

**The use of the mitochondrial cytochrome c oxidase subunit I (COI) gene to identify and discriminate spider species of eThekweni region**

By

**Siphesihle Joel Mthethwa**

BSc (*Hons*) Genetics

Submitted in fulfilment of the academic requirements for the degree of Master of Science in

Discipline of Genetics

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa




**UNIVERSITY OF <sup>TM</sup>  
KWAZULU-NATAL**

---

**INYUVESI  
YAKWAZULU-NATALI**

**Supervisor: Dr S. Willows-Munro**

**Signature:** 

## **Preface**

The research contained in this dissertation was completed by the candidate while based in the discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by the EM-UKZN Joint Research Partnership: KwaZulu-Natal Sandstone Sourveld Research Programme.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.



---

Signed: Dr S. Willows-Munro

Date: 13 January 2017

## College of Science and Agriculture

### Declaration of Plagiarism

I, Siphesihle Joel Mthethwa, declare that:

- (i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;
- (ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;
- (iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;
- (iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - a) their words have been re-written but the general information attributed to them has been referenced;
  - b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;
- (v) where I have used material for which publications followed, I have indicated in detail my role in the work;
- (vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;
- (vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

---

Signed: Siphesihle Joel Mthethwa

Date: 13 January 2017

## **Abstract**

**DNA barcoding compliments traditional morphology-based taxonomic approaches and is a molecular-based method for rapid species identification and flagging of potential new species. A fragment of 658-bp of the mitochondrial cytochrome c oxidase subunit I (COI) is used as the standard DNA barcode region for animals. While DNA barcoding has been applied with success across a wide range of eukaryotes, including spiders, in some animal groups, for example Diptera, COI has failed to reliably separate recognized species. This suggests that for particular groups DNA barcoding may not provide accurate species identification. Lack of gap between interspecific and intraspecific genetic distances complicates accurate species delimitation using DNA barcoding alone. This is particularly problematic when DNA barcoding is used differentiate among species which are closely related. The main aim of this study is to test the utility of COI to accurately discriminate among species of South African spiders collected from a small regional area and so we expect many of the species to be closely related. The study took place within the eThekweni municipality and surrounding areas. The municipality includes the city of Durban, and is situated within the globally-important Maputaland-Pondoland-Albany biodiversity hotspot. This region is characterized by a large human population (3.55 million) and a high rate of urbanisation (92%). In this study spiders (order Araneae) were chosen as indicators of diversity. Spiders are a hyper-diverse group of arthropods, with 40 000 species described world-wide. There are currently, 2 170 described species from 71 families recorded from South Africa, but this number is expected to rise as additional species continue to be described. Spiders are important as bioindicators, and have been used to study the effect of urbanisation on biodiversity. In this study DNA barcoding is used to aid in the rapid identification of spider species. By using the DNA-based tool, the effect of urbanisation on spider species diversity along an urbanisation gradient will be assessed. The study also intends to design and test the utility of mini-barcodes for species identification when the full DNA barcode region cannot be amplified i.e. when DNA is degraded.**

## **Acknowledgements**

My earnest thanks to my supervisor, Dr Sandi Willows-Munro for her guidance throughout this project. I would like to acknowledge the EM-UKZN Joint Research Partnership: KwaZulu-Natal Sandstone Sourveld Research Programme for funding this project. I would also like to thank the University of Guelph in Canada for assistance with this project. The Conservation Genetics lab members are also thanked for sharing ideas and previous experience, their input has made a valuable contribution towards shaping this project and for making the lab a fun place to be. The SANBI interns (2011-13) are thanked for helping with specimen collection, my family for their relentless support, emotionally and financially, and last but not least, Ms NR Maduna for being there for me through this entire journey since my undergraduate, she believed in me and always encouraged me to reach for the stars. You are appreciated!

## Contents

Preface.....	ii
Declaration of Plagiarism .....	iii
Abstract.....	iv
Acknowledgements.....	v
List of Figures .....	ix
List of Tables .....	xi
1. Urbanisation and biodiversity.....	2
1.1 Arthropods in conservation .....	3
1.2 Spiders.....	3
2 DNA barcoding.....	4
2.1 How DNA barcoding works?.....	4
2.2 Is COI a good marker for DNA barcoding?.....	6
2.3 DNA barcoding of Spiders.....	8
2.4 Current Advances in DNA barcoding .....	8
3 Aims of this thesis .....	9
Chapter Two: Construction of a Preliminary Reference Library for Spiders of eThekweni ...	11
Abstract.....	11
2.1 Introduction.....	12
2.1.1 DNA barcoding as an identification tool.....	13
2.1.2 DNA Barcoding Gap as a measure of identification accuracy .....	13
2.1.3 DNA barcoding as tool to monitor biodiversity .....	14
2.1.4 Database.....	14
2.1.5 Spiders .....	15
2.2 Materials and Methods.....	16
2.2.1 Study area.....	16
2.2.2 Sampling of specimens .....	16

2.2.3	Morphological sorting of specimens .....	17
2.2.4	Genomic DNA extraction and Polymerase Chain Reaction amplification .....	18
2.2.5	Sequence data processing .....	19
2.2.6	Phylogenetic analysis of the spider DNA barcode reference library.....	19
2.2.7	Threshold optimization and measures of identification accuracy .....	20
2.2.7.1	Measures of identification accuracy .....	20
2.2.7.2	The Barcoding gap .....	21
2.3	Results.....	22
2.3.1	Threshold optimization and measures of identification accuracy .....	24
2.3.3	Phylogenetic analysis .....	28
2.4	Discussion.....	31
2.5	Conclusion .....	32
Chapter Three: Assessing the Spatial Distribution of Spider Species in eThekweni using DNA Barcoding.....		33
Abstract.....		33
3.1	Introduction.....	34
3.1.1	Spiders as bioindicators.....	34
3.1.2	DNA barcode data to evaluate levels of biodiversity.....	35
3.1.3	Vegetation type and biodiversity as a function of urbanisation .....	36
3.2	Aims.....	36
3.3	Materials and Methods.....	37
3.3.1.1	Species accumulation curves .....	37
3.3.2	Spatial distribution of spider diversity .....	39
3.4	Results.....	40
3.4.1	Species richness and diversity .....	44
3.4.2	Spatial distribution and species richness .....	47
3.4.3	The effect of area size and distance from the city on species richness .....	49

3.5	Discussion.....	50
	Abstract.....	54
4.1	Introduction.....	55
4.2	Materials and Methods.....	56
4.2.1	Designing of mini-barcodes, manually.....	56
4.2.1.1	Phylogenetic analysis of mini-barcodes generated manually.....	57
4.2.2	Sliding Windows Analysis on R.....	57
4.3	Results.....	58
4.3.1	Manually designed mini-barcodes.....	58
4.3.2	Sliding Windows Analysis.....	59
4.4	Discussion.....	62
4.4.1	Manually created mini-barcodes.....	62
4.4.2	Sliding Windows Analyses.....	63
4.5	Conclusion.....	63
	Chapter Five: General Discussion.....	65
6.	References.....	68
	Appendix1. A neighbor-joining tree of the 1 153 specimens of spiders collected from the eThekwini region and surrounding areas.....	67
	Appendix2. A neighbor-joining tree of the optimal 219-bp mini-barcode as predicted by the sliding windows analysis.....	67



## List of Figures

- Figure 2.1** Partial map of the KwaZulu-Natal province showing all the study sites within the eThekweni Municipal area and in outlying areas. Species richness was studied across the transect starting from areas adjacent to Durban city up to inner parts of KwaZulu Natal (UKZN & Bisley Nature reserve).....17
- Figure 2.2** Examples of photographs taken of specimens using a Digital Image Microscope, A: HamGr6Arach12 and B: HamGr6Arach1401, taken after sorting into morphospecies, prior to DNA extraction. The body size was measured using standard graph paper. ....18
- Figure 2.3** Barplot showing the number of species already represented in BOLD (greater than 95% similarity) and new species added to the database (less than 95% similarity).....22
- Figure 2.4** Barplot of false positives (light grey) and false negative (dark grey) rate of identification for the Araneae of eThekweni data set using a range of threshold values from 1.0 to 5.0%. As the percentage threshold for species delimitation increases, the number of false negatives increases. ....25
- Figure 2.5** Barplot of the success rate using ThreshID identification criterion and the Best Close Match at 1-5% thresholds. The optimal threshold for identification for both criteria was 3%.....27
- Figure 2.6** Distribution of pairwise divergence values calculated using the K2P model. There is no overlap between the intra- and interspecific distances. The gap occurred between the 0.020 - 0.035 range of K2P distance.....28
- Figure 2.7** Bayesian tree of eThekweni spiders including 48 species from 45 genera. The tree is midpoint rooted and only posterior probabilities above 0.5 are presented in the tree. Only major family groups, represented by at least three genera, are highlighted. ....29
- Figure 2.8** Maximum likelihood tree including 48 species of eThekweni spiders from 45 genera. The tree is midpoint rooted and bootstrap values greater than 70% are shown. Only major family groups, represented at least by three genera, are highlighted. ....30
- Figure 3.1** Haplotype accumulation curve for 1 153 DNA barcodes generated on SPIDER, using 1000 “random permutations” and the number of sequences set to 2000. The poor gradient indicates the need for additional sampling. ....41
- Figure 3.2** Haplotype accumulation curves for Iphithi, Palmiet, Springside nature reserves and Drummond grassland. The numbers on the y-axis represent the number of haplotypes observed and on the x-axis is the number of sequences barcoded from each

locality. The species accumulation curve is the solid black line, which represents the average of 100 random draws, sampling with replacement, at each level of abundance. The shaded envelope represents a symmetric 95% bootstrap confidence interval, calculated from the estimated variance of the random draws. ....42

**Figure 3.3** Haplotype accumulation curves for Hamiton grassland, Vernon Crookes, Msini and New Germany Nature Reserves. The numbers on the y-axis represent the number of haplotypes observed and on the x-axis is the number of sequences barcoded from each locality. The species accumulation curve is the solid black line, which represents the average of 100 random draws, sampling with replacement, at each level of abundance. The shaded envelope represents a symmetric 95% bootstrap confidence interval, calculated from the estimated variance of the random draws. ....43

**Figure 3.4** Species accumulation curves for 1 153 specimens using three extrapolation methods, where S (obs) is the number of observed species while Chao 2, Jack 1 and MM are estimates of true species richness. ....44

**Figure 3.5** Graph showing the number of specimens, BIN clusters and morphospecies per sampling site. The sampling sites are sorted in an increasing geographic distance from the harbour; Msinsi nature reserve is closest to the urban hub and Bisley nature reserve the farthest. ....45

**Figure 3.6** Dendograms based on the Bray-Curtis similarity, showing the similarity in spider assemblages between study sites based on (a) abundance data (to compare the distribution of spider species throughout the entire sampling region) and (b) presence/absence data (distribution of species as affected by vegetation type). Sites sharing more species are drawn together.....48

**Figure 3.7** Bar graph showing the distribution of genera across the eThekwini study sites. Only three genera, *Leucuage*, *Hypsosinga* and *Oxyopes*, were found in at least 50% of the sampling sites whereas the rest of the genera occurred only in a few sites.....49

**Figure 4.1** Sliding Windows Analysis carried on R environment using SPIDER package; window length = 218-bp; interval = ‘codons’. The top graph, first column shows the mean distances of the sequences. The third graph, column one shows the proportion of summary diagnostic nucleotides, characters informative for species identification. The second graph on the second column shows the proportion of congruent trees between mini-barcodes and the full length dataset. ....60

## List of Tables

<b>Table 3.1</b> Detail of sampling conducted at 16 sites through a period of 2011-2015. The coordinates of the exact locations are presented, the size of each open space, vegetation classification, desistance from city hub as well as the number of times each open space was sampled.....	38
<b>Table 3.2</b> Biodiversity estimation indices where $S$ is the total number of COI BINs, $N$ is the total number of individual, $D$ is Simpson's index, $H'$ Shannon's index, $d$ Margalef's index and $\alpha$ Fisher's $\alpha$ .....	46
<b>Table 4.1</b> Summary statistics for all the mini-barcode and full-length barcode alignments computed on MEGA5. The names of the mini-barcode alignments indicate their size (length in bp) and their position relative to the 5' end of the mitochondrial COI gene. ..	58
<b>Table 4.2</b> Comparison of maximum likelihood trees generated from mini-barcodes against the full length barcode tree. ....	59
<b>Table 4.3</b> The results of the Sliding Windows Analyses summarized, showing potential mini-barcode segments and their positions in the full-length alignment. ....	61

## **Chapter One: An Introduction to DNA Barcoding and Invertebrates in Urban Areas**

### **1. Urbanisation and biodiversity**

Populations of native plants, animals and other organisms are rapidly declining around the world, leading to a threat to global biodiversity known as the “biodiversity crisis”. Along with the spread of invasive species, habitat loss is considered among the greatest threats to biodiversity (Moorhead and Philpott, 2013). Urbanisation, the process by which urban ecosystems are created (Jones and Leather, 2012), is the leading cause of habitat loss and habitat fragmentation (Jones and Leather, 2012; Magura et al., 2008; McKinney, 2002, 2008; Moorhead and Philpott, 2013; Shochat et al., 2010). According to the 2014 Global Health Observatory data, human populations living in urban settings accounted for 54% of the total global population, this value has increased from 48% in 2004 (WHO, 2014). The global urban population is expected to grow approximately 1.84% per year between 2015 and 2020 and 1.63% per year between 2020 and 2025. With this projected population growth it is anticipated that an even greater proportion of natural habitat will become fragmented or lost completely through urbanisation (WHO, 2014). As a result massive extinctions of species at both local and global scales have been witnessed (McKinney, 2008; Shochat et al., 2010).

With the on-going urbanisation, green spaces within urban areas are becoming more important as wildlife conservation areas and movement corridors (Jones and Leather, 2012). Not all native biota can thrive in these green spaces within cities. Fragmentation of suitable habitat and land use transformation will favour certain generalist or fast adapting taxa, and over time species composition will be changed (Jones and Leather, 2012; Lowe et al., 2014; Shochat et al., 2010). As urbanisation is set to continue, understanding the impact it has on native wildlife in urban settings has become increasingly important if we are to conserve biodiversity. The most commonly used taxa for measuring biodiversity are multicellular organisms (Purvis and Hector, 2000), particularly invertebrates, such as insects and other arthropods (Wilson, 1988). Invertebrates comprise an important component of biodiversity in many ecosystems and as such are key taxa to include in comprehensive biodiversity surveys (Taylor and Doran, 2001). The impact of urbanisation on invertebrates has been studied previously in Coleoptera (Niemela et al., 2002), terrestrial isopods (Magura et al., 2008), and Araneae (Magura et al., 2010; Shochat et al., 2004). Some studies reported no significant difference in overall species richness along an urbanisation gradient (Jones and Leather,

2012; Magura et al., 2008). In contrast, other studies show urbanisation is the primary cause for decline in arthropod populations (Horvath et al., 2009; Jones and Leather, 2012; Magura et al., 2010; Magura et al., 2008).

### **1.1 Arthropods in conservation**

Arthropods (phylum arthropoda) are invertebrates and include insects, arachnids, myriapods and crustaceans. Given their great diversity, which includes rare and economically important species (Jones and Leather, 2012), coupled with their ability to occupy the widest possible diversity of ecosystems, and their ecological importance, invertebrates provide a good model of environmental change and are useful bioindicators (Jones and Leather, 2012; Kremen et al., 1993). In addition, many arthropod species respond to environmental changes more rapidly than vertebrates, which makes them beneficial for management purposes (Kremen et al., 1993).

Despite the increased awareness of their importance in global conservation planning (Ferrier et al., 2004; Kremen et al., 1993), relatively little attention has been devoted to the inventory and monitoring of terrestrial arthropods (Smith et al., 2005). Studies have, however, been conducted on some insect taxa for example Diptera (Whitworth et al., 2007), Lepidoptera (Wiemers and Fiedler, 2007), Odonata (Rach et al., 2008) and Plecoptera (DeWalt et al., 2012). Invertebrates make up a large portion of terrestrial biodiversity (Andersen et al., 2004b) and studies have been conducted on ants (Andersen et al., 2004a; Smith et al., 2005), ground dwelling beetles (Magura et al., 2008; McKinney, 2002) and spiders (Horvath et al., 2009; Magura et al., 2010).

### **1.2 Spiders**

Spiders (order Araneae) are a hyper-diverse group of arthropods, with approximately 40 000 species described worldwide with 2 170 species in 71 families recorded from South Africa (Dippenaar-Schoeman et al., 2015). Of these species 1 220 (>60%) are endemic to South Africa (Dippenaar-Schoeman et al., 2015; Foord et al., 2011). This is thought to represent only a fraction of the actual diversity present, with many species yet to be formally described (Barrett and Hebert, 2005a). Morphological characters used to define spider species are often

difficult for non-experts to interpret and most diagnostic characters only appear at certain stages of life (Barrett and Hebert, 2005a). Also given the complex sexual system that occurs in spiders, many species tend to show pronounced sexual dimorphism, which can further complicate species identification (Coddington and Levi, 1991). Thus spider taxonomy remains challenging. African species are, in particular, poorly studied with most attention paid to trapdoor and baboon spiders rather than the entire order of spiders (McGeoch et al., 2011). If the current knowledge on South African spiders is to be used to benefit conservation and planning initiatives particularly in urban environments then extra identification tools need to be developed to speed-up and improve species description and characterization. In this study a molecular tool ‘DNA barcoding’ for rapid and accurate species identification was tested. This molecular tool has been successfully used in a variety of animal groups (Virgilio et al., 2010).

## **2 DNA barcoding**

DNA barcoding is a molecular tool for rapid species identification (distinguish between known species) and discovery (delimitation of new lineages) using standardized DNA regions or ‘DNA barcodes’ as tags (DeSalle et al., 2005; Hebert et al., 2003a; Tavares and Baker, 2008; Valentini et al., 2008). The goal of DNA barcoding involves the development of molecular-based species identification system which allows unknown individuals to be assigned to species whilst enhancing the discovery of new species (Austerlitz et al., 2009; Tavares and Baker, 2008).

### **2.1 How DNA barcoding works?**

DNA barcoding makes use of short sequence tags, from standardized common genomic regions, as diagnostic barcodes that can be used to identify and document species. Because this is a DNA-based approach, it has a number of advantages over the traditional taxonomic methods. Factors like phenotypic plasticity, cryptic speciation, and sexual dimorphism (particularly in spiders) can complicate identification based on morphology alone. DNA barcoding could compliment classic morphology-based taxonomy (Krishnamurthy and Francis, 2012), by providing systematists with an additional tool to quickly and reliably sort specimens into species.

The mitochondrial cytochrome c oxidase I (COI) gene has been used as the standard molecular marker in most animal groups, including bovids (Ibrahim et al., 2012), birds (Hebert et al., 2004b; Hebert et al., 2004b; Tavares and Baker, 2008), ants (Smith et al., 2005), bees (Koch, 2010; Schmidt et al., 2015), and butterflies (Burns et al., 2007; Hajibabaei et al., 2006; Wiemers and Fiedler, 2007) among others. DNA barcoding works on the premise that the genetic distance among individuals that belong to the same species is markedly lower than the genetic distance between different species thus enabling successful identification of species using clustering algorithms based on genetic distance (Elias et al., 2007). While DNA barcoding has been successfully applied in most orders of insects, the utility of COI to accurately delimit species can vary significantly across different taxonomic groups. For example, low identification success has been reported in studies using the barcode marker COI in Diptera (Meier et al., 2006; Whitworth et al., 2007; Yassin et al., 2010), Lepidoptera (Elias et al., 2007; Wiemers and Fiedler, 2007) and Orthoptera (Trewick, 2007).

DNA barcoding has grown through the work of an international initiative, the Consortium for the Barcode of Life (CBOL), whose main objective has been to promote global standards and co-ordinate research in DNA barcoding (Valentini et al., 2008). To achieve these two goals, CBOL built the Barcode of Life Data systems (BOLD) which is used to store and curate all generated DNA barcodes. These barcode records are compiled into a publically available reference library, which also contains DNA barcode sequences linked to reference samples of known species, photographs of individual specimens barcoded and the GPS coordinates of each specimen collected and barcoded (Austerlitz et al., 2009; Ratnasingham and Hebert, 2007). The addition of spatial information allows for local assessment of biodiversity by ecologists and could facilitate the monitoring of biodiversity (BOLDSYSTEMS, 2013). The database BOLD also provides bioinformatics tools (such as blast search) that can be used for species identification using molecular data.

## 2.2 Is COI a good marker for DNA barcoding?

DNA barcoding involves the large-scale screening of standardized reference gene or genes across a wide taxonomic range of organisms (Hebert et al., 2003; Hollingsworth et al., 2011; Moritz and Cicero, 2004; Stoeckle, 2003). The success of this technology relies on selecting standardized portions of the genome, which are conservative enough that they can be amplified from a wide taxonomic range of organisms, but these markers also need to have a mutation rate appropriate for delimitation of organisms to species level (Valentini et al., 2008).

Although a range of molecular markers (e.g., allozymes, rDNA and mtDNA) have been used to clarify phylogenetic relationships among species. In animals, the mitochondrial COI gene has been found to be superior to any other mitochondrial marker as it matches the two fundamental prerequisites of a successful DNA barcoding marker. The amino acid sequence of the mitochondrial COI gene is conserved allowing for broad applicability of primers thus enabling analysis for most animal groups (Hebert et al., 2003; Moritz and Cicero, 2004), but at the nucleotide level, the mutation rate is appropriate for the discrimination of species including those that are closely related (Hebert et al., 2003; Hebert et al., 2003a; Hebert et al., 2003b). The mitochondrial COI gene, however, is only useful in animals. For groups such as plants and fungi other portions of the genome are more useful. Until recently COI has been considered by the Consortium for the Barcode of Life as the default marker for fungi. Studies, however, show COI is more reliable in a few clades of closely related species and results in the few groups examined are experimentally inconsistent and cloning was often required (Dentinger et al., 2011). In some fungal clades such as *Neocallimastigomycota*, the mitochondrion is absent (Robideau et al., 2011; Schoch et al., 2012). Thus in place of COI the internal transcribed spacer (ITS) region is used, it has the highest probability of successful identification for the broadest range of fungi, with most clearly defined gap between inter- and intraspecific variation (Schoch et al., 2012). Supplementary barcodes to complement ITS region may be developed in the future for narrowly described taxonomic groups.



The rate of nucleotide substitution in plant mitochondrial genomes is generally low preventing the use of COI as a universal plant barcode (Hollingsworth et al., 2011). Instead many researchers propose various combinations of seven plastid markers to obtain adequate species discrimination. These are *rbcL*, *trnH-psbA*, *rpoB*, *rpoC1*, *matK*, *atpF-H* and *psbK-I* (Hollingsworth et al., 2011). *In silico* assessment of the resolving power of different marker combinations showed several combinations performed equally well and the *rbcL+matK* combination was highlighted as having the best potential for use as a core-barcode. The two-marker plastid barcodes gave better discrimination than single barcodes and no other 2-marker or multi-marker plastid barcode gave greater species resolution than the *rbcL+matK* combination (CBOL-PWG, 2009). This result is however different from the findings by Costion et al. (2011) who reports all combinations which included the *trnH-psbA* locus performed better at species discrimination than *matK* and *rbcLa* combined.

Despite its success in many animal groups, DNA barcoding using a single marker system has been pointed out by some critics as insufficient for species identification (Krishnamurthy and Francis, 2012; Moritz and Cicero, 2004). Factors such as introgression due to hybridisation, heteroplasmy, infection with *Wolbachia* bacteria in arthropods (Smith et al., 2012) and the occurrence of nuclear pseudogenes of mitochondrial origin have been identified as major limits to mtDNA use (Jinbo et al., 2011; Krishnamurthy and Francis, 2012; Waugh, 2007). These limits are not specific to barcoding but are a major consideration for any study that relies on a single marker system. As such DNA barcoding cannot be used to replace traditional morphology-based taxonomy (Krishnamurthy and Francis, 2012), rather it should be viewed as a useful tool to complement traditional taxonomy (Hajibabaei et al., 2007). Although the taxonomic limits of COI barcoding have not been fully understood in all animal groups (Gunnarsson, 1990) the diagnostic system has proven useful in determining species in most groups tested with many studies reporting relatively high species identification rates of up to and exceeding 95% (Waugh, 2007). For example, Hebert et al. (2003) claims a 100% success rate in identifying Lepidopterans, while Hubert et al. (2008) report a 93% success rate in identifying Canadian freshwater fish. The identification success in other less well-studied animal groups, however, remains to be quantified.

### **2.3 DNA barcoding of Spiders**

Due to their extensive morphological variation, spiders are taxonomically challenging (Gaikwad et al., 2016), a number of studies have tested the success of DNA in species identification i.e. Barrett and Hebert (2005a) and Robinson et al. (2009). These studies suggest DNA barcoding can be used to successfully identify species, however, to a certain extent. This was indicated by the presence/absence of barcoding gaps in certain families. Because spiders are hyperdiverse, most studies only focussed on certain groups while others used larger geographic scales (Blagoev et al., 2013) and others mostly on local fauna (Barrett and Hebert, 2005a). Given the success rate of using DNA barcoding to identify spiders, attempts to barcode spiders are being made all over the world, however, it is important to note, till now no attempts were made to DNA barcode South African spiders despite their rich diversity.

### **2.4 Current Advances in DNA barcoding**

Given the utility of DNA barcodes to delimit species and aid in the description of undescribed species, many institutions and countries around the world have embarked on constructing DNA barcode reference libraries to catalogue the world's biodiversity (Hajibabaei et al., 2006b). The majority of current reference libraries focus mainly on the analysis of fresh or recently collected specimens, as these specimens contain the richest source of DNA. However, as the reference library grows, it is becoming increasingly difficult to find fresh specimens of species that require barcoding to complete the reference library (Miller et al., 2013). Museums are an indispensable source of biodiversity information relevant to ecology, evolutionary biology, and conservation biology (Hajibabaei et al., 2006b). They harbor millions of specimens that have been identified by taxonomic experts, and tapping into these sources could provide a cost-effective way of building barcode libraries (Hajibabaei et al., 2006b), particularly for rare species or species that are difficult to collect.

Until recently, DNA barcoding has not been fully applied to museum collections as obtaining the full length barcode of 658-bp is often difficult (Miller et al., 2013). The most popular methods of insect preservation used in museums include pinning and drying of specimens; in very old specimens the amount of DNA recovered is low and there is decreased PCR success

rate over time (Jinbo et al., 2011; Miller et al., 2013). As specimens age DNA becomes fragmented and longer fragments show relatively low amplification success. Amplification is still possible using primers that amplify shorter 200-bp fragments (Hajibabaei et al., 2006b). To address the low amplification success of archival museum material, two strategies have been proposed (i) to identify specimens based only on short fragments ‘mini-barcodes’ that are easily amplified, (ii) obtaining a full barcode length by concatenating all the short fragments (Jinbo et al., 2011). The latter is, however, time consuming and costly, which violates the claims of DNA barcoding as a method for rapid, cheap method of identification. Mini-barcodes have been developed for fish (Meusnier et al., 2008), butterflies (Hajibabaei et al., 2006b); fruit-flies (Fan et al., 2009); and snakes (Dubey et al., 2011). These studies show that short portions of the barcode marker are effective for species identification (Dubey et al., 2011; Fan et al., 2009; Hajibabaei et al., 2006b). The use of mini-barcodes further extends the application of DNA barcoding, allowing for rapid assessment of biodiversity from collections of environmental samples which may also contain degraded DNA, such as animal scats (Kress et al., 2015). This relatively new method is known as meta-barcoding (Deagle et al., 2014; Yu et al., 2012).

### **3 Aims of this thesis**

Given the importance of invertebrates in biodiversity and conservation planning, particularly in urban areas, the current study sets out to study the distribution of spider species and spider community assemblages across the eThekweni region through the use of DNA barcoding. A preliminary reference library representing spider diversity of the eThekweni region will be constructed and tested for reliability and accuracy of species identification. At present, only 1,507 South African specimens are present on BOLD database (BOLDSYSTEMS, 2015) and this study aims to substantially improve the number of records, as well as the taxonomic and geographic coverage currently available.

The aims of the study involve:

- (i) testing the utility of the COI DNA barcode marker in South African spiders, particularly for species in the eThekweni region, this involves construction and evaluation of a barcode reference library for the spiders of eThekweni,

- (ii) use the generated barcode data to assess the diversity and community assemblages of spiders collected from the open spaces of the eThekweni region, and also investigate the effect area size and distance from city center (urbanisation gradient) have on spiders diversity,
- (iii) using the COI reference library, this study will design and test (*in silico*) mini-barcodes for use in future South African spider species identification.

## **Chapter Two: Construction of a Preliminary Reference Library for Spiders of eThekwini**

### **Abstract**

DNA barcoding has been proposed as a method for quick and reliable species delineation. Accurate species identification of unknown species can be achieved by blasting the COI gene of the unknown specimen against the publically available data stored in the Barcode of Life Data System (BOLD). Successful species assignment, however, rests on the inclusiveness of identified taxa represented in the database. In this chapter I present a preliminary reference library for spiders of eThekwini region. In total 1 153 barcode records are analysed. The reference library constructed includes at least 90 genera and 30 families. I evaluate the usefulness of DNA barcoding in identification of spider species collected from the eThekwini region and surrounding areas. While most studies only present tree-based (neighbor-joining) methods for assessing identification success, in this study I test both tree-based (maximum-likelihood and Bayesian-Inference) methods and distance-based methods (Best Close Match and ThreshID). The findings of the current study confirm the usefulness of DNA barcoding in identifying spiders of eThekwini (100% identification success). The COI marker was sufficiently variable between different species, maximum intraspecific = 0.020; minimum interspecific = 0.035, and as a result a barcoding gap was present. A threshold of 3% genetic distance is reported as the optimal threshold for accurate species identification. Additional sampling and taxonomic expertise are required to identify voucher specimens to increase the number of species represented in the database. Including a large number of individuals per species will also improve the reliability of the results.

## 2.1 Introduction

The accurate identification and characterization of living things is fundamental to biological sciences and understanding how many species occur in a given area is vital in the development of effective conservation strategies and biodiversity surveillance (Waugh, 2007; Witt et al., 2006). Despite 250 years of hard work dedicated to identification and classification of organisms, the number of described species falls far short of actual diversity present on earth and a large number of species remain to be identified (Krishnamurthy and Francis, 2012; Waugh, 2007; Witt et al., 2006). In the past most species identification has relied solely on traditional taxonomy, which is morphology-based. Although this method is well entrenched in the literature, it is difficult and time-consuming particularly in less well-studied groups such as invertebrates. As such this group contains the largest number of undescribed species (Blaxter, 2003).

In invertebrates, identification of species using morphological characters alone is often challenging, due to several limitations, including phenotypic plasticity of the trait being examined, the existence of cryptic species that can only be separated by subtle morphological differences, different developmental stages (e.g., different life forms or development of certain diagnostic traits), and sexual dimorphism (Valentini et al., 2008; Witt et al., 2006). DNA barcoding is a molecular-based approach, which has been used to identify species where morphology is insufficient or complicated (i.e. morphological keys are only applicable at particular life stage or gender (Barrett and Hebert, 2005a; Hajibabaei et al., 2011; Kress et al., 2015; Moritz and Cicero, 2004)).

Given the two major problems facing conservation, (i) the loss of biodiversity, which is occurring at an increasingly rapid rate in comparison to the discovery of new species, and (ii) the greatly diminishing population of taxonomists (Krishnamurthy and Francis, 2012), a fast and accurate species delimitation method could help accelerate species discovery. DNA barcoding has the potential to provide biologists with an inexpensive, simple tool to aid both identification of known species and the recognition of undescribed taxa (Witt et al., 2006).

### **2.1.1 DNA barcoding as an identification tool**

DNA barcoding involves the sequencing of mitochondrial COI gene (for animals) to aid in species identification and discovery (Hebert et al., 2003; Ratnasingham and Hebert, 2007). The approach has been applied to various taxa with high success (90% and above in some groups), even in hyper-diverse groups (Waugh, 2007), including fish (Ward et al., 2005), birds (Hebert et al., 2004b; Hebert et al., 2004b), bees (Schmidt et al., 2015), ants (Smith et al., 2005), and butterflies (Hebert et al., 2004a). Most importantly for this study, spider species have also been successfully delimited using the COI gene (Barrett and Hebert, 2005b; Candek and Kuntner, 2015; Paquin and Hedin, 2004; Robinson et al., 2009).

In some cases DNA barcoding has been found to be less successful in species identification, for instance in Diptera (Meier et al., 2006; Whitworth et al., 2007) and Lepidoptera (Wiemers and Fiedler, 2007). Several factors have been suggested to affect the success of identification using DNA barcoding. These factors include the phylogeographic scale of sampling (Bergsten et al., 2012), also the insufficient intraspecific sampling in favour of greater taxonomic coverage (Chapple and Ritchie, 2013; Zhang et al., 2012). These two factors interfere with species identification, intraspecific distances may vary greatly as a result of the scale of sampling, Bergsten et al. (2012) report that in some cases genetic divergence of individuals of same species at different geographic locations can be higher than expected because of below species-level phylogeographic structuring. This compromises the ability of DNA barcoding to accurately identify species, as each geographic cluster would be assigned to a separate species group overestimating species richness.

### **2.1.2 DNA Barcoding Gap as a measure of identification accuracy**

Identifications using DNA barcodes are made based on genetic divergence of sequences. The assumption being that interspecific genetic divergence is always higher than intraspecific divergence (Barrett and Hebert, 2005a; Robinson et al., 2009). This distinction between intra- and inter-specific divergences is known as the 'barcoding gap'. The presence of a barcoding gap is essential for species delimitation as it enables unknown sequences to be assigned with confidence to an existing species cluster or flagged as new species (Austerlitz et al., 2009; Chapple and Ritchie, 2013).

The absence of the barcoding gap (overlap between intra- and inter-specific) is often attributed to issues with the existing taxonomy (presence of cryptic species or unrecognized species complexes), the quality of the reference database (sparse sampling), as well as the geographic scale of sampling and the intensity of intraspecific sampling (Austerlitz et al., 2009; Bergsten et al., 2012; Chapple and Ritchie, 2013). Establishing the presence of the barcoding gap is essential when DNA barcoding is done on a small regional scale such as in the current study, where there is a high likelihood that species included will be closely related. However, intensive intraspecific sampling is required, this way proper sequence divergence threshold values and mean genetic distances for species delimitation can be clearly established.

### **2.1.3 DNA barcoding as tool to monitor biodiversity**

A decline in biodiversity has a negative impact on ecosystem services and other biodiversity related services, and so monitoring biodiversity should be both a conservation and urban planning priority (Gordon et al., 2009; Krishnamurthy and Francis, 2012). However, attempting to catalogue and monitor biodiversity even at a small regional scale is a very difficult task. As a result, knowledge of both global and local levels of biodiversity remains insufficiently understood (Krishnamurthy and Francis, 2012). In efforts to rapidly describe and monitor biodiversity, biologists and ecologists have turned to DNA barcoding (Blaxter et al., 2005). DNA barcoding seeks to provide an invaluable tool for (i) rapid global and regional biodiversity assessment, by facilitating identification of species already defined by taxonomic expertise, (ii) and to facilitate the description of new species. Most importantly, DNA barcoding involves the creation of an electronic reference library that can be used by policy makers (such as local government) to guide urban development and monitor biodiversity at a small regional scale.

### **2.1.4 Database**

The success of DNA barcoding rests on the availability of a reliable, searchable DNA barcode reference library containing sequences of all described species (Krishnamurthy and Francis, 2012; Sonet et al., 2013). The DNA barcode reference library contains reference sequences linked to photographs of voucher specimens, sequencing trace files, and additional information such as primers used and collection data (GPS coordinates) and the taxonomist



who identified the specimens (Milton et al., 2013). Having this data stored in a single database allows for quality control i.e. validation of taxonomy through morphology assessment (Ratnasingham and Hebert, 2007; Sonet et al., 2013).

DNA barcode record provides useful information for different stakeholders in conservation: researchers could identify species more quickly, taxonomists could determine groups of species that require more detailed studies, and policy makers could use barcode data to determine appropriate scales for conservation (Krishnamurthy and Francis, 2012). Including DNA data in biodiversity inventories allows for rapid biodiversity assessment at both global and local levels, and could streamline the identification of areas that are in great need of conservation actions or protection. Also DNA barcoding allows for continuous monitoring of biodiversity, this provides an opportunity to evaluate the success of conservation strategies or actions (Krishnamurthy and Francis, 2012).

### **2.1.5 Spiders**

Cataloguing and monitoring biodiversity for conservation purposes is a difficult task thus ecologists have relied on bioindicators that can be used to provide indications of ecosystem health (Andersen et al., 2004a; McGeoch et al., 2011; Paoletti, 1999). Invertebrates constitute a substantial proportion of terrestrial and freshwater biodiversity and provide critical ecosystem functions and are thus often used as bioindicators (Ferrier et al., 2004; Kremen et al., 1993; McGeoch et al., 2011). Spiders (Araneae) have been used as bioindicators (Otter et al., 2012; Pearce and Venier, 2006) as they have the ability to occupy a wide range of terrestrial habitats, they are very diverse and provide a good model of environmental change which is important in conservation planning (Candek and Kuntner, 2015; Kremen et al., 1993). However, because of the very distinct behaviour and morphological characters many species possess, it becomes difficult applying a single reliable identification method to all spiders species (Candek and Kuntner, 2015). In this chapter, the utility of DNA barcoding in identifying spider species of the eThekwini region is tested for by means of testing for the presence of a barcode gap, also finding the optimal threshold for species identification. The barcode reference library for the spiders of eThekwini was constructed and its reliability evaluated,

## **2.2 Materials and Methods**

### **2.2.1 Study area**

This project is part of the KwaZulu-Natal Sandstone Sourveld Research Programme, an interdisciplinary, joint research partnership between the eThekweni Municipality and the University of KwaZulu-Natal. The study took place primarily within the eThekweni Municipal area, although other surrounding areas outside the eThekweni region (Bisley nature reserve, UKZN grassland in Pietermaritzburg) were also included. The eThekweni region is 2 297 km<sup>2</sup> in size and represents approximately 1.4% of the KwaZulu-Natal province. The municipality includes the city of Durban, which is characterized by large human population 3.55 million, with 0.9% predicted population growth rate (Govender, 2014) and high rate of urbanisation 92% (Govender, 2014). The region is rich in biodiversity and contains three of the country's eight terrestrial biomes namely; savanna, forest and grassland (Govender, 2014) and eight of the recognized vegetation types (Eastern Valley Bushveld, KwaZulu-Natal Coastal Belt, KwaZulu-Natal Hinterland Thornveld, KwaZulu-Natal Sandstone Sourveld, Ngongoni Veld, Scarp Forest, Northern Coastal Forest and, Mangroves, (Govender, 2014). The municipality is situated within the globally-important Maputaland-Pondoland-Albany biodiversity Hotspot (Govender, 2014).

### **2.2.2 Sampling of specimens**

During the summer months of 2011-2015, a total of 1 446 spider specimens were collected using active sampling (sweep nets) from 16 localities in and around the eThekweni region (Figure 2.1). Sampling sites include protected areas (Bisley, Drummond, Giba Gorge, Iphithi, Hazelmere dam, Kenneth Stainbank, North Park, Springside and Vernon Crookes Nature Reserves) and undeveloped spaces within the region (Bartlett estate, Hamilton, and High Meadows), and two localities outside the eThekweni region, the grassland on the University of KwaZulu-Natal, Pietermaritzburg campus (UKZN grassland) and Bisley Nature Reserve. Killing jars containing 99% ethanol were used to kill and preserve specimens prior to sorting.

### 2.2.3 Morphological sorting of specimens

Specimens were sorted into morphospecies using published taxonomic keys (Dippenaar-Schoeman et al., 2013; Dippenaar-Schoeman and Jocque, 1997; Filmer, 1999; Leroy and Leroy, 2003). Where possible, five individuals of each morphospecies per locality were targeted for DNA extraction and amplification. Each specimen selected for DNA analysis was photographed using a USB Digital Microscope 2.0, (for example of photographs taken see Figure 2.2). Specimen associated information (photographs, taxonomy, GPS coordinates, collection date, sequencing traces; forward and reverse primers used, details of the person who collected specimens) was uploaded to the Araneae of eThekweni project in Barcode of Life Data-systems (BOLD). Specimens were stored as voucher specimens in the collection housed at the University of KwaZulu-Natal in the Conservation Genetics lab.

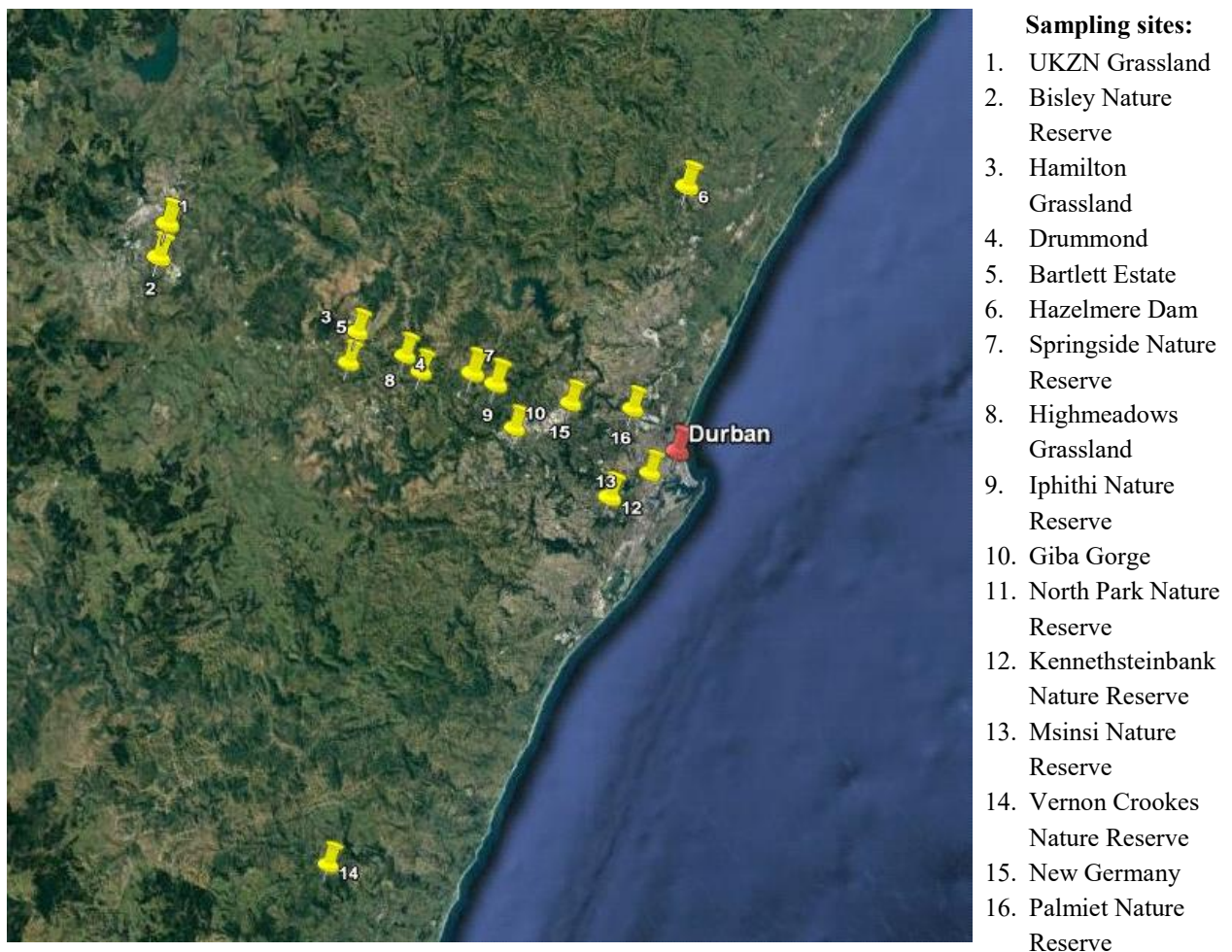


Figure 2.1 Partial map of the KwaZulu-Natal province showing all the study sites within the eThekweni Municipal area and in outlying areas. Species richness was studied across the transect starting from areas adjacent to Durban city up to inner parts of KwaZulu Natal (UKZN & Bisley Nature reserve).

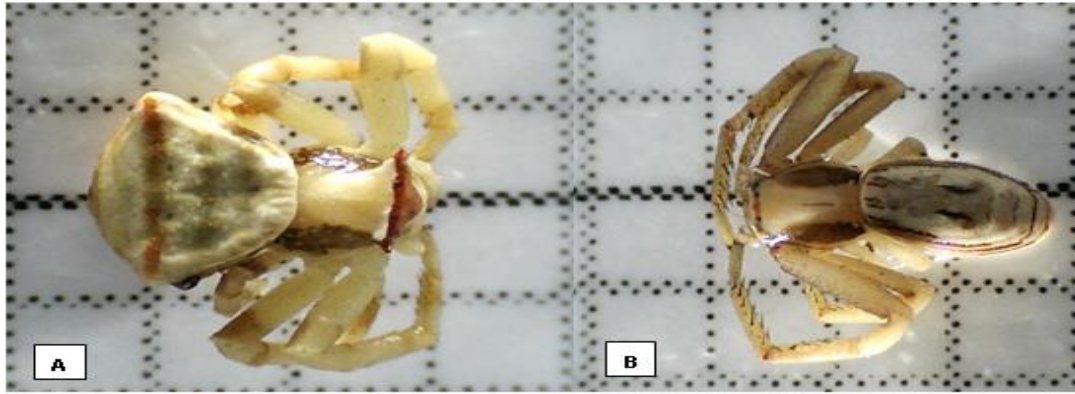


Figure 2.2 Examples of photographs taken of specimens using a Digital Image Microscope, A: HamGr6Arach12 and B: HamGr6Arach1401, taken after sorting into morphospecies, prior to DNA extraction. The body size was measured using standard graph paper.

#### 2.2.4 Genomic DNA extraction and Polymerase Chain Reaction amplification

The DNA barcoding was performed following standard protocols (Hajibabaei et al., 2005). A leg from each specimen was excised and used for DNA extraction. DNA was extracted using the ZR-96 Tissue & Insect DNA Kit<sup>TM</sup> (ZYMO Research). The kit uses chemically inert bashing beads together with an organic denaturant to lyse tissue samples, rapidly and efficiently. The DNA is isolated and purified, effectively removing PCR inhibitors by using the fast-spin column technology. The quality and concentration of the isolated genomic DNA was estimated using the NanoDrop Spectrophotometer 2000 (Thermo Scientific). From the extracted genomic DNA, the mitochondrial COI gene was amplified using universal primers (Folmer et al., 1994); LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAATCA-3'). For each 25 $\mu$ l PCR reaction, at least 20ng/ $\mu$ l of template DNA was used. Each PCR tube consisted of 2 $\mu$ l of diluted DNA (or 5 $\mu$ l of DNA in case of very low concentrations), 2 $\mu$ l of 10X DreamTaq Buffer with 20mM of MgCl<sub>2</sub>, 0.5 $\mu$ l of dinucleotide triphosphates (dNTPs) (10 $\mu$ M), 0.5 $\mu$ l of each primer (10 $\mu$ M), 2 $\mu$ l of MgCl<sub>2</sub> (25 $\mu$ M) and 1 $\mu$ l of Bovine Serum Albumin (BSA) (1mg.ml<sup>-1</sup>) then 16.4 $\mu$ l of distilled water and 0.1 $\mu$ l (5U/ $\mu$ l) of DreamTaq polymerase (manufactured by Thermo Fisher Scientific). Thermocycler amplification was carried out using the conditions described by Folmer et al. (1994) with the initial denaturation at 95°C for 3 min, and a total number of 35 cycles of denaturation at 95°C for 1 min, annealing at 40°C for 1 min, extension at 72°C for 1:30 min, followed by a final extension step at 72°C for 7 min. In some cases, however, optimization was needed to ensure that clean crisp bands were recovered.

This was done by changing the annealing temperature, which ranged from 42 to 56°C. PCR products were visualized and size estimated using 0.8 % (w/v) agarose gel stained with Ethidium Bromide (*EtBr*) against a 100-bp molecular weight ladder (Solis Biodyne). Gels were viewed under UV light using the MiniBis Pro gel capture instrument (Bio-Imaging Systems). Successfully amplified PCR products were sent to the Central Analytical Facility at Stellenbosch University for sequencing.

### **2.2.5 Sequence data processing**

Barcode compliant sequences (>500-bp of length, no stop codons or contamination) generated in the present study (n = 202) were aligned with the sequence data downloaded from BOLD (n = 966) using ClustalX 2.1 (Larkin et al., 2007). This alignment was also optimized by eye to ensure homology. Summary statistics including the number of variable characters, the number of parsimony informative characters and the average nucleotide composition were estimated for the whole alignment using MEGA5.2 (Tamura et al., 2011).

### **2.2.6 Phylogenetic analysis of the spider DNA barcode reference library**

To replicate BOLD's taxon ID tree construction (which can be performed on BOLD data through the bioinformatics platform), MEGA5.2 was used to create a neighbor-joining tree, using the COI data available through BOLD for eThekwini spiders and COI data generated in the present study (tree shown in Appendix 1). This neighbor-joining tree was constructed using the Kimura-2-Parameter (K2P) as substitution model, this is the standard model used on BOLD and in most DNA barcode research (Hebert et al., 2004b; Ward et al., 2005). This method was used to group sequences into distinct clusters known as Barcode Index Numbers (BINs). Barcode Index Numbers have been found to closely correspond to species (Ratnasingham and Hebert, 2013; Schmidt et al., 2015). To determine how this project has contributed towards the global barcoding initiative, one individual was picked randomly from each BIN cluster and blasted against the BOLD database. Sequences were matched to other records already in BOLD using sequence similarity. Sequences with sequence similarity values above 95% were considered already present in the BOLD database and where possible species name or provisional genus-level names were assigned to those specimens. Although most barcoding studies present only the neighbor-joining tree (Collins and Cruickshank, 2013) in this study maximum likelihood analysis was performed as well, using the program

RAxML v1.3 (Gunnarsson, 1990) and Bayesian inference, using the program MrBayes v3.2.2 (Huelsenbeck and Ronquist, 2001). Given that the K2P substitution model may not be the best-fit model for my data, the program jModelTest 2 (Darriba et al., 2012) was used to estimate the best-fit model of evolution. This model was then used in maximum likelihood (tree search method = ML + thorough bootstrap and number of replicates = 1000) and Bayesian analyses (number of generations = 20 million, number of chains = 4). The convergence of runs (ESS values, traces) and burn-in in Bayesian analysis was evaluated using Tracer v1.5 (Rambaut, 2009; Rambaut and Drummond, 2007), and the 50% majority rule consensus tree was generated using Phylip v3.6 (Felsenstein, 2005). Branch support was assessed using 1000 bootstrap replicates in the maximum likelihood analysis and posterior probabilities in the Bayesian analysis.

### **2.2.7 Threshold optimization and measures of identification accuracy**

Research on DNA barcoding has demonstrated that the accuracy of species delimitation is influenced by several factors including incomplete lineage sorting (Chapple and Ritchie, 2013). In situations of incomplete lineage sorting, tree-based identification methods will result in ambiguous or incorrect identifications (Lowenstein et al., 2010). Identification success rates are difficult to quantify when tree-based analytical approaches are used to assign specimens to known species (Little and Stevenson, 2007). Thus, in addition to the phylogenetic analyses (mentioned above), measures of identification accuracy of the DNA barcode data (see below) were carried out on R statistical programming environment, using the Species Identity and Evolution in R (SPIDER) package (Brown et al., 2012).

#### **2.2.7.1 Measures of identification accuracy**

The threshOpt function in the SPIDER package was used to find the optimal sequence divergence threshold for the identification of spiders in the eThekwini data. Unlike the 1% threshold used by BOLD to define BINs (Brown et al., 2012), in SPIDER the user can manipulate the threshold to best-fit the given data. The function returns the number of true positive, false negative, false positive and true negative identifications at a given threshold. Using this information the cumulative error is calculated as the sum of false negative and false positive error. The optimal threshold for the data is the one that minimises error rates (false negatives and false positives). To confirm the optimal threshold, the localMinima

function was used, unlike the threshOpt function in threshold optimization, it does not require prior knowledge of species identity to get an indication of potential threshold values.

Identification accuracy was assessed using three methods; Nearest Neighbour (Austerlitz et al., 2009), Meier's Best Close Match (Meier et al., 2006) and ThreshID (Brown et al., 2012) method. For all three methods, each individual in the data set is treated as an unknown while the remaining sequences form the reference library used for identification (Brown et al., 2012). The Nearest Neighbour criterion finds the closest individual(s) to the target and returns the species index for that individual. The result is 'TRUE' if the nearest species index or allocation is the same as the individual being tested. A 'FALSE' result would indicate that the species allocations of the target and the nearest neighbour are not the same. Similar to the Nearest Neighbour criterion are Meier's Best Close Match (similar to the method of specimen identification used by BOLD) and the ThreshID function, they find the nearest neighbour or neighbours to the query, and incorporate thresholds (percentage sequence divergence) which can be optimized to fit the given data (see above). The results indicate if the query and the nearest neighbour have been assigned to the same species (correct), the closest match and the query have been assigned to different species (incorrect), if the result is ambiguous (if query is equidistant between two species groups) or if the query is very different there will be no matches found. However, the ThreshID differs from Meier's Best Close Match in that it includes all matches within the threshold whereas Meier's Best Close Match operates upon the single nearest neighbour match (Austerlitz et al., 2009). The Best Close Match and ThreshID analyses were carried out using a range of sequence divergence values from 1-5% as threshold values (Brown and Collins, 2011; Brown et al., 2012).

### **2.2.7.2 The Barcoding gap**

Using the identification tool on BOLD, species names and provisional names were assigned to all the barcode data present in our database based on percentage sequence similarity. Sequences with similarity ranging from 97-100% were assigned to the same species, those with >95% (but below 97%) similarity were typically assigned to the same genus. To test for the presence of the 'barcode gap', intra- and inter-specific distances were calculated and compared in SPIDER using the K2P distance model. The barcode gap is calculated by subtracting the maximum intraspecific distance from the minimum interspecific distance (Barrett and Hebert, 2005a; Meier et al., 2006; Robinson et al., 2009).



### 2.3 Results

The Araneae order in BOLD currently has a total of 68 968 published records. Only 17 032 have been assigned species names and represent 2 932 species. Of these 1 057 specimens (7 614 BINs) were contributed by South Africa (with specimens collected primarily from KwaZulu-Natal, Gauteng and Eastern Cape provinces). This study has added an additional 1 153 records to BOLD (Araneae of eThekweni), representing 288 species (BINs) of spiders collected in eThekweni and surrounding areas. A total of 51 BINs could be identified to species level, while 5 could only be identified to genus and 232 to family level. Maximum likelihood and Bayesian analyses were performed on representative of specimens that could be identified to species level.

The alignment consisted of 279 individuals with COI fragments of 654-bp in length with 355 variable characters of which 338 were parsimony informative. The nucleotide composition shows the sequences are AT rich (T = 43.0; A = 25.7; G = 18.7; C = 12.6).

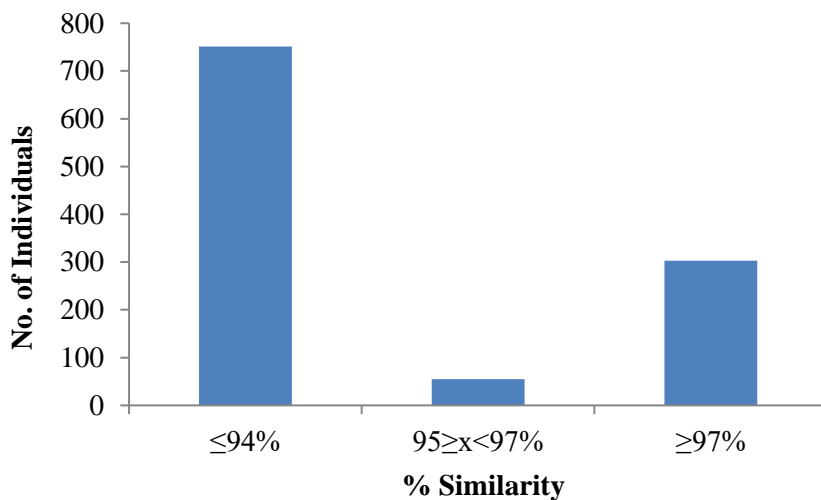


Figure 2.3 Barplot showing the number of species already represented in BOLD (greater than 95% similarity) and new species added to the database (less than 95% similarity).

A total of 51 (303 specimens) out of 288 (1 153 specimens) barcode clusters exhibited a 97% sequence similarity or higher match when using BOLD's specimen identification tool. Only 5 BIN clusters showed 95% or higher but less than 97% sequence similarity, while and the rest of the barcode clusters had a below 95% sequence similarity match (Figure 2.3). This suggests that only 56 BINs or genetic species of 288 are already represented in the database.



Table 2.1 The results of BLAST search against BOLD using one representative per BIN. Only matches 95% and above are shown in the table.

BIN Number	Family	Top search result (BOLD)	% Match	Taxonomic assignment
ABX7722	Araneidae	<i>Araneus</i>	97.45	<i>Araneus_sp.MA22</i>
ABY1256	Araneidae	<i>Argiope bruennichi</i>	99.21	<i>Argiope_bruennichi</i>
AAQ2634	Araneidae	<i>Argiope bruennichi</i>	97.28	<i>Argiope_bruennichi</i>
ACH1546	Araneidae	<i>Caerostris sexcupidata</i>	95	<i>Caerostris_sexcupidata</i>
ABX8168	Araneidae	<i>Cyrtophora</i>	100	<i>Cyrtophora_sp.AR2013</i>
ACG0526	Araneidae	<i>Cyrtophora</i>	97.84	<i>Cyrtophora_sp.</i>
ACF6522	Araneidae	<i>Cyclosa deserticola</i>	97	<i>Cyclosa_deserticola</i>
ACB3425	Araneidae	<i>Hypsosinga</i>	99.03	<i>Hypsosinga_sp.HY25</i>
ACB3650	Araneidae	<i>Hypsosinga</i>	100	<i>Hypsosinga_sp.HY500</i>
ACB1958	Araneidae	<i>Larinia</i>	99.85	<i>Larinia_sp.</i>
ACH1367	Araneidae	<i>Parawixia</i>	96	<i>Parawixia_sp.</i>
ACB3509	Clubionidae	<i>Clubiona</i>	99.69	<i>Clubiona_sp.</i>
ABY1687	Corinnidae	<i>Hortipes mesembrinus</i>	99.2	<i>Hortipes_mesebrinus</i>
ACF7562	Corinnidae	<i>Pronophea atlantica</i>	99.85	<i>Pronophea atlantica</i>
ACB2081	Eutichuridae	<i>Cheiracanthium inclusum</i>	99.37	<i>Cheiracanthium_inclusum</i>
ACC7131	Mimetidae	<i>Mimetus</i>	99.84	<i>Mimetus_sp.</i>
ACG2148	Miturgidae	<i>Cheiramiona</i>	100	<i>Cheiramiona_sp.</i>
ACC7109	Oxyopidae	<i>Oxyopes javanus</i>	96	<i>Oxyopes_javanus</i>
ABX8173	Oxyopidae	<i>Oxyopes heterophthalmus</i>	95.46	<i>Oxyopes_heterophthalmus</i>
ACB1954	Philodromidae	<i>Tibellus</i>	97.36	<i>Tibellus_sp.</i>
ACH1763	Philodromidae	<i>Suemus</i>	100	<i>Suemus_sp.</i>
ABY0205	Salticidae	<i>Evarcha</i>	99.84	<i>Evarcha_sp.</i>
ACC7234	Salticidae	<i>Asemonea clara</i>	100	<i>Asemonea_clara</i>
ACB1955	Salticidae	<i>Heliophanus</i>	100	<i>Heliophanus_sp</i>
ACB3559	Salticidae	<i>Baryphas ahenus</i>	97	<i>Baryphas_ahenus</i>
ACC7285	Salticidae	<i>Brancus mustelus</i>	100	<i>Brancus_mustelus</i>
ABX8936	Salticidae	<i>Baryphas</i>	98.93	<i>Baryphas_sp.</i>
ACG1003	Salticidae	<i>Festucula</i>	100	<i>Festucula_sp.</i>
ABX9604	Salticidae	<i>Heliophanus orchestra</i>	100	<i>Heliphanus_orchestra</i>
ACB3695	Salticidae	<i>Heliophanus</i>	100	<i>Heliophanus_sp.12</i>
ABX7720	Salticidae	<i>Hyllus dotatus</i>	99.68	<i>Hyllus_dotatus</i>
ACC7108	Salticidae	<i>Hyllus argyrotoxis</i>	100	<i>Hyllus_argyrotoxis</i>
ACB3612	Salticidae	<i>Myrmarachne lulengana</i>	97.31	<i>Myrmarachne_lulengana</i>
ACH4432	Salticidae	<i>Thyenula</i>	100	<i>Thyenula_sp.MT-051</i>
ACF6523	Salticidae	<i>Pseudicius</i>	97.08	<i>Pseudicius_sp.</i>

Table 2.1 Continued

BIN Number	Family	Top search result (BOLD)	% Match	Taxonomic assignment
ACC7093	Salticidae	<i>Tusitala barbata</i>	100	<i>Tusitala_barbata</i>
ABY0987	Scytodidae	<i>Scytodes</i>	99.84	<i>Scytodes_sp.</i>
ABX7721	Selenopidae	<i>Anyphops</i>	97.63	<i>Anyphops_sp</i>
ACB3622	Sparassidae	<i>Palystes</i>	99.69	<i>Palystes_sp.</i>
ABY1258	Tetragnathidae	<i>Leucauge</i>	100	<i>Leucauge_sp.LE5800</i>
ABY1818	Tetragