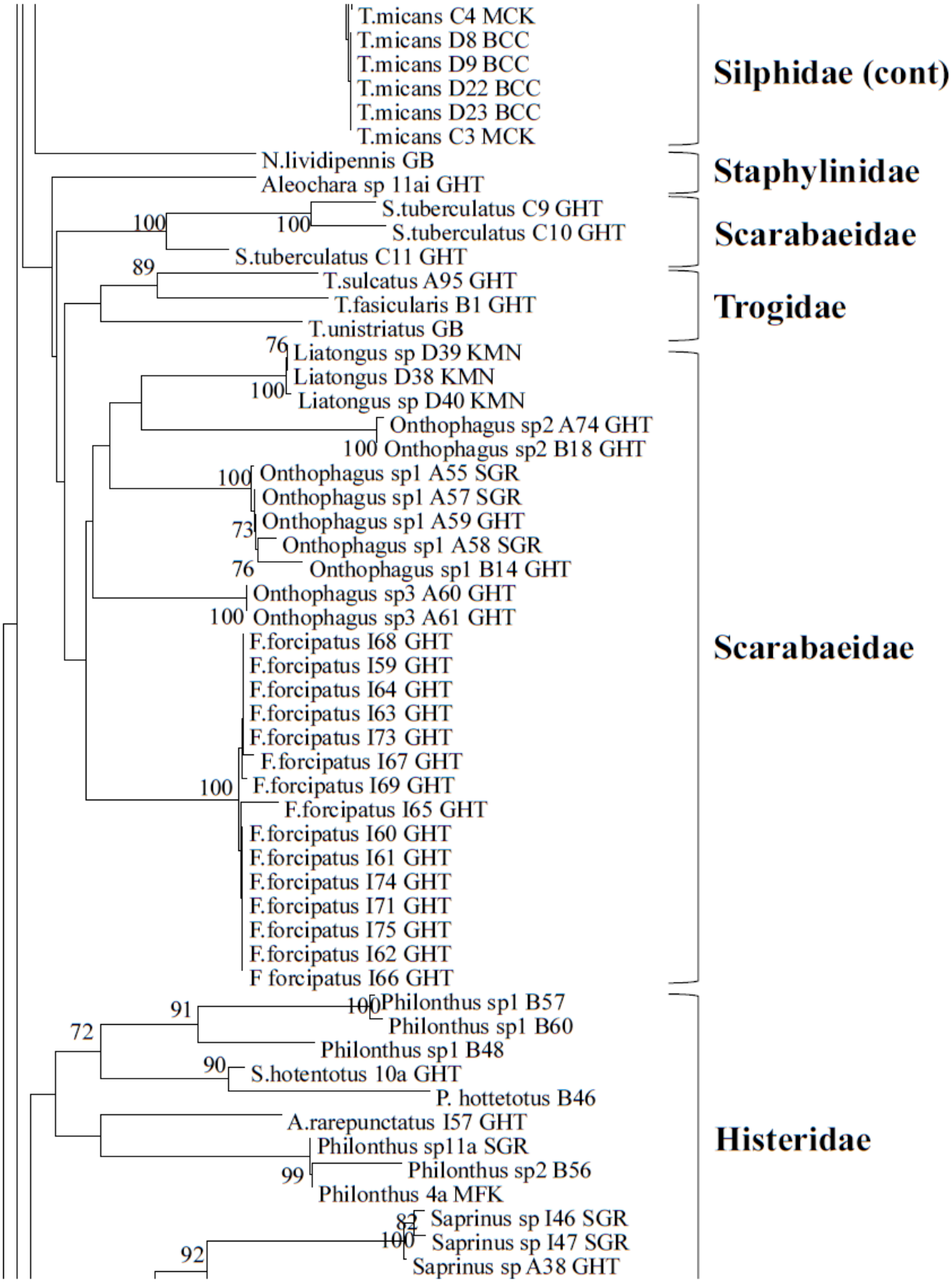

Figure 3.13: continued.

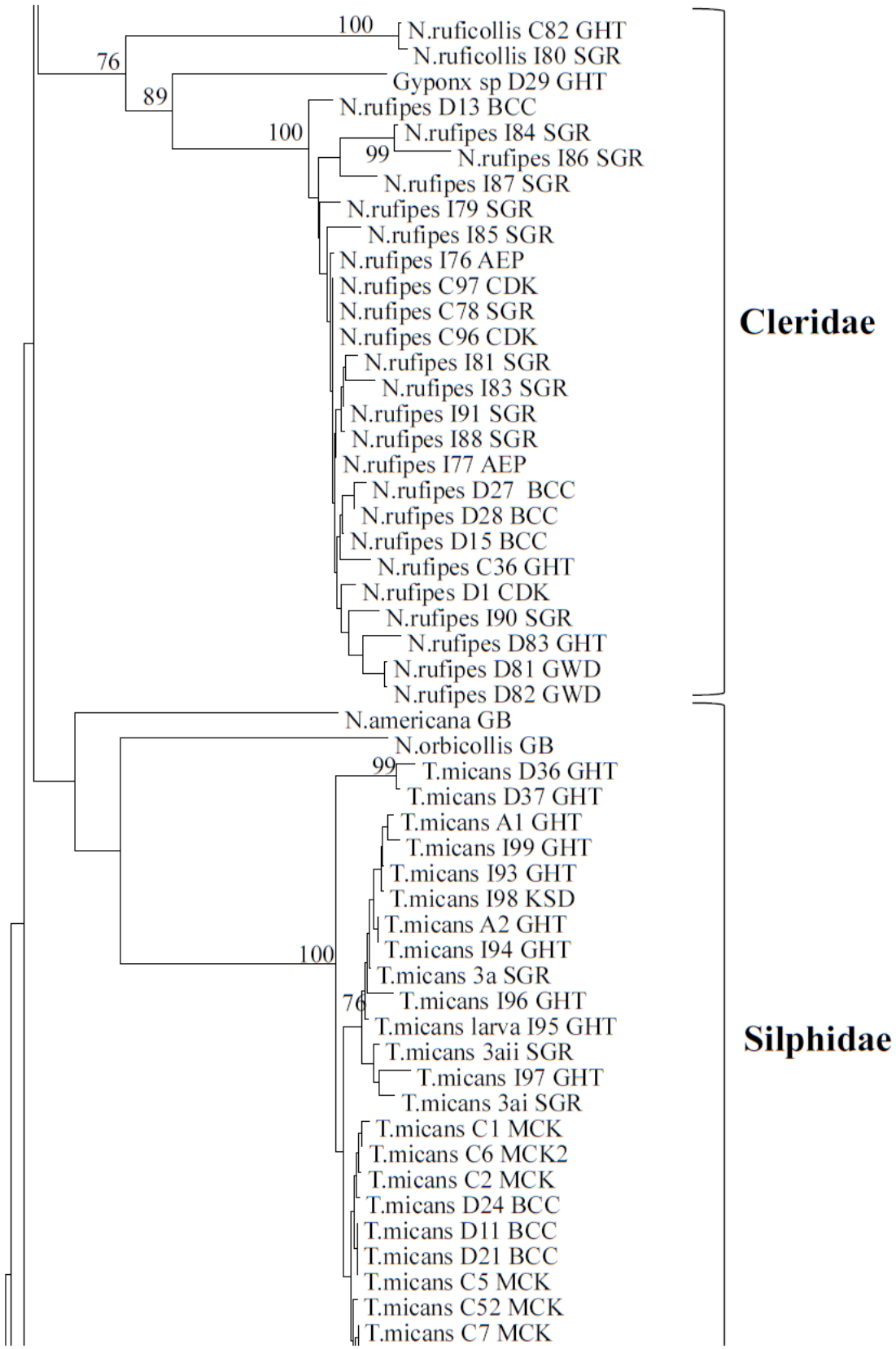

Figure 3.13: continued.

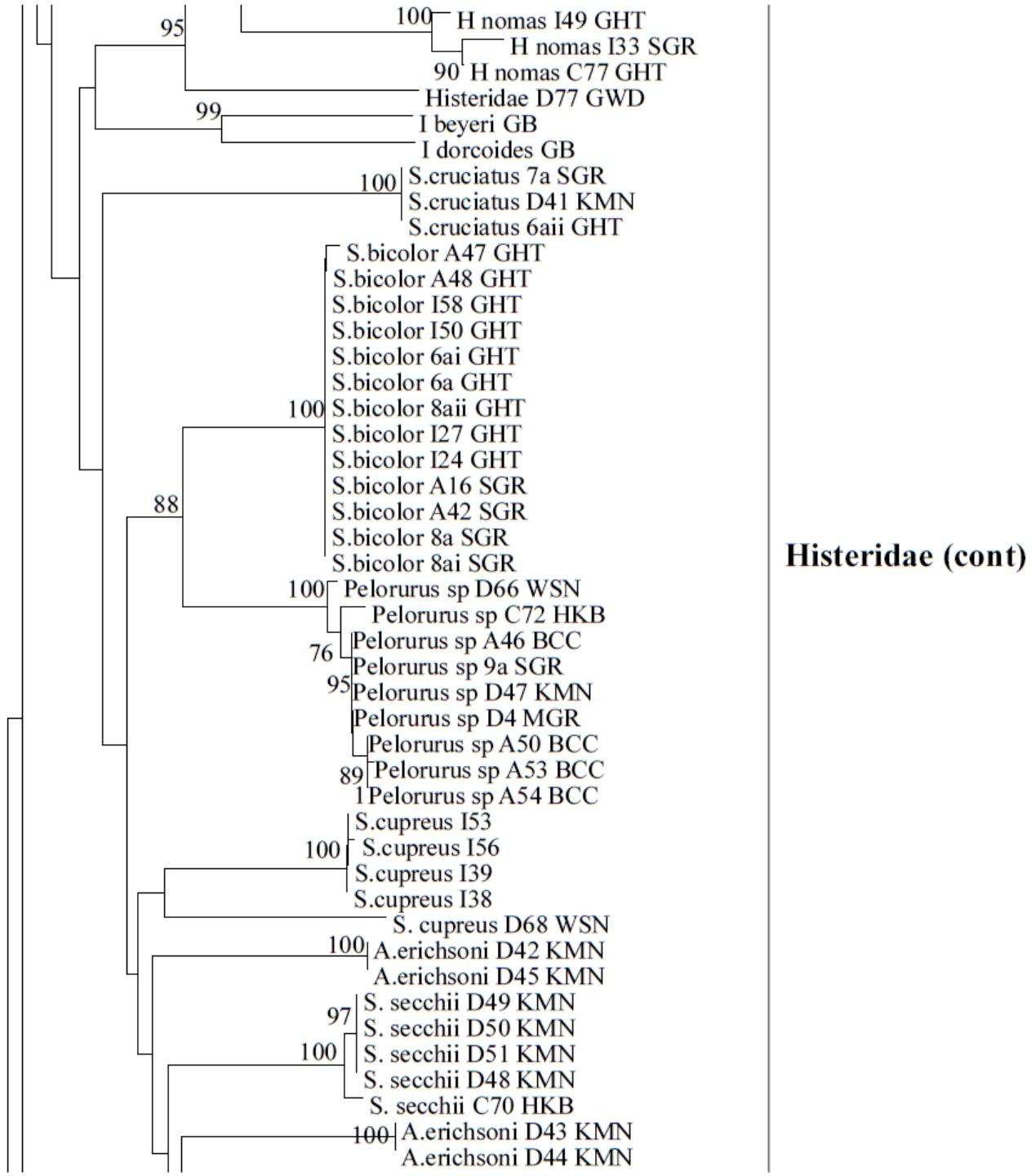

Figure 3.13: continued.

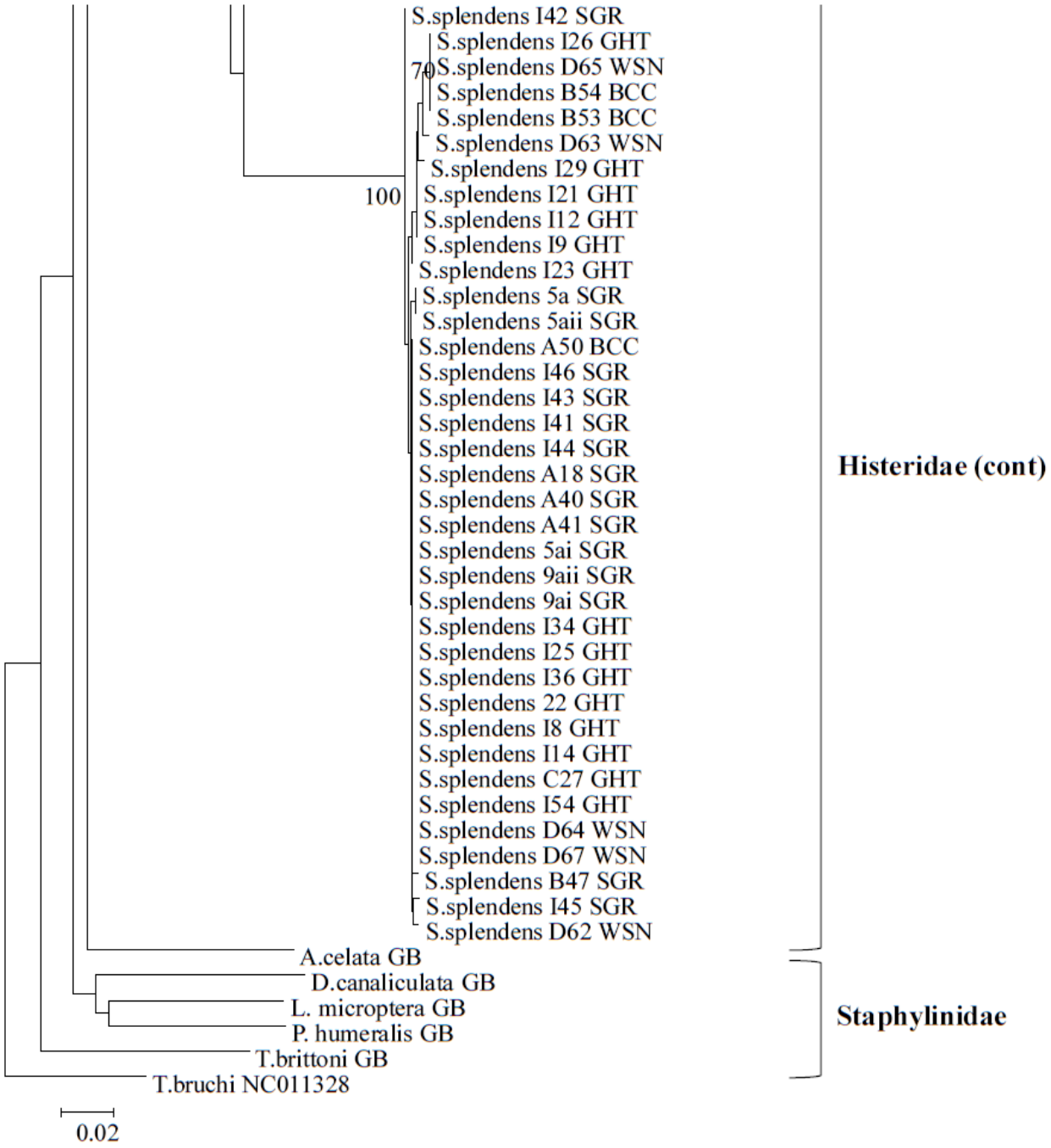

Figure 3.13: continued.

4 A comparison of nucleotide diversity across COI-COII to that of DNA barcodes in forensically significant southern African beetles

4.1 Introduction

In the previous chapter it was established that the DNA barcoding concept is not an entirely satisfactory tool for identifying insects in a forensic setting because its failure rate is problematic for court presentation. This raises the question of whether there are any molecular markers that are more reliable.

Barcodes are not always likely to resolve identification of all closely-related species as barcode phylogenies only infer the relationships for the particular fragment of the gene. It therefore becomes necessary to have a multi-gene approach with identification studies where information from other regions incorporated with of the barcode verifies identification (McDonagh, 2009; van Nieukerken *et al*., 2009). Barcoding aims for consistency of a gene region for identifying species. Mitochondrial regions have traditionally been used for insect identification including other regions of COI (Brunner *et al*., 2002, Kohlmayer *et al*., 2002), 16S rRNA (Brown *et al*., 2002) and cytochrome *b* (Khemakhem *et al*., 2002). Other mtDNA regions such as Cytochrome Oxidase II (COII), NADH dehydrogenase (ND) 1, ND 2, ND 4, ND 5 and 12S rRNA are often used for identification purposes (Loxdale & Lushai, 2009; Caterino *et al*., 2000). Nuclear genes are less commonly used for identification but the rRNA internal transcribed spacer regions (ITS1 and ITS2) have also been used. These genes have not been used across all taxa and lack the standardisation required by DNA barcodes. These have been used for identification of some Diptera and Coleoptera (Tables 4.1 and 4.2), but have not been used consistently across these orders. The bicoid (Bcd) (Park *et al*., 2012) and period (Per) (Williams & Villet, 2013; Guo *et al*., 2013) nuclear genes have also been less commonly used to differentiate forensically important blowfly species, although they have higher discriminatory power in dipterans than in other insects. Since beetles have a forensic appeal, finding an additional region of DNA to identify them is valuable. Several potential alternative genes and reasons for focussing on mitochondrial genes were discussed in Chapter 1. The evaluation of COI and COII together as identification markers has been evaluated in a number of studies (Alessandrini *et al*., 2008; Tan *et al*., 2010; Preativatanyou *et al*., 2010). COII has traditionally been used in population genetics and evolutionary studies as a result of its high

variation (Junquiera *et al*., 2002), but Ying *et al.* (2007) and Guo *et al.* (2010) identified flies of forensic interest using only COII sequences. Roe & Sperling (2007) examined the nucleotide divergence over COI-COII sequences and did not identify any optimally informative part of these genes. They also assessed the effect of fragment length from the beginning, end and midpoint of their sequences and found the results to conflict. Similar findings were reported for a fragment used for the identification of chironomid midges (Ekrem *et al*., 2007).

Here, similar analyses are carried out on the region of mtDNA downstream of the barcoding region including the remaining fragment of COI, $tRNA^{Leu}$ and most of COII to determine whether a more informative mitochondrial region exists and whether there is an optimal length on which to base identifications. Representatives of four carrion-feeding beetle families were used to assess the efficacy of barcoding pertaining to these insects and indicate whether a new region of mtDNA will be more useful in terms of identification when calculating a PMI. This depends on a more informative region being taxon-wide and whether flanking regions are sufficiently conserved to allow for the design of effective primers.

Degradation has been found to limit the amplification of long sequences (>200 bp) from museum specimens older than ten years old (Whitfield, 1999) and it is difficult to recover their barcodes cheaply and quickly (Hajibabaei *et al*., 2005). To examine the efficacy of short fragments of DNA within the barcoding region to accurately identify species, Hajibabaei *et al*. (2006) performed analyses on short sequences (~100 bp) of museum wasp and moth specimens and found them to be as effective as full-length barcodes. They were less effective in delineating fishes, but were generally successful as long as they were placed within closely-related species assemblages. Meusnier *et al*. (2008) went on to determine the shortest fragment that would delineate species as well as the full-length barcode and also concluded that a region of ~150 bp would be adequate. An additional aim is to determine the effects of fragment length for identification of beetles.

4.2 Materials and Methods

The extracts used in Chapter 2 were used again to obtain longer sequences for four species of staphylinids and two species each of histerids, scarabs and clerids (Table 4.3). Two specimens from ten different species were sequenced.

The 2.2kb region of mtDNA including the barcoding region (between LCO1490/TY-J-1460 and HCO2198) and the rest of the COI gene, $tRNA^{leu}$ and most of the COII gene was sequenced. The barcode sequences from Chapter 2 were concatenated with primer pairs C1-J-2183 (5'-CAACATTTATTTTGATTTTTTGG-3')/C1-N-3014 (5'-

TCCAATGCACTAATCTGCCATATTA-3') (Simon, *et al*., 1994) and C1-J-2756 (5'-ACATTTTTCCCCCAACATTT-3') or TL2-J-3033 (5'-AATATGGCAGATTAGTGC-3')/TK-N-3786 (5'-GTTTAAGAGACCATTACTT-3') (Simon, *et al*., 1994). The TL2-J-3033 primer was used in cases where C1-J-2756 would not amplify. The three regions that were concatenated with each other extend from around 1451 to 3710 on the *Drosophila yakuba* mitochondrial genome (Clary & Wolstenholme, 1985).

4.2.1 Amplification of C1-J-2183/C1-N-3014 region

The second fragment was amplified using the primers mentioned above in a total volume of 25 µl and the recipe was as follows: 12.5 µl Promega PCR Master Mix, 1.5 µl of each primer (10 µM), 2 µl of magnesium chloride (50mM), 5.5 µl PCR grade water and 2-3 µl DNA extract (55-100 ng/µl). A Thermohybaid Pf2 (ThermoScientific, Boston, Massachusetts) thermocycler was used for the PCR with an initial denaturing step of 94°C for 5 minutes. This was followed by 40 cycles of 94°C for 1 minute, 50°C for 40 seconds and 72°C for 30 seconds. The reaction was completed by a final elongation at 72°C for 10 minutes.

4.2.2 Amplification of C1-J-2756 or TL2-J-3033/ TK-N-3786 region

This fragment was amplified using the same recipe and under the same conditions as above besides an annealing temperature of 48°C. For samples amplified with TL2-J-3033, the annealing temperature was also 50°C.

4.2.3 Sequencing

The products were run on 1% agarose gels and viewed by UV transillumination. Forward and reverse strands were sequenced by Macrogen (Seoul, Korea) on an ABI3730 XL automatic DNA sequencer, using the same primers as used in PCR.

4.2.4 Alignment

Editing of sequences was done using BioEdit 7.1.3.0. (Hall, 1999). MEGA 5.05 (Tamura *et al*., 2011) was used to align sequences.

4.2.5 Data analysis

4.2.5.1 Data sets

For assessment of sequence variation within species between geographic regions, the intraspecific diversity and divergence was calculated using DnaSP version 5.10.0 (Librado &

Rozas, 2009). The sequence variation was mapped graphically for the whole region, but both values were converted to percentages for ease of comparison.

Nucleotide polymorphism patterns were calculated using sliding window analysis performed on the barcoding region using DnaSP version 5.10.0 (Librado & Rozas, 2009). The regions amplified by the three primer sets, namely (Dy# 1460-2198, Dy# 2183-3014 and Dy# 2756-3786) were analysed in conjunction with one another and separately. COI was considered separately and also in combination with COII. Windows of 600 bp with a step size of 5 bp were chosen for analysis as it is comparable to the size of the barcoding fragment.

4.2.5.2 Sliding window analysis

For assessment of sequence variation within species between geographic regions, the interfamily diversity (p) (Nei, 1987) and divergence (K) (Tajima, 1983) was calculated using DnaSP version 5.10.0 (Librado & Rozas, 2009). The sequence variation was mapped graphically for the whole region, but both values were converted to percentages for ease of comparison. Default settings, including a Jukes-Cantor correction to divergence estimates, were used to analyse nucleotide substitution patterns as indicated in the text for the analyses to be comparable to those of Sperling & Roe (2007).

4.2.5.3 Phylogenetic analysis

Neighbour-joining trees with 1000 bootstrap replicates were constructed in MEGA 5.05 (Tamura *et al*., 2011). They were created using windows of 200 bp along the length of the sequenced fragment with a step size of 100 bp to determine whether the species would be differently organised on the tree depending on the location along the fragment. A window of 200 bp was used as it is a minimum length that would be used to construct a phylogenetic tree and 600 bp were used to represent the size of the DNA barcode.

4.2.5.4 Fragment length

To determine whether fragment length has an effect on the level of divergence over the region, the nucleotide diversity of fragments increasing by 100 bp increments from both ends of the sequenced region and the midpoint was calculated. It was primarily done to show the optimal sequence length needed to accurately delineate species.

By comparing the nucleotide diversity of short fragments along the concatenated sequences, it was possible to estimate the minimum length needed to produce viable identifications. The nucleotide divergence for fragments of 100 bp increments was calculated

from three starting points, being the 5' end of COI, the 3' end of COII and the midpoint of the whole sequence. These were displayed graphically.

4.3 Results

4.3.1 Alignment

The alignment in MEGA 5.05 (Tamura *et al*., 2011) contained two specimens for ten species within five families (Table 4.1). The *Necrobia rufipes* sequences contained deletions of 67 bp (Dy#1851-1918) in the barcoding region and another downstream of 65 bp (Dy# 2254-2319), both in COI. The deletion removes over 40 amino acids and mutates about 112 intervening amino acids so the sequences would only be accepted onto GenBank as NUMTs. The sequences were still similar enough to the remaining bases to be aligned easily.

4.3.2 Phylogenetic reconstruction

Twenty-one neighbour-joining trees constructed using 200 bp windows with 100 bp step sizes and 17 trees of 600 bp windows with 100 bp step sizes were created. All the trees showed consistent delineation of species. Branch lengths varied slightly from tree to tree and the order of families were swapped occasionally, but this is not a consideration for DNA barcoding. A representative tree is included (Figure 4.1).

4.3.3 Fragment length analysis

Starting from the 5' end, more variation is found in the 100, 200, 300, 400 and 700 bp fragments than for longer sequences. The line representing fragment length from the 3' end shows more variation at 700 bp than the fragment of 900 bp and the midpoint start also shows decreasing resolving ability after 700 bp. There is no common trend for variation to increase with fragment length, but all lines seem to have a consensus level of variation at 600 bp (Figure 4.2).

4.3.4 Sliding window analysis

Regions of overlap denote the diversity of the family represented in both the 'Family x Family' conditions (Table 4.4). The ~645 bp DNA barcoding region lies between Dy#1490-2198. In the sliding window analysis, distribution of nucleotide diversity in the two dermestids was highest in the region between 2723-3210 bp. The region between 2146-2747 bp had highest diversity in the clerids, histerids showed most diversity between 2497-2732 bp and

scarabs between 2563-2797 bp (Table 4.5). The Cleridae are the only family where the region of highest diversity overlapped the barcoding region. The other families all have increased nucleotide diversity downstream of the barcoding region, suggesting that the barcoding region is not the most informative for these families (Figure 4.3).

4.3.5 Regional variability

All families show a degree of variability, but most lay between the scarabs (*Frankenbergerius forcipatus* and *Sarophorus tuberculatus*). The highest variability lay downstream from the barcoding region. For the families for which sister species were compared, variability was low over the whole region. The region of highest variability was between the 2724 and 3524 bp positions (800 bp) (Figure 4.3).

4.4 Discussion

A longer fragment of mitochondrial DNA inclusive of the barcoding region was sequenced to assess levels of variability along COI-COII to determine whether a more informative region than the barcode exists along the same region of mtDNA that could be used in preference to the standard barcode for the forensically important beetles analysed. A single-locus mtDNA barcode would be a more effective marker because of the relatively smaller population size in comparison to the nuclear genome and additional benefits such as a general absence of repetitive DNA, pseudogenes and large spacer sequences. Here, there was a general inclination for the region downstream of the barcode to be more informative than the barcode region itself.

Variation was apparent across the region sequenced in these four families. The interfamily sliding window analysis showed the majority of families to increase in variability in the COII region, downstream of the barcode (Figure 4.3). Exact values are given as percentages in Tables 4.4 and 4.5. The exception to this was the Cleridae whose COI region showed more variation than the COII region. This suggests that a fragment in the latter half of the sequence (COII) may be more useful when it comes to identifying a reliable species barcode. Most variation lies downstream of the proposed barcoding region between Dy# 2724-3524 (Figure 4.5).

Species variation was compared within families to show regions of highest divergence (Figure 4.4). There was not any clear tendency in the interspecific variation as most of the species did not differ significantly besides the Scarabaeidae. A probable reason for this is that the pairs of species of dermestids, clerids and histerids that were used in this study each belong

to the same genus and are therefore not highly differentiated. The scarabs belong to two different but closely related genera, and so show more differentiation. As with the interfamily variation, the most diverse region is slightly downstream from the standard barcode. The increased level of variation downstream from the barcoding region (Dy# 2724-3710) is also shown in the number of informative sites per 200 bp with the increased variation outside of the barcoding region (Figure 4.5). All of these analyses indicate that the standardised barcode is not the most ideal fragment for these, and perhaps other, beetles. Cox *et al*. (2013) showed a region close to the 3' end of COI between Dy# 2183-3014 to be appropriate for identifying stag beetles.

Roe & Sperling (2007) found that the regions of highest divergence varied among organisms. The choice of ideal marker may depend on the taxon being studied. For example, Derocles *et al*. (2011) used the long wavelength rhodopsin (LWRh) gene as well as COI to identify and delineate European aphids. Most teleost fishes can be identified using CytB and rhodopsin (Sevilla *et al*., 2007) suggesting that COI barcodes may not be needed at all if more than one region is being used. In spite of successful species identification using DNA barcodes, there have been cases that have showed COI to not give enough resolution and results could be misleading (Hurst & Jiggins, 2005; Linares *et al*., 2009; Whitworth *et al*., 2007). A multi-locus approach thus seems increasingly necessary (van Nieukerken *et al*., 2012).

If it becomes routine to verify barcodes with information from other sequences, it would be ideal to locate a fragment that has similar characteristics to those of COI barcodes (short, sufficiently variable, conserved flanking regions), albeit nuclear. Having an extra region to sequence somewhat defeats the object of a species barcode, especially if it is taxon-specific. It is likely that one or more regions would not have the same capacity for taxon-wide standardisation of primers and PCR cycling conditions that is required for a robust DNA identification system. Also, using multiple extra regions increases the time spent in the laboratory and the chances of all the reactions working first time are slim. Having single short sequences are also less cumbersome to use. However, inconsistencies in the barcoding approach may be the deciding factor. When implementing a validating sequence it is not always possible to determine whether the results from COI are bogus (Moulton *et al*., 2010).

From the string of studies having showed barcoding to be reliable in species delineation, it is apparent that molecular information needs to be taken into account in modern taxonomy. However, DNA should not be the only source of species-differentiating characteristics and barcodes will not displace the need to properly assess all aspects of species theories (Wheeler *et al*., 2004). Indeed, barcodes should be inclusive of all diagnostic characters from both

morphology and molecular data. This is what the 'integrated taxonomy' concept embodies (Godfray, 2002). Integrating these two forms of information will help to reconcile the 'Tower of Babel' experienced by insect molecular systematics due to lack of collaborative effort on the part of researchers (Caterino *et al*., 2000).

One would expect the diversity to be positively correlated with fragment length as a longer sequence should have more stochastic variation than a shorter one. This was not the case with these data. Fragment length increased from any of the starting points for more informative sites suggesting that any length of sequence (100-1000 bp) will be adequate to distinguish between these species (Figure 4.2). Trees constructed with 200 bp and 600 bp fragments showed identical delineation of species in spite of varying branch lengths and order of families. Meusnier *et al.* (2008) performed an analysis on barcodes to determine the minimum length needed to give accurate species identification and found that sequences of 150 bp were long enough to do this. There is merit in pursuing a reliable mini-barcode as it will ultimately be cheaper to sequence samples for large barcoding projects, and samples with highly fragmented DNA may be sequenced (Shokralla *et al*., 2011). This would be of particular use in forensic entomology as old and degraded DNA is a major limiting factor in the amplification process. Although a longer sequence is probably more reliable, the advantages of having a short barcode make them a good candidate for use as a forensic marker. A 189 bp fragment of COII was successfully used to identify forensically important Chinese and Egyptian flies (Guo *et al*., 2010; Aly *et al*., 2012), so the same may apply to the beetles used in this study. There was no optimal length of fragment so when identifying a species in which DNA is degraded, it should be possible to get at least a short fragment on which to base identification. A common condition of using short fragments in these studies is the option of comparing the sequence to longer ones of known identification. Having these barcodes will be a backbone for short fragment comparison and may have use when puparia or larval exuviae are the main source of information on which to calculate a PMI. Short fragments may be equally useful provided there are longer fragments of known identification to compare them to. Nevertheless, using a small fragment of DNA may fail to produce an accurate representation of the total variation across that gene and could result in misrepresenting inter- and intraspecific divergence between closely related species. This would lead to inaccurate species identification (Roe & Sperling, 2007).

Longer sequences are considered to be more reliable, but are not practical in terms of a forensic barcode. The current length of the barcode (~658 bp) was selected so that it could be sequenced in a single reaction in conventional cycle-sequencing platforms. A short sequence

is less expensive, but DNA sequencing costs have decreased dramatically since the technique was first developed in 1970s. However, shorter sequences are still cheaper. The methods for sequencing short fragments of DNA (<1400 bp) are simpler and more robust so it is possible to have a standard simple method or set of methods that can be used with the same degree of accuracy anywhere in the world. With longer sequences the interpretation and the probability of amplification errors or differential amplification is worse. This means that the sequencing process would require more skilled technicians.

Although using another fragment of DNA to verify the barcode may address ambiguously or wrongly identified specimens, barcoding does have a reasonable amount of success in identifying species (Ratnasingham & Hebert, 2013). The present barcoding region will suffice if used in conjunction with other characteristics. The idea of 'integrated taxonomy' is considered to be the only incontestable way of identifying species (Damm *et al*., 2010; Paial *et al*., 2010; Schlick-Steiner *et al*., 2010). Using a combination of characters is the only way of successfully diagnosing species as it takes the various species concepts into consideration and incorporates more than just the genetic aspect of species-complexity.

Table 4.1: Mitochondrial genes used for species identification in Diptera and Coleoptera.

Gene region	Coleoptera	Diptera
COI barcode	Schilthuizen *et al*., 2010 Cai *et al*., 2011 Davis *et al*., 2011 Zhuang *et al*., 2011	Wells *et al*., 2001 Zehner *et al*., 2004 Saigusa *et al*., 2005 Harvey *et al*., 2008 Song *et al*., 2008 Guo *et al*., 2010 Meiklejohn *et al*., 2011 Tan *et al*., 2010 Boehme *et al*., 2011 Brodin *et al*., 2012 Mazzanti *et al*., 2012 Jordaens *et al*., 2013
COI (other)	Marinho *et al*., 2012	-
COII		Guo *et al*., 2010 Aly *et al*., 2012
CytB	Sembene, 2006	Zaidi *et al*., 2011 Su *et al*., 2013
12S	-	Stevens & Wall, 1996
16S	Marinho *et al*., 2012 Tang *et al*., 2012	Li *et al*., 2010 Xinghua *et al*., 2010
ND1	Elven *et al*., 2010	Besansky *et al*., 1997
ND2	-	He *et al*., 2007
ND5	Osawa *et al*., 2005	Zehner *et al*., 2004

Table 4.2: Nuclear genes used for species identification in Diptera and Coleoptera.

Gene region	Diptera	Coleoptera
Bcd	Park *et al*., 2013	
ITS1	Ratcliffe *et al*., 2003	Szalanski, 2000
ITS2	Marinho *et al*., 2012	Sembene, 2006
PepCK		Sota & Vogler, 2001; 2003
Per	Williams & Villet, 2013; Guo *et al*., 2013	
28S	Mamrinho *et al*., 2012	

Table 4.3: Specimens used in this study for determining the optimal COI-COII fragment for identification.

Family	Species	No.	Location	Latitude	Longitude
Cleridae	*Necrobia ruficollis*	C82	Grahamstown	33°16'17"S	26°27'25"E
	Necrobia ruficollis	I80	Shamwari	33°16'08"S	26°07'18"E
	Necrobia rufipes	I88	Shamwari	33°16'08"S	26°07'18"E
	Necrobia rufipes	D1	Fish River	31°55'23"S	25°24'01"E
Dermestidae	*Dermestes haemorrhoidalis*	B27	Culture	33°18'34"S	26°31'11"E
	Dermestes haemorrhoidalis	B28	Culture	33°18'34"S	26°31'11"E
	Dermestes maculatus	C59	Hondeklipbaai	30°18'37"S	17°16'31"E
	Dermestes maculatus	C62	Saldanha Bay	32°48'37"S	18°10'02"E
Staphylinidae	*Thanatophilus micans*	D22	Blue Canyon Res	24°21'17"S	31°03'01"E
	Thanatophilus micans	I97	Grahamstown	33°16'17"S	26°27'25"E
Scarabaeidae	*Sarophorus tuberculatus*	C10	Grahamstown	33°18'48"S	26°31'07"E
	Sarophorus tuberculatus	C11	Grahamstown	33°18'48"S	26°31'07"E
	Frankenbergerius forcipatus	I62	Grahamstown	33°19'49"S	26°26'16"E
	Frankenbergerius forcipatus	I68	Grahamstown	33°18'48"S	26°31'07"E
Histeridae	*Saprinus bicolor*	A16	Shamwari	33°26'11"S	26°04'27"E
	Saprinus bicolor	A47	Grahamstown	33°20'13"S	26°40'49"E
	Saprinus splendens	A18	Shamwari	33°26'11"S	26°04'27"E
	Saprinus splendens	A44	Blue Canyon Res	24°21'17"S	31°03'01"E

Table 4.4: Interfamily nucleotide divergence (%) for six family pairs for the COI-COII region. Sliding windows of 600 bp were analysed for the whole region.

Family	N	Sliding window analysis					
		Max	Midpoint	Min	Midpoint	COI	COI-COII
Histeridae x Scarabaeidae	8	0.1236	2663	0.0911	2208	0.1090	0.1123
Histeridae x Cleridae	8	0.1236	2667	0.0867	1823	0.1109	0.1138
Scarabaeidae x Cleridae	8	0.1111	2652	0.0681	3262	0.1008	0.0925
Dermestidae x Histeridae	8	0.1486	3036	0.0597	2433	0.1048	0.1067
Dermestidae x Scarabaeidae	8	0.1475	3022	0.0597	2433	0.1049	0.1066
Dermestidae x Cleridae	8	0.1475	3021	0.0619	2191	0.1064	0.1077

Table 4.5: Within-family nucleotide diversity (%) for 4 families for the COI-COII region.

Family	N	Total%	Max%	Midpoint	Min%	Midpoint	BC (%)	2183 (%)	2756 (%)
Histeridae	4	0.113	0.124	2364-2963	0.091	1909-2508	0.100	0.118	0.122
Scarabaeidae	4	0.089	0.111	2354-2953	0.067	2979-3578	0.092	0.099	0.085
Cleridae	4	0.084	0.092	1999-2598	0.065	2664-3263	0.088	0.081	0.080
Dermestidae	4	0.107	0.150	1201-1806	0.060	611-1210	0.095	0.124	0.150

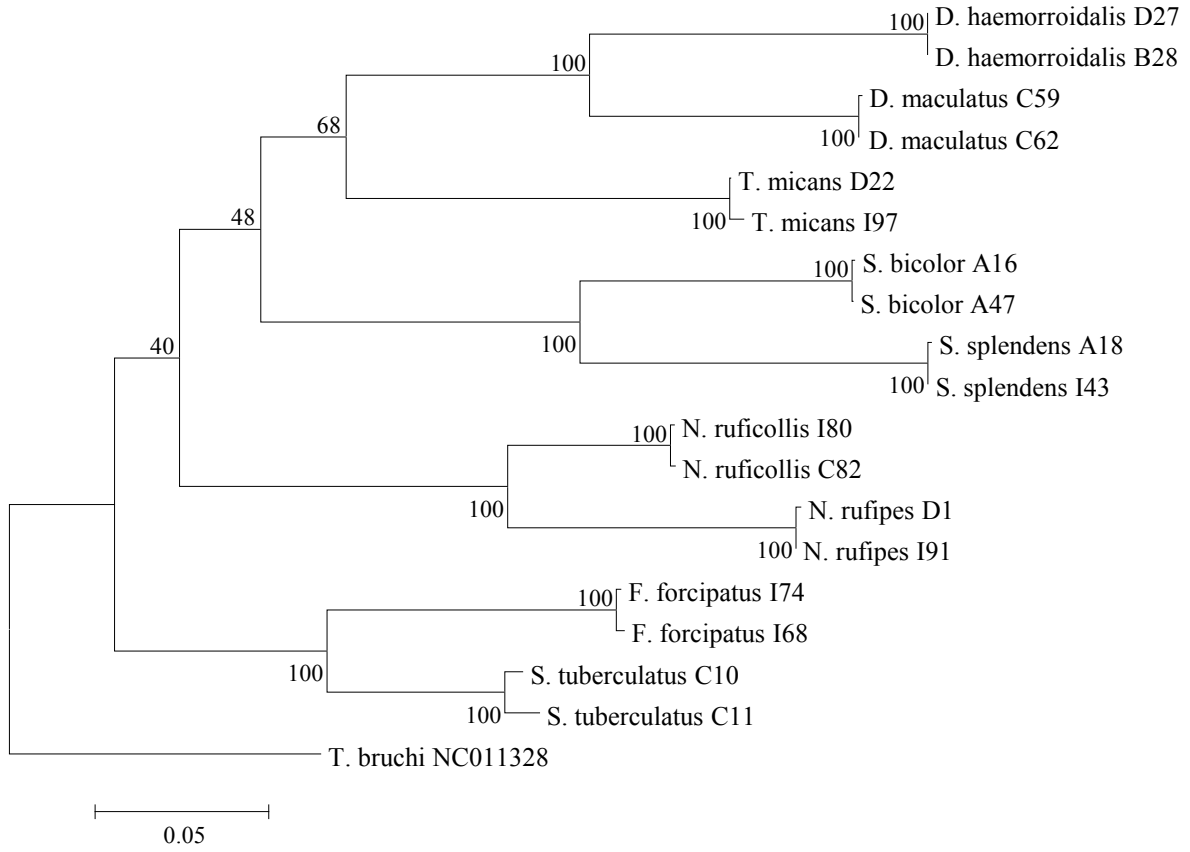

Figure 4.1: Neighbour-joining tree (K2P) with 1000 bootstrap replicates performed using 200 bp fragments. The tree is rooted with *T. bruchi* (Ommatidae).

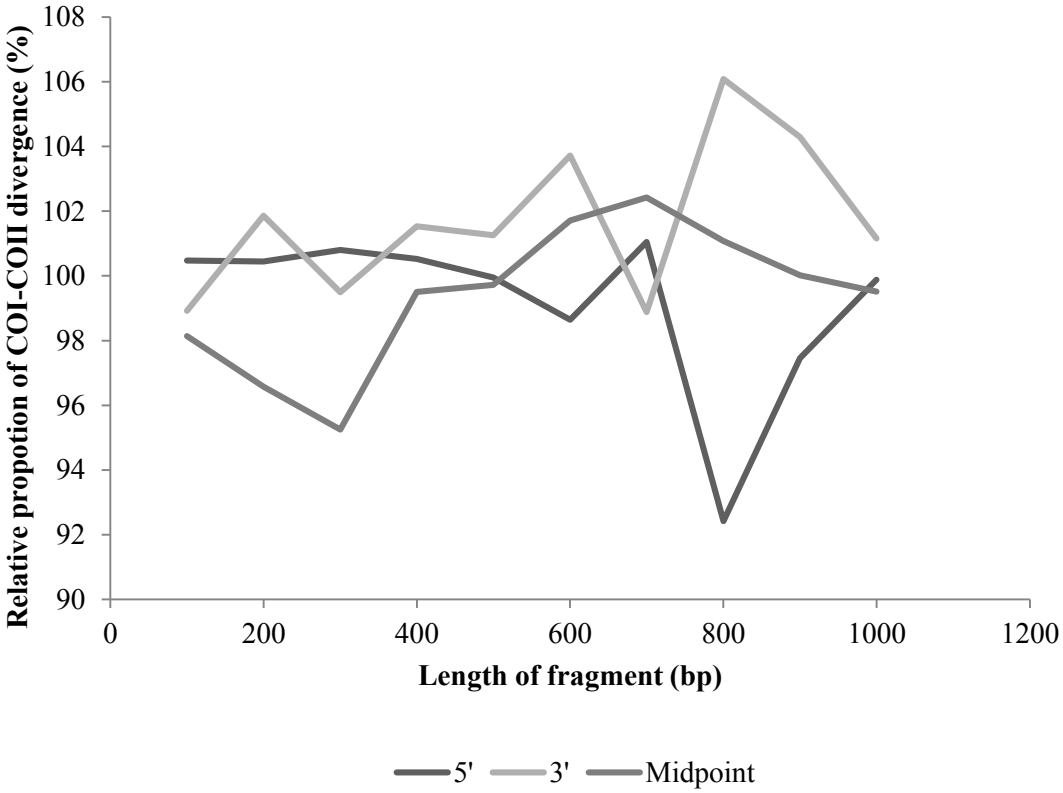

Figure 4.2: Percentage divergence for fragments increasing in length by 100 base pairs for concatenated COI-COII sequences. Fragments were increased starting from the 5' end, the 3' end or the midpoint of the whole sequence.

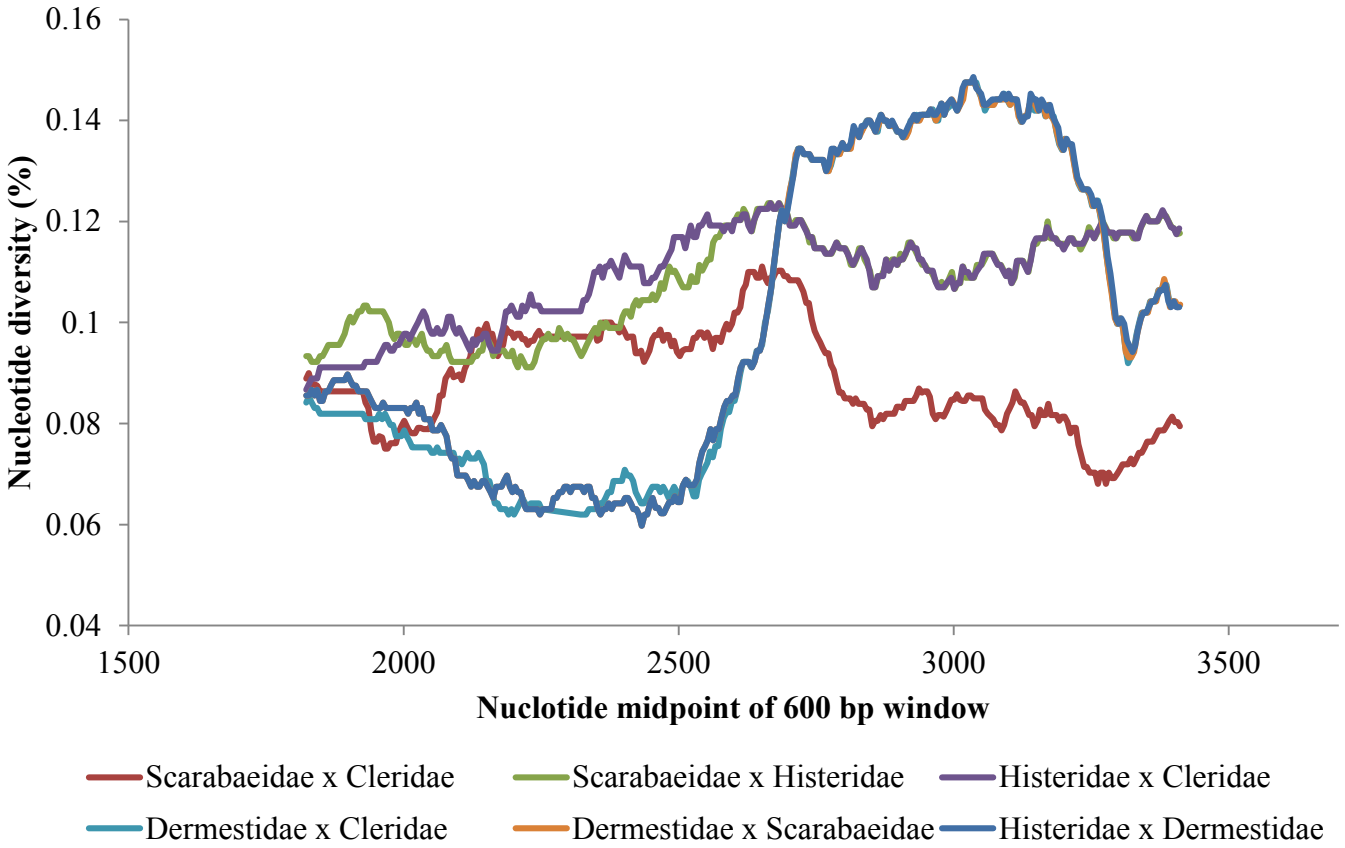

Figure 4.3: Sliding window analysis showing interfamily divergence across 4 carrion beetle families from region Dy# 1524-3710.

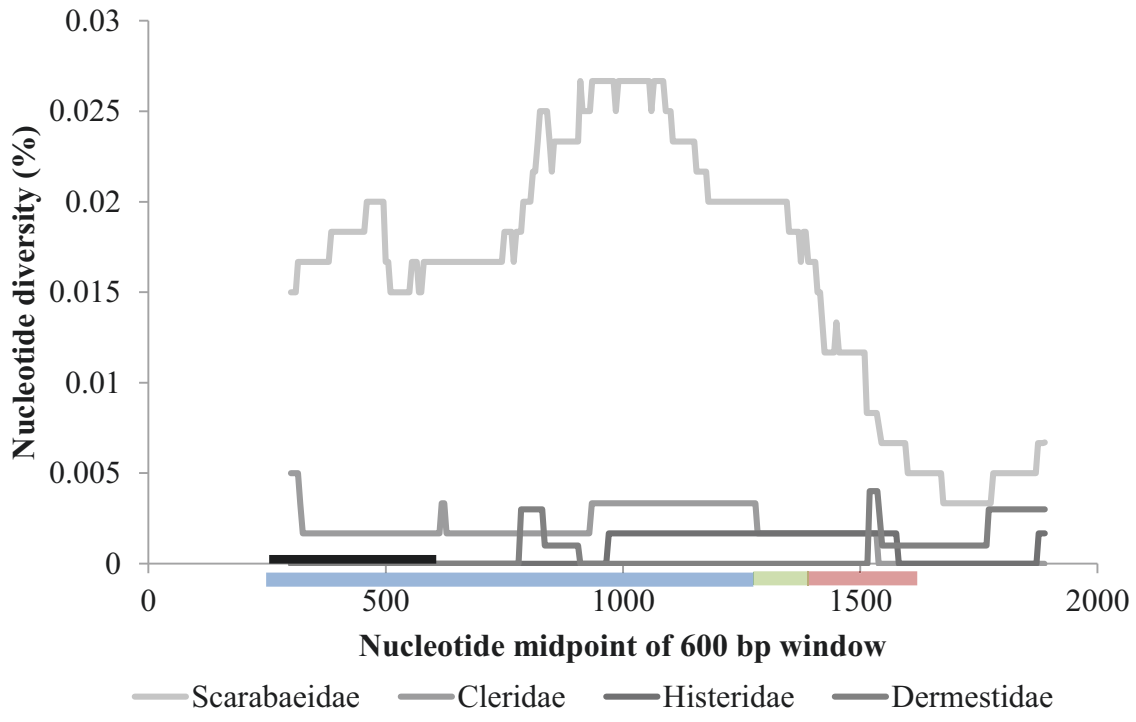

Figure 4.4: Interspecific nucleotide divergence for four species pairs over the COI-COII region. Sliding window profiles of 600 bp with a step size of 5 were used. The black line indicates the barcoding region. The blue region indicates the COI fragment used in this study, the green region indicates the region coding for leucine and the red region shows COII.

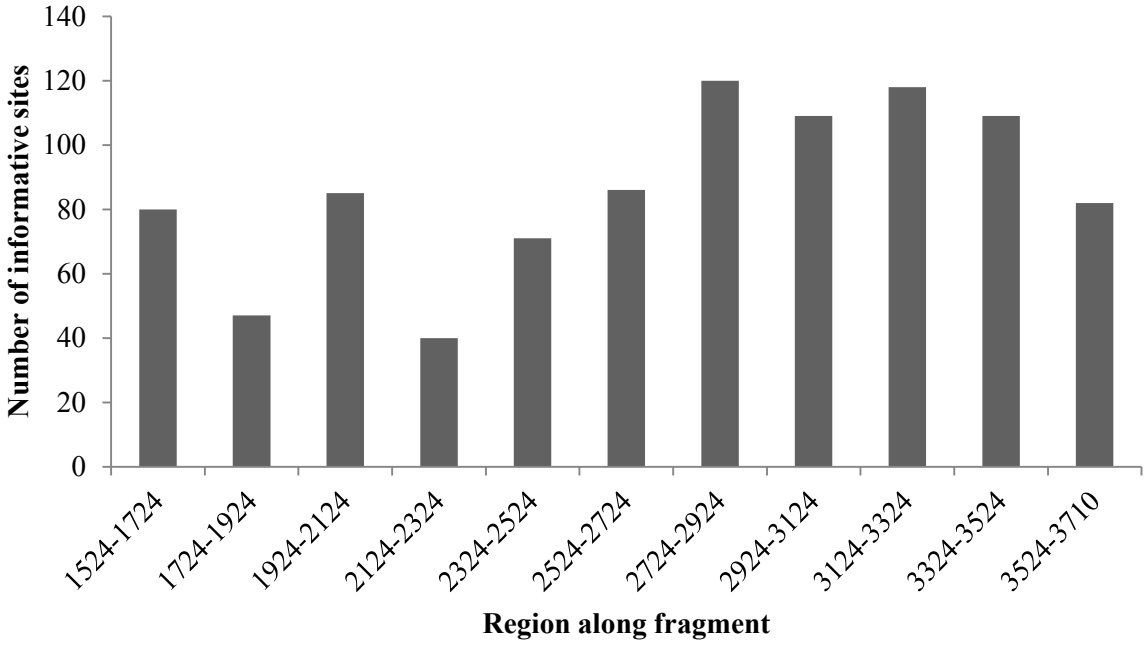

Figure 4.5: Mean number of informative sites per 200 bp along the 2189 bp region of COI-COII (Dy# 2724-3710) for the families sampled.

5 General Conclusion

5.1 Introduction

Various molecular methods have been used in insect taxonomy (see Chapter 1). The DNA barcoding approach (See Chapter 2) has been both celebrated and dismissed for its simplicity since Hebert *et al*. (2003a, b) encouraged its use. Barcoding has proved to be useful in species identification of many groups but as much as 25% of species in some groups cannot be delineated using mtDNA sequences (Funk & Omland, 2003; Elias *et al*., 2007; Weimers & Fiedler, 2007). This study set out to explore a number of issues relating to the use of barcode for identifying beetles as forensic evidence.

5.2 Providing barcodes

Firstly, do barcodes adequately identify beetles used for forensic purposes? The general answer is 'yes', but there are some taxa (e.g. Trogidae) that could not be successfully amplified, let alone identified. A 1.3% failure rate of identification at the family level was recorded and 39% of the specimens failed to amplify. Thus, this study showed that DNA barcoding does not consistently meet forensic standards (cf. Bandelt & Salas, 2012). However, the juvenile specimens were all linked to their adult stages correctly by their barcodes. Barcodes can provide a powerful tool for identifying immature stages.

One of the postulated benefits of barcoding was that it will allow inexperienced DNA taxonomists or trained technicians who are not specialist taxonomists to sequence and identify many animal species, reducing the workload of the world's comparatively few professional taxonomists. Unfortunately, primer failure can occur (e.g. in the Trogidae and Cleridae in this study) probably due to mutations in the primer binding sites. Amplification failure can also occur because of DNA degradation. Also, amplification protocols may require (sometimes extensive) optimising before a usable product can be sequenced. This is contrary to the barcoding aspiration that even non-specialists could apply the standard protocol.

5.3 Interpreting variation

It is often suggested that insects may provide evidence that a body has been moved long distances. Bergsten *et al.* (2012) showed DNA barcodes from conspecifics to vary on a global scale and could thus be used to indicate the species' origin. The geographical origin of the

species used in this study cannot be reliably inferred from their barcodes as haplotypes are shared between different locations within southern Africa (Chapter 4). Consequently, barcodes cannot be relied upon to determine the region of origin of a corpse if relocation is suspected. Microsatellites would be more appropriate here.

5.4 Placing species not represented in barcode databases

The barcode databases may also not hold a reference sample for an unidentified species at all. A limiting factor for barcoding seems to be a lack of sequences in reference libraries (Nagy *et al*., 2013), especially for infrequently studied organisms, and barcoding will only be truly assessed once there are a sufficient number of barcodes to be compared.

The BLAST and BOLD searches revealed forensically relevant beetle species to be poorly represented. The results presented here imply that there is no reliable 'barcoding gap' that might help to detect species not represented in the existing databases.

5.5 Comparing barcodes to other regions of COI and COII and mini-barcodes

Another aim was to examine the diversity across COI-COII and to determine whether a more informative region exists for species identification. For all but the clerids, diversity increased in the latter portion of COI extending into COII (Chapter 4). This suggests that a more robust barcode region exists downstream from the present one, at least for these beetles. This region may work successfully for other species and should be addressed in other members of the Coleoptera.

Approaches employing the 'mini-barcode' notion (using 100 to 180 bp of COI for identification) have been tested to overcome DNA degradation issues associated with museum specimens (Hajibabaei *et al*., 2006; Meusnier *et al*., 2008; Shokralla *et al*., 2011). The fragment length analysis performed in this study (see Chapter 4) showed short sequences (>200 bp) to be as effective at distinguishing among species as long sequences (>600 bp), which has good implications for forensic entomology since specimens with degraded DNA may yield a good mini-barcode that would reduce the cost of sequencing markedly.

5.6 Conclusions

There is no doubt that the DNA barcode has the power to delimit some species accurately, but the studies that have shown the barcode to work may lead to confirmation bias

on the part of barcoding advocates. The fact that barcodes *are* able to successfully discriminate some species incentivises them to show that it works for as many species as possible under standard conditions. The species that have not been successfully identified do not get as much attention and it is possible that many specimens that are not successful are ignored and omitted from publications. Focussing on only positive results skews perspectives on what is considered to be the power of barcoding. Barcoding proponents should document all of the failures - not only the sequences that failed to cluster with conspecifics, but also those that failed to amplify with the barcoding primers (taking DNA degradation into account, of course). This would give a more exact estimate of the accuracy of barcoding and how heavily it can be relied on, especially for legal purposes. Identification used to calculate a PMI needs to be unequivocal. "Best guesses" cannot be taken to court and forensic investigators will have to implement another means necessary to yield a compelling identification.

It is likely that barcodes will require verification with another gene sequence and a set of morphological, ecological and biogeographical cues as suggested by those who advocate an integrated approach to species discovery and identification (e.g. Johnson *et al*., 2005; Will *et al*., 2005; Damm *et al*., 2010; Paial *et al*., 2010; Schlick-Steiner *et al*., 2010) to be of indisputable forensic use. It may even be found that it is not possible to discriminate all animals based on a common region of DNA and that barcode regions specific to particular higher taxa will have to be discovered and assessed.

Taxonomy is a complex field which requires the inclusion of input other than genetic data, especially for species discovery (DeSalle *et al*., 2005). Hebert *et al*. (2003a, b) never intended barcoding to replace traditional taxonomy, only to complement it (Hebert & Gregory, 2005; Hajibabaei *et al*., 2007). An attempt at a shortcut for taxonomy, DNA barcoding may miss subtle differences that would only be noticed using the traditional morphological methods or more extensive sequencing. Lipscomb *et al*. (2003) was concerned that DNA barcoding would "reduce taxonomy to a mere technical service", but if used properly to augment other species data in integrated approaches, barcoding could be an extremely useful addition to taxonomy.

5.7 References

Al-Barrak, M., Loxdale, H.D., Brookes, C.P., Dawah, H.A., Biron, D.G., and Alsagair, O. 2004. Molecular evidence using enzyme and RAPD markers for sympatric evolution in British species of *Tetramesa* (Hymenoptera: Eurytomidae). *Biological Journal of the Linnean Society* 83: 509-525.

Alessandrini, F., Mazzabti, M., Onofri, V., Turchi, C., and Tagliabracci, A. 2008. MtDNA analysis for genetic identification of forensically important insects. *Forensic Science International: Genetics Supplement Series* 1: 584-585.

Aliabadian, M., Beentjes, K., Roselaar, K.C.S., van Brandwijk, H., Nijman, V., and Vonk, R. 2013. DNA barcoding of Dutch birds. *ZooKeys* 365:25–48.

Allendorf, F.W., 2012; Luikart, G., and Atken, S.N. 2012. *Conservation and the Genetics of Populations*. Wiley-Blackwell, Chichester. 602 pp.

Aly, M.S., Wen, J., Wang, X., and Cai, J. 2012. Cytochrome oxidase II gene "short fragment" applicability in identification of forensically important insects. *Romanian Journal of Legal Medicine* 20: 231–236.

Amendt, J., Campobasso, C.P., Gaudry, E., Reiter, E., Leblanc, H.N., and Hall, M.J.R. 2007. Best practice in forensic entomology - standards and guidelines. *International Journal of Legal Medicine* 121: 90-104.

Anderson, G.S., and Cervenka, V.J. 2002. Insects associated with the body: their use and analysis. In: Haglund, W.D., and Sorg, M.H., (eds), *Advances in Forensic Taxonomy - Method, Theory and Archaeological Perspectives*. CRC, Boca Raton, pp. 173–200.

Ardura, A., Planes, S., and Garcia-Vazquez, E. 2013. Applications of DNA barcoding to fish landings: authentication and diversity assessment. *ZooKeys* 365: 49–65.

Avise, J.C. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts. 447 pp.

Avise, J.C., 2004. *Molecular Markers, Natural History, and Evolution*, second ed. Sinauer Associates, Sunderland, Massachusetts. 684 pp.

Ball, S.L., Hebert, P.D.N., Burian, S.K., and Webb, J.M. 2005. Biological identification of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society* 24: 245–255.

Balloux, F. 2010. The worm in the fruit of the mitochondrial DNA tree. *Heredity* 104: 419-420.

Bandelt HJ, Salas, A. 2012. Current next generation sequencing technology may not meet forensic standards. *Forensic Science International: Genetics* 6: 143–145.

Beckenbach, A.T., Wei, Y.W., and Liu, H. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Molecular Biology and Evolution* 10: 619--634.

Beebe, N.W., Whelan, P.I., Hurk, A. F. V. D., Ritchie, A.A., Corcoran, S., and Cooper, R.D. 2007. A polymerase chain reaction-based diagnostic to identify larvae and eggs of

container mosquito species from the Australian region. *Journal of Medical Entomology* 44: 376–380.

Benecke, M., 1998. Random amplified polymorphic DNA (RAPD) typing of necrophageous insects (Diptera, Coleoptera) in criminal forensic studies: validation and use in practice. *Forensic Science International* 98: 157–68.

Benecke, M., 2001. A brief history of forensic entomology. *Forensic Science International* 120: 2–14.

Bensasson, D., Zhang, D. and Hartl, D.L. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecological Evolution* 16: 314–321.

Bergsten, J.O.B., Ilton, D.A.T.B., Ujisawa, T.O.F., Lliott, M.I.E., Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G. and Vogler, A.P. 2012. The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology* 61: 851–69.

Bickford, D., Lohman, D. J., Sodhi, N. S., Meier, R., Winker, K., Ingram,K.K., and Das, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22: 148–155.

Blaxter, M. 2003 Counting angels with DNA. *Nature* 421: 122–124.

Boehme, P., Amendt, J., and Zehner, R. 2011. The use of COI barcodes for molecular identification of forensically important fly species in Germany. *Parasitology Research* 110: 2325-2332.

Bouchard, P.B., Grebennikov, V.V., Smith, A.B.T., and Douglas, H. 2009. Biodiversity of Coleoptera. In: Footit, R.G., and Adler, P.H., (eds), *Insect Biodiversity – Science and Society*. Wiley-Blackwell, Chichester. pp. 265-303.

Bourel, B., Tournel, G., Hédoin, V., Goff, M.L., and Gosset,D. 2001. Determination of drug levels in two species of necrophagous Coleoptera reared on substrates containing morphine. *Journal of Forensic Sciences* 46: 600–603.

Bourel, B., Tournel, G., Hédouin, V., and Gosset, D. 2004. Entomofauna of buried bodies in northern France. *International Journal of Legal Medicine* 118: 215–220.

Braack, L.E.O. 1981. Visitation patterns of principal species of the insect-complex at carcasses in the Kruger National Park. *Koedoe* 24: 33–49.

Braack, L.E.O. 1986. Arthropods associated with carcasses in the northern Kruger National Park. *South African Journal of Wildlife Research* 16: 91–98.

Braack, L.E.O., and de Vos, V. 1987. Seasonal abundance of carrion-frequenting blow-flies (Diptera: Calliphoridae) in the Kruger National Park. *Onderstepoort Journal of Veterinary Research* 54: 591-597.

Brohede, J., Primmer, C.R., Moller, A., and Ellegren, H. 2002. Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Research* 30: 1997–2003.

Brown, S., McLaughlin, W., Jerez, I.T. and Brown, J. K. 2002 Identification and distribution of *Bemisia tabaci* (Gennadius; Homoptera: Aleyrodidae) haplotypes in Jamaica. *Journal of Tropical Agriculture* 79: 140–149.

Brunner, P.C., Fleming, C. and Frey, J.E. 2002 A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR–RFLP-based approach. *Agricultural and Forest Entomology* 4: 127–136.

Buhay, J.E. 2009. "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology* 29: 96–110.

Burns, J.M., Hanzen, D.J., Hajibabei, M. and Hallwachs, W. 2010. Of many similar species in the Neotropical genus *Porphyrogenes* (Lepidoptera: Hesperidae), a new one, repeatedly reared in Costa Rica, is relatively distinct. *Proceedings of the Entomological Society of Washington* 112: 32–42.

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W., and Hebert, P.D.N. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservation Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences* 105: 6350-6355.

Byrd, J.H., and Castner, J.L., 2001. Insects of Forensic Importance. In: Byrd, J.H., and Castner, J.L., (eds), *Forensic Entomology: The utility of arthropods in legal investigations*. CRC Press. Boca Raton, pp 43 – 80.

Cameron, S., Rubinoff, D., and Will, K. 2006. Who will actually use DNA barcoding and what will it cost? *Systematic Biology* 55: 844–847.

Campobasso, C.P., Di Vella, G., and Introna, F. 2001. Factors affecting decomposition and dipteran colonization. *Forensic Science International* 120: 18–27.

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., and De Mattia, F. 2010. DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in Bioinformatics* 11: 440–453.

Caterino, M.S., and Tishechkin, A.K. 2006. DNA identification and morphological description of the first confirmed larvae of Hetaeriinae (Coleoptera: Histeridae). *Systematic Entomology* 31: 405–418.

Caterino, M.S., Cho, S., and Sperling, F.A.H. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology* 45: 1-54.

Caterino, M.S., Vogler, A.P. 2002. The phylogeny of the Histeroidea (Coleoptera: Staphyliniformia). *Cladistics* 18: 394-415.

Catts, E.P., and Goff, M.L. 1992. Forensic entomology in criminal investigations. *Annual Review of Entomology* 37: 253–272.

Catts, E.P., and Haskell N.H. 1990. *Entomology and death: A procedural guide*. Joyce's Print Shop, Clemson, South Carolina.180 pp.

Choat, H. 2006. Phylogeography and reef fishes: bringing ecology back into the argument. *Journal of Biogeography* 33: 967–968.

Clary, D.O., and Wolstenholme, D.R. (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *Journal of Molecular Evolution* 22: 252-271.

Clement, M., Posada, D., and Crandall, K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1660.

Coleman, A.W., and Vacquier, V.D., 2002. Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*). *Journal of Molecular Evolution* 54: 246– 257.

Collins, R.A., and Cruickshank, R.H. 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* 13: 969-75.

Collins, R.A., Armstrong, K.F., Meier, R., Yi, Y., Brown, S.D.J., Cruickshank, R.H., Keeling, S., and Johnston, C., 2012. Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. *PLoS One* 7, e28381.

Cowan, R.S., Chase, M.W., Kress, W.J., and Savolainen, V. 2006. 300,000 Species to identify: Problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55: 611–616.

Cox, K., Thomaes, A., Antonini, G., Zilioli, M., De Gelas, K., Harvey, D., Solano, E., Audisio, P., McKeown, N., Shaw, P., Minetti, R., Bartolozzi, L., and Mergeay, J. 2013. Testing the performance of a fragment of the COI gene to identify western Palaearctic stag beetle species (Coleoptera, Lucanidae). *ZooKeys* 365: 105–26.

Coyne, J.A., and Orr, H.A. 2004. *Speciation*. Sinauer and Associates, Sunderland, Massachusetts. 545 pp.

Damm, S., Schierwater, B., and Hadrys, H. 2010. An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Molecular Ecology* 19(18): 3881–93.

Dapporto, L. 2007. Cuticular lipid diversification in *Lasiommata megera* and *Lasiommata paramegaera*: the influence of species, sex, and population (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* 91: 703–710.

Dasmahapatra, K.K., and Mallet, J. 2006. Taxonomy: DNA barcodes: Recent successes and future prospects. *Heredity* 97:254–255.

Dayrat, B., 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–415.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard, D.J., and Berlocher, S.H., (eds), *Endless forms: Species and speciation*. Oxford University Press., Oxford. pp. 57–75.

Derocles, S.A.P., Le Ralec, A., Plantegenest, M., Chaubet, B., Cruaud, C., Cruaud, A. and Rasplus, J.-Y. 2012. Identification of molecular markers for DNA barcoding in the Aphidiinae (Hym. Braconidae). *Molecular Ecology Resources* 12: 197–208.

Derycke, S., Vanaverbeke, J., Rigaux, A., Backeljau, T. and Moens, T. 2010. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. *PloS One* 5: e13716.

DeSalle R. 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. *Conservation Biology* 20: 1545-1547.

DeSalle, R., and Templeton, A.R. 1988. Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila*. *Evolution* 42: 1076-1084.

DeSalle, R., Egan, M. G. and Siddall, M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B* 360: 1905–1916.

Dowling, D.K., Friberg, U., and Lindell, J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology and Evolution* 23: 546–554.

Eaton, M.J., Meyers, G.L., Kolokotronis, S., Leslie, M.S., Martin, A.P., and Amato, G. 2009. Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. *Conservation Genetics* 11: 1389–1404.

Ekrem, T., Willassen, E., and Stur, E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* 43: 530–542.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J., and Jiggins, C.D. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B* 274: 2881–2889.

Emerson, B. C., and Wallis, G. P. 1995. Phylogenetic relationships of the Prodontria (Coleoptera; Scarabaeidae; Melolonthinae), derived from sequence variation in the mitochondrial cytochrome oxidase II gene. *Molecular Phylogenetics and Evolution* 4: 433–447.

Floyd, R., Abebe, E., Papert, A., and Blaxter, M. 2002. Molecular barcodes for soil nematode identification. *Molecular Ecology* 11: 839–850.

Floyd, R.M., Wilson, J.J., and Hebert, P.D.N. 2009. DNA barcodes and insect biodiversity. In: Foottit, R.G., and Adler, P.H., (eds), *Insect Biodiversity: Science and Society*. Wiley-Blackwell, Chichester. pp. 417–432.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3(5): 294–9.

Fontaine, K.M., Cooley, J.R., and Simon, C. 2007. Evidence for paternal leakage in hybrid periodical cicadas (Hemiptera: *Magicicada* spp.). *PLoS One* 2: e892.

Freedman, J. 2005. *The Applications and Limitations of Taxonomy (in Classification of Organisms): An Anthology of Current Thought*. The Rosen Publishing Group, New York. 158 pp.

Frézal, L., and Leblois, R. 2008. Four years of DNA barcoding: current advances and prospects - infection, genetics and evolution. *Journal of Molecular Eepidemiology and Evolutionary Genetics in Infectious Diseases* 8: 727–36.

Frolov, A.V. and Scholtz, C.H. 2005. Revision of the southern African genus *Frankenbergerius* Balthasar with description of new taxa (Coleoptera: Scarabaeidae: Scarabaeinae). *Journal of Natural History* 39: 2355–2377.

Fujita, M.K., Leach, E.A.D., Burbrink, F.T., McGuire, J.A., and Moritz, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* 27: 480–488.

Funk, D.J., and 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.

Galtier, N., Nabholz, B., Glémin, S., and Hurst, G.D.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* 18: 4541–50.

Gissi, C., Reyes, A., Pesole, G., and Saccone, C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. *Molecular Biology and Evolution* 17: 1022–1031.

Godfray, H.C.J. 2002. Challenges for taxonomy. *Nature* 417: 17–19.

Goldstein, P.Z., and DeSalle, R. 2010. Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *Bioessays* 33: 135–147.

Gorham, J.R. 1987. Insect and Mite Pests in Food: an Illustrated Key. *U. S. Department of Agriculture*, Handbook No 665.

Gossner, M.M., and Hausmann, A. 2009. DNA barcoding enables the identification of caterpillars feeding on native and alien oak (Lepidoptera: Geometridae). *Mitteilungen der Münchner Entomologischen Gesellschaft* 99: 135-140.

Grimaldi, D., and Engel, M.S. 2005. *Evolution of the Insects*. Cambridge University Press, New York. 390 pp.

Gullan, P.J., and Cranston, P.S. 2010. *The Insects: an Outline of Entomology*. Wiley-Blackwell, Chichester. 565 pp.

Guo, Y., Zha, L., Yan, W., Li, P., Cai, J., and Wu, L. 2013. Identification of forensically important sarcophagid flies (Diptera: Sarcophagidae) in China based on COI and period gene. *International Journal of Legal Medicine* 128: 221–8.

Guo, Y.D., Cai, J.F., Li, X., Xiong, F., Su, R.N., Chen, F.L., Liu, Q.L., Wang, X.H., Chang, Y.F., Zhong, M., Wang, X., and Wen, J.F. 2010. Identification of the forensically important sarcophagid flies *Boerttcherisca peregrina*, *Parasarcophaga albiceps* and *Parasarcophaga dux* (Diptera: Sarcophagidae) based on COII gene in China. *Tropical Biomedicine* 27: 451-60.

Gyllensten, U., Wharton, D., Josefsson, A., and Wilson, A.C. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* 352: 255– 257.

Hajibabaei, M., de Waard, J. R., Ivanova, N. V., Ratnasingham, S., Dooph, R. T., Kirk, S. L., Mackie, P. M. and Hebert, P. D. N. 2005. Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B* 360: 1959–1967.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences* 103: 968–971.

Hajibabaei, M., Singer, G.A., and Hebert, P.D.N. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23: 167–172.

Hall, A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41**:** 95-98.

Hall, H.G. 1990. Parental analysis of introgressive hybridization between African and European honeybees using nuclear DNA RFLPs. *Genetics* 125: 611- 621.

Hanner, R., Becker, S., Ivanova, N.V. and Steinke, D. 2011. FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* 22: 106-122.

Harrison, R. G. 1998. Linking evolutionary pattern and process. The relevance of species concepts for the study of speciation. In: Howard, D.J., and Berlocher, S.H., (eds) *Endless forms: species and speciation*. Oxford University Press, Oxford. pp. 19–31.

Hazkani-Covo, E., Zeller, R.M. and Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (NUMTs) in sequenced nuclear genomes. *PLoS Genetics* 6: p.e1000834.

Hebert P.D.N., Stoeckle, M.Y., Zemlak T S, and Francis, C.M. 2004b. Identification of birds through DNA barcodes. *Plos Biology* 2: 1657-1663.

Hebert, P.D.N., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54: 852–859.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society* 270: 313-321.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14812–14817.

Hebert, P.D.N., Ratnasingham, S., and deWaard, J.R. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society* 270: S96-S99.

Henn, B.M.M Ginoux, C.R., and Feldman, M.W. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Molecular Biology and Evolution* 26: 217–230.

Hinton, H.E. 1945. *A monograph of the beetles associated with stored products*. London: British Museum (Natural History).

Ho, S.Y., Phillips, M.J., Cooper, A., and Drummond, A.J. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22: 1561–1568.

Hoeh, W.R., Blakely, K.H., and Brown, W.M. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251: 1488–1490.

Huelsenbeck, J.P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Hunt, T., Bergsten, J., Levkanicova, Z., St John, O., Wild, R., Hammond, P., Ahrens, D., Balke, M., Caterino, M. S., Gómez- Zurita, J., Ribera, I., Barraclough, T. G., Bocakova, M., Bocak, L., and Vogler, A. P. 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318: 1913–1916.

Hurst, G.D.D., and Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B* 272: 1525–1534.

Huson, D.H., and Bryant, D. 2006: Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Ikeda, H., Nishikawa, M. and Sota, T., 2012. Loss of flight promotes beetle diversification. *Nature Communications* 3: 648.

Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708–713.

Introna, F., Campobasso, C.P., and Goff, M.L. 2001. Entomotoxicology. *Forensic Science International* 120: 42–47.

Jinbo, U., Kato, T., and Ito, M., 2011. Current progress in DNA barcoding and future implications for entomology. *Entomological Science* 14: 107–124.

Johnson, N.K., and Cicero, C. 2002. The role of ecological diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): Multigene evidence from mtDNA. *Molecular Ecology* 11: 2065–2081.

Johnson, S.B., Waren, A. & Vrijenhoek, R.C. 2008. DNA barcoding of *Lepetodrilus* limpets reveals cryptic species. *Journal of Shellfish Research* 27: 43–51.

Jones, Y.L., Peters, S.M., Weland, C., Ivanova, N. V and Yancy, H.F. 2013. Potential use of DNA barcodes in regulatory science : identification of the U.S. Food and Drug

Administration's "Dirty 22 " contributors to the spread of foodborne pathogens. *Journal of Food Production* 76: 144–149.

Junqueira, A.C.M., Lessinger, A.C., Azeredo-Espin, A.M.L. 2002. Methods for the recovery of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Medical and Veterinary Entomology* 16: 39–45.

Kartavtsev, Y.P., and Lee J-S. 2006. Analysis of nucleotide diversity at the cytochrome *B* and cytochrome oxidase 1 genes at the population, species, and genus levels. *Russian Journal of Genetics* 42: 341-362.

Kengne, P., Antonio-Nkondjio, C., and Awono-Ambene, H.P. 2007. Molecular differentiation of three closely related members of the mosquito species complex, *Anopheles moucheti*, by mitochondrial and ribosomal DNA polymorphism. *Medical and Veterinary Entomology* 21: 177-182.

Khemakhem, M. M., Hanem, M. and Mohamed, M. 2002. PCR–RFLP identification of mitochondrial markers of *Mayetiola* (Diptera: Cecidomyiidae) on cereals. *Annales Societe Entomologique de France* 38: 277–282.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.

Kocárek, P. 2003. Decomposition and Coleoptera succession on exposed carrion of small mammal in Opava, the Czech Republic. *European Journal of Soil Biology* 39: 31-45.

Kohlmayr, B., Riegler, M., Wegensteiner, R. and Stauffer, C. 2002. Morphological and genetic identification of the three pine pests of the genus *Tomicus* (Coleoptera Scolytidae) in Europe. *Agricultural and Forrest Entomology* 4: 151–157.

Kovarik, P.W., and Caterino, M.S. 2001. Histeridae. In: Arnett, R.H., and Thomas, M.C., (eds), *American Beetles. Volume 1: Archostemata, Myxophaga, Adephaga, Polyphaga: Staphyliniformia*. CRC Press LLC, Boca Raton. pp. 212-227.

Kulshrestha, P., and Satpathy, D.K. 2001. Use of beetles in forensic entomology. *Forensic Science International* 120: 15-17.

Kurland, C.G., Canback, B., and Berg, O.G. 2003. Horizontal gene transfer: a critical view. *Proceedings of the National Academy of Sciences* 100: 9658–9662.

Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–5.

Laiho, J., and Stahls, G. 2013. DNA barcodes identify Central-Asian *Colias* butterflies (Lepidoptera, Pieridae). *ZooKeys* 365: 175–96.

Lefebure, T., Douady, C.J., Gouy, M., and Gibert, J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* 40: 435-447.

Li, X., Schuler, M.A., and Berenbaum, M.R. 2007. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology* 52: 231e253.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Linares, M.C., Soto-Calderón, I.D., Lees, D.C. and Anthony, N.M. 2009. High mitochondrial diversity in geographically widespread butterflies of Madagascar: a test of the DNA barcoding approach. *Molecular Phylogenetics and Evolution* 50: 485–95.

Little, D.P. 2011. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS ONE* 6: e20552.

Little, D.P., and Stevenson, D.W. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 23: 1–21.

Lopez, J.V., Culver, M., Stephens, J.C., Johnson, W.E. and O'brien, S.J. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Molecular Biology and Evolution* 14: 277–286.

Lowenstein, J.H., Burger, J., Jeitner, C.W., Amato, G., Kolokotronis, S.O., and Gochfeld, M. 2010. DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. *Biology Letters* 6: 692–695.

Loxdale, H.D., and Lushai, G., 2009. Molecular markers in entomology. *Bulletin of Entomological Research* 88: 577.

Malgorn, Y., and Coquoz, R.1999. DNA typing for identification of some species of Calliphoridae: an interest in forensic entomology. *Forensic Science International* 102: 111-119.

Mardulyn, P., and Whitfield, J. B. 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids. *Molecular Phylogenetic Evolution* 12: 282–294.

Mardulyn, P., Milinkovitch, M.C., and Pasteels, J.M. 1997. Phylogenetic analyses of DNA and allozyme data suggest that *Gonioctena* leaf beetles (Coleoptera; Chrysomelidae)

experienced convergent evolution in their history of host-plant family shifts. *Systematic Biology* 46: 722–747.

Martin, A.P. 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. *Molecular Biology and Evolution* 12: 1124–1131.

May, R.M., and Harvey, P.H. 2009. Species uncertainties. *Science* 323: 687.

Mayden, R.L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge, M.F., Dawah, H.A. and Wilson, M.R., (eds), *Species: The units of biodiversity*. Chapman and Hall, London. pp. 381-424.

Mayr, E., 1942. *Systematics and the Origin of Species*. Columbia University Press, New York. 463 pp.

McDonagh, L., Thornton, C., Wallman, J.F., and Stevens, J.R., 2009. Development of an antigen-based rapid diagnostic test for the identification of blowfly (Calliphoridae) species of forensic significance. *Forensic Science International Genetics* 3: 162–5.

McKelvey, B. 1982. *Organizational Systematics – Taxonomy, Classification, Evolution*. University of California Press. 511 pp.

Meier, R., Shiyang, K., Vaidya, G., and Ng, P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–28.

Meier, R., Zhang, G. and Ali, F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology* 57: 809–13.

Meiklejohn, K.A., Wallman, J.F., and Dowton, M. 2013. DNA barcoding identifies all immature life stages of a forensically important flesh fly (Diptera: Sarcophagidae). *Journal of Forensic Sciences* 58: 184–7.

Meusel, M.S., and Moritz, R.F.A. 1993. Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. *Current Genetics* 24: 539-543.

Meusnier, I., Singer, G.A.C., Landry, J.-F., Hickey, D.A., Hebert, P.D.N and Hajibabaei, M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9: 214.

Meyer, C.P., and Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* 3: 2229–2238.

Michu, E. 2007. A short guide to phylogeny reconstruction. *Plant, Soil and Environment* 2007: 442–446.

Midgley J.M., Richards, C.S., and Villet. M.H. 2010. The utility of Coleoptera in forensic investigations. In: Amendt, J., Campobasso, C.P., Goff, M.L., and Grassberger, M., (eds) *Current Concepts in Forensic Entomology*. Heidelberg: Springer. pp. 57-68.

Midgley, J.M., and Villet, M.H. 2008. The effect of killing method on post-mortem change in length of larvae of *Thanatophilus micans* (Fabricius, 1794) (Coleoptera: Silphidae) stored in 70% ethanol. *International Journal of Legal Medicine* 123: 103-108.

Midgley, J.M., Collett, I.J., and Villet, M.H. 2012. The distribution, habitat, diet and forensic significance of the scarab *Frankenbergerius forcipatus* (Harold, 1881) (Coleoptera: Scarabaeidae). *African Invertebrates* 53: 745-749.

Miller, M.A., Pfeiffer, W., and Schwartz, T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE),* New Orleans .pp 1-8.

Monaghan, M.T., Wild, R., and Elliot, M. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* 58: 298–311.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. 2011. How many species are there on Earth and in the ocean? *PLoS Biology* 9: e1001127.

Moulton, M.J., Song, H., and Whiting, M.F. 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology* 10: 615-627.

Mourier, T., Hansen, A.J., Willerslev, E., and Arctander, P. 2001. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Molecular Biology and Evolution* 18: 1833–1837.

Nagy, Z.T., Sonet, G., Mortelmans, J., Vandewynkel, C., and Grootaert, P. 2013. Using DNA barcodes for assessing diversity in the family Hybotidae (Diptera, Empidoidea). *ZooKeys* 365: 263–78.

Nei, M. 1987. *Molecular evolutionary genetics*. New York, Columbia University Press. 186 pp.

Nelson, L.A., Lambkin, C.L., Batterham, P., Wallman, J.F., Dowton, M., Whiting, M.F., Yeates, D.K., and Cameron, S.L. 2012. Beyond barcoding: a mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene* 511: 131–42.

Newton, A.F., Thayer, M.K., Ashe, J.S., and Chandler, D.S., 2000. Staphylinidae Latreille, 1802. In: Arnett, R.H., and Thomas, M.C., (eds), *American Beetles, vol. 1,*

Archostemata, Myxophaga, Adephaga, Polyphaga: Staphyliniformia. CRC Press, LLC, Boca Raton. pp. 272–418.

Ogden, T.H., and Rosenberg, M.S. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Molecular Phylogenetics and Evolution*, 42: 817–26.

Paial, J.M., Miralles, A., De la Riva, I., and Vences, M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* 7: 1–14.

Pamilo, P., Viljakainen, L., Vihavainen, A. 2007. Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* 24: 1340-1346.

Paquin, P., and Hedin, H. 2004. The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Cicurina* (Araneae: Dictynidae) from Texas caves. *Molecular Ecology* 13: 3239-3255.

Park, S. H., Zhang, Y, Piao, H., Yu, D.H., Jeong, G.Y., Yoo, U., Chung, T.H., Jo, H.G., and Hwang, J.J. 2009. Use of cytochrome c oxidase subunit I (COI) nucleotide sequences for identification of the Korean Luciliinae fly species (Diptera: Calliphoridae) in forensic investigations. Journal of Korean Medicine *Science* 24: 1058-1063.

Parr, R.L., Maki, J., Reguly, B., Dakubo, G.D., Aguirre, A., Wittock, R., Robinson, K., Jakupciak, J.P., and Thayer, R.E. 2009. The pseudo-mitoochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics* 7: 185.

Parsons, Y.M., and Shaw, K.L. 2001. Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology* 10: 1765–1772.

Peacock, E.R. 1993. Adults and larvae of hide, larder and carpet beetles and their relatives (Coleoptera: Dermestidae) and of derodontid beetles (Coleoptera: Derodontidae). In: Dolling, W.R., (ed), *Handbooks for the identification of British insects*. Royal Entomological Society of London, London. 144 pp.

Peng, W., Lin, H., Chen, C., and Wang, C. 2003. DNA identification of two laboratory colonies of the weevils, *Sitophilus oryzae* (L.) and *S. zeamais* Motschulsky (Coleoptera: Curculionidae) in Taiwan. *Journal of Stored Products Research* 39: 225–235.

Penman, D.J., Mair, G.C., McAndrew, B.J., Hussain, M.G., Basavaraju, Y., and Tuan, P.H. 2002. Genetic management of non-indigenous carps of low-input aquaculture systems in Asia. *Aquatic News* 28: 19–20.

Pons, T., Gonzalez, B., Ceciliani, F., and Galizzi, A. 2006. FlgM anti- sigma factors: identification of novel members of the family, evolutionary analysis, homology

modelling, and analysis of sequence-structure-function relationships. *Journal of Molecular Modelling* 12: 973-983.

Powell, J. R. 1994. Molecular techniques in population genetics: a brief history. In: Schierwater , B.B., Streit, G.P., Wagner, and DeSalle, R., (eds), *Molecular Ecology and Evolution: Approaches and Applications*. Burkhauser Verlag , Basel, Switzerland . pp. 131-156

Preativatanyou, K., Sirisup, N., Payungporn, S., Poovorawan, Y., Thavara, U., Tawatsin, A., Sungpradit, S., and Siriyasatien, P. 2010. Mitochondrial DNA- based identification of some forensically important blowflies in Thailand. *Forensic Science International* 202: 97-101.

Prins, A.J. 1983. Morphological and biological notes on some South African arthropods associated with decaying organic matter. I: Chilopoda, Diplopoda, Arachnida, Crustacea, and Insecta. *Annals of the South African Museum* 92: 53–112.

Prins, A.J. 1984a. Morphological and biological notes on some South African arthropods associated with decaying organic matter II: the predatory families Carabidae, Hydrophilidae, Histeridae, Staphylinidae and Silphidae (Coleoptera). *Annals of the South African Museum* 92: 295–356.

Prins, A.J. 1984b. Morphological and biological notes on some South African arthropods associated with decaying organic matter. III: the families Dermestidae, Cantharidae, Melyridae, Tenebrionidae, and Scarabaeidae (Coleoptera). *Annals of the South African Museum* 94: 203–304.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21: 1864–1877.

Quesada, H., Wenne, R., and Skibinski, D.O.F. 1995. Differential introgression of mitochondrial DNA across species boundaries within the marine mussel genus *Mytilus*. *Proceedings of the Royal Society* 262: 51-56.

Rajendran. S., and Hajira Parveen, K.M. 2005. Insect infestation in stored animal products. *Journal of Stored Product Research* 41: 1-30.

Ratnasingham, S., and Hebert, P.D.N. 2007. BOLD: the barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364.

Ratnasingham, S., and Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. *PLoS ONE* 8: e66213.

Raupach, M.J., Astrin, J.J., Hannig, K., Peters, M.K., Stoeckle, M.Y., and Wägele, J.-W., 2010. Molecular species identification of Central European ground beetles (Coleoptera:

Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* 7: 26.

Ridgeway, J., Midgley, J.M., Collett, I.J., and Villet, M.H. 2014. Advantages of using development models of the carrion beetles *Thanatophilus micans* (Fabricius) and *T. mutilatus* (Castelneau) (Coleoptera: Silphidae) for estimating minimum post mortem intervals, verified with case data. *International Journal of Legal Medicine* 128: 207-220.

Roe, A.D. and Sperling, F.A.H. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* 44: 325–45.

Ronquist, F., and Huelsenbeck, J.P. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Rot, C., Goldfarb, I., Ilan, M., and Huchon, D. 2006. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evolutionary Biology* 6: 71.

Rubinoff, D., Cameron, S., and Will, K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for 'barcoding' identification. *Journal of Heredity* 97: 581–594.

Ruiz, F.J. 2010. Systematics of the *Anopheles Nyssorhynchus Oswaldoi* group in Latin America. PhD Thesis, Canterbury Christ Church University.

Schawaller, W. 1981. Taxonomie und Faunistik der Gattung *Thanatophilus* (Coleoptera: Silphidae). Stuttgarter Beiträge zur Naturkunde A (Biologie) 351: 1-21.

Schawaller, W. 1987. Faunistische und systematische Daten zur Silphiden-Fauna Südafrikas (Coleoptera, Silphidae). *Entomofauna* 8: 277-288.

Schilthuizen, M., Scholte, C., Wijk, R.E.J. Van, Dommershuijzen, J., Horst, D. Van Der, Meijer, M., Lievers, R., Groenenberg, D.S.J., van Wijk, R.E.J., van der Horst, D., and zu Schlochtern, M.M. 2011. Using DNA-barcoding to make the necrobiont beetle family Cholevidae accessible for forensic entomology. *Forensic Science International* 210: 91–5.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R.H. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 431–438.

Scholtz CH. 1983. A new species of *Trox* F. (Coleoptera: Trogidae) from Senegal with new distribution records for other African species. *Journal of the Entomological Society of Southern Africa* 46: 205-208.

Scholtz, C.H. 1980. Monograph of the genus *Trox* F. (Coleoptera: Trogidae) of subsaharan Africa. *Cimbebasia Memoir* 4: 1-104.

Scholtz, C.H. 1982. Catalogue of world Trogidae (Coleoptera: Scarabaeoidea). *Entomology Memoir of the Department of Agriculture of the Republic of South Africa* 54: 1-27.

Sevilla, R., Diez, A., and Noren, M. 2007. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Molecular Ecology Notes* 7: 730–734.

Shokralla, S., Zhou, X., Jansen, D.H., Hallwachs, W., Ladry, J.-F., Jacobus, L.M., and Hajibabei, M. 2011. Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE* 7: 1–4.

Shufran, K.A., Burd, J.D., Anstead, J.A., and Lushai, G. 2000. Mitochondrial DNA sequence divergence among greenbug (Homoptera: Aphididae) biotypes: evidence for host-adapted genes. *Insect Molecular Biology* 9: 179–184.

Simon, C., Buckley, T.T., Frati, F., Stewart, J.B. and Beckenbach, A.T. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics* 37: 545–579.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and 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.

Smit, J., Reijnen, B., and Stokvis, F. 2013. Half of the European fruit fly species barcoded (Diptera, Tephritidae); a feasibility test for molecular identification. *ZooKeys* 365: 279–305.

Smith, K.G.V. 1986. *A Manual of Forensic Entomology*. British Museum (Natural History), London, United Kingdom. 205 pp.

Sonet, G., Jordaens, K., Braet, Y., and Desmyter, S. 2012. Why is the molecular identification of the forensically important blowfly species *Lucilia caesar* and *L. illustris* (family Calliphoridae) so problematic? *Forensic Science International* 223: 153–159.

Sonet, G., Jordaens, K., Braet, Y., Bourguignon, L., Dupont, E., Backeljau, T., de Meyer, M., and Desmyter, S. 2013. Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France. *ZooKeys* 365: 307–28.

Song, Z., Wang, X., and Liang, G. 2008. Species identification of some common necrophagous flies in Guangdong province, southern China based on the rDNA internal transcribed spacer 2 (ITS2). *Forensic Science International* 175: 17-22.

Sota, T., Ishikawa, R., Ujiie, M., Kusumoto, F., and Vogler, A.P. 2001. Extensive trans-species mitochondrial polymorphisms in the carabid beetles *Carabus* subgenus *Ohomopterus* cause by repeated introgressive hybridization. *Molecular Ecology* 10: 2833-2847.

Sperling, F.A.H., and Roe, A.D. 2009. Molecular dimensions of insect taxonomy. In: Foottit, R.G., and Adler, P. H., (eds.), *Insect Biodiversity: Science and Society.* Blackwell Publishing: Oxford. pp. 397–415.

Streiff, R., Veyrier, R., Audiot, P., Meusnier, S., and Brouat, C. 2005. Introgression in natural populations of bioindicators: a case study of *Carabus splendens* and *Carabus punctatoauratus*. *Molecular Ecology* 14: 3775–3786.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maxi- mum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.

Tan, D.S.H., Ang, Y., Lim, G.S., Ismail, M.R., and Meier, R., 2010. From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). *Zoologica Scripta* 39: 51–61.

Taylor, H.R. and Harris, W.E., 2012. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, 12: 377–88.

Thalmann, O., Hebler, J., Poinar, H.N., Paabo, S., and Vigilant, L. 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology* 13: 321-335.

Thayer, M.K. 2005. Staphylinidae Latreille, 1802. In: Beutel, R.G., and Leschen, R.A.B., (eds.), *Coleoptera, Beetles Volume 1: Morphology and Systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim).* Walter de Gruyter, Berlin. pp. 296–344.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The ClustalX window interface: flexible strategies for multiple sequence alignment aided by quality and analysis tools. *Nucleic Acids Research* 24: 4876–4882.

Timm, A. E., Stamper, T., Dahlem, G. A., Villet, M. H., and DeBry, R. W. 2014. Utility of COI, COII and ND4 genes for identifying forensically important North American

carrion flies (Insecta: Diptera). International Journal of Legal Medicine. Unpublished Work.

Tourle, R., Downie, D.A. and Villet, M.H. 2009. Flies in the ointment: a morphological and molecular comparison of *Lucilia cuprina* and *Lucilia sericata* (Diptera: Calliphoridae) in South Africa. *Medical and Veterinary Entomology* 23: 6-14.

Trewick, S.A. 2008. DNA barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* 24: 240-254.

Trontelj, P., and Fišer, C. 2009. Cryptic species diversity should not be trivialised. *Systematics and Biodiversity* 7:1–3.

Tshikae, B.P., Davis, A.L.V., and Scholtz, C.H. 2008. Trophic associations of a dung beetle assemblage (Scarabaeidae: Scarabaeinae) in a woodland savanna of Botswana. *Environmental Entomology* 37: 431–441.

Tsutsui, N.D., Suarez, A.V., Holway, D.A., and Case, T.J. 2001. Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Molecular Ecology* 10: 2151–2161.

Valentini, A., Pompanon, F., Taberlet, P. 2009. Barcoding for ecologist. *Trends in Ecology and Evolution*, 24, 110–117.

Van Der Bank, H., Herbert, D., Greenfield, R., and Yessoufou, K. 2013. Revisiting species delimitation within the genus *Oxystele* using DNA barcoding approach. *Zookeys*, 365: 337–354.

Van der Merwe, Y., and Scholtz, C.H. 2005. New species of flightless Trogidae from South African relict forest fragments. *African Entomology* 13: 181-186.

Van Laerhoven S. L., and Anderson G. S. 1999. Insect succession on buried carrion in two biogeoclimatic zones of British Columbia. *Journal of Forensic Sciences* 44: 32–43.

Van Nieukerken, E.J., Doorenweerd, C., Stokvis, F.R., and Groenenberg, D.S.J. 2012. DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-A: two are better than one in recognising cryptic species. *Contributions to Zoology* 81: 1-24.

Villet, M.H. 2011. African carrion ecosystems and their insect communities in relation to forensic entomology: a review. *Pest Technology* 5: 1-15.

Villet, M.H., and J. Amendt. 2011. Advances in entomological methods for estimating time of death. In: Turk, E.E., (ed), *Forensic Pathology Reviews*. Heidelberg: Humana Press. pp. 213- 238.

Villet, M.H., Richards, C.S., and Midgley, J.M. 2010. Contemporary precision, bias and accuracy of minimum post-mortem intervals estimated using development of carrion-feeding insects. In: Amendt, J., Campobasso, C.P., Goff, M.L., and Grassberger M., (eds), *Current concepts in Forensic Entomology*. Springer Netherlands. pp. 109-137.

Virgilio, M., Backeljau, T., Nevado, B., and De Meyer, M. 2010. Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* 11: 206.

Wahlberg, N., Oliveira, R., and Scott, J.A. 2003. Phylogenetic relationships of *Phycoides* butterfly species (Lepidoptera: Nymphalidae): Complex mtDNA variation and species delimitations. *Systematic Entomology* 28: 1–17.

Walker, G., and Leedham-Green, E. 2010. *Identity*. Cambridge University Press, Cambridge. 228 pp.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., and Hebert, P.D.N. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* 360: 1847–1857.

Weekers, P.H.H., De Jonckheere, J.F., and Dumont, H.J. 2001. Phylogenetic relationships inferred from ribosomal ITS sequences and biogeographic patterns in representatives of the genus *Calopteryx* (Insecta: Odonata) of the West Mediterranean and adjacent West European zone. *Molecular Phylogenetics and Evolution* 20: 89–99.

Wells, J. D., Wall, R. and Stevens, J. R. 2007. Phylogenetic analysis of forensically important Lucilia flies based on cytochrome oxidase I sequence: a cautionary tale for forensic species determination. *International Journal of Legal Medicine* 121: 229-233.

Wells, J.D., and Stevens, J.R. 2008. Application of DNA-based methods in forensic entomology. *Annual Review of Entomology* 53: 103–20.

Wheeler, Q.D., Raven, P.H., and Wilson, E.O. 2004. Taxonomy: impediment or expedient? *Science* 303: 285–285.

Whitfield, J. B., and Kjer, K.M. 2008. Ancient rapid radiations of insects: Challenges for phylogenetic analysis. *Annual Review of Entomology* 53: 449-472.

Whitfield, J.B. 1999. Destructive sampling and information management in molecular systematic research: an entomological perspective. In: Byers, S., and Metsger, D., (eds), *Managing the Modern Herbarium: An Interdisciplinary Approach.* Society for Preservation of Natural History Collections and Royal Ontario Museum, Ontario. pp. 271-290.

Whitworth, T. L., Dawson, R. D., Magalon, H., and Baudry, E. 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B* 274: 1731–1739.

Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M. A, Winterton, S.L., and Yeates, D.K. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biology* 7: 34.

Wiemers, M. and Fiedler, K. 2007. Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* 4: 8.

Wild, A.L., and Maddison, D.R. 2008. Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution* 48: 877-891.

Will, K.W., and Rubinoff, D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47–55.

Will, K.W., Mishler, B.D., and Wheeler Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54: 844– 851.

Williams, J. G., Hanafey, M.K., Antoni Rafalski, J., and Tingey, S. V. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods in Enzymology* 218: 704–740.

Williams, K.A., and Villet, M.H. 2013. Ancient and modern hybridization between *Lucilia sericata* and *Lucilia cuprina* (Diptera: Calliphoridae). *European Journal of Entomology* 110: 187-196.

Williams, S.T. and Knowlton, N. 2001. Mitochondrial pseudo- genes are pervasive and often insidious in the snapping shrimp *Alpheus*. *Molecular Biology and Evolution* 18: 1484–1493.

Wilson-Wilde, L. 2010. Wildlife crime: A global problem. *Forensic Science, Medicine, and Pathology* 6: 221–222.

Yassin, A., Markow, T.A., Narechania, A., O'grady, P.M., and DeSalle, R. 2010. The genus *Drosophila* as a model for testing tree and character based methods of species identification using DNA barcoding. *Molecular Phylogenetics and Evolution* 57: 509–517.

Ying, B.W., Liu, T.T., Fan, H., Wei, D., Wen, F.Q., Bai, P., Huang, J. and Hou, Y.P. 2007. The application of mitochondrial DNA cytochrome oxidase II gene for the identification of forensically important blowflies in western China. *American Journal of Forensic Medicine and Pathology* 28: 308–313.

Young, I., and Coleman, A.W. 2004. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Molecular Phylogenetics and Evolution* 30: 236–242.

Zaidi, F., Wei, S.-J., Shi, M., and Chen, X.X. 2011. Utility of multi-gene loci for forensic species diagnosis of blowflies. *Journal of Insect Sciences* 11: 59.

Zakharov, E. V., and Hellmann, J.J. 2008. Genetic differentiation across a latitudinal gradient in two co-occuring butterfly species: revealing population differences in a context of climate change. *Molecular Ecology* 17: 189–208.

Zhang, A.B., and Sota, T. 2007. Nuclear gene sequences resolve species phylogeny and mitochondrial introgression in *Leptocarabus* beetles showing trans-species polymorphisms. *Molecular Phylogenetics and Evolution* 45: 534-546.

Zhang, A.B., Muster, C., and Liang, H.B. 2012. A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular Ecology* 21: 1848–1863.

Zhang, D.X., and Hewitt, G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 12: 563–584.

Zhantiev, R.D., and Volkova, T.G. 1998. Larvae of dermestid beetles of the genus *Dermestes* (Coleoptera, Dermestidae) of Russia and adjacent countries: 1. Subgenus *Dermestes. Entomological Review* 78: 962-968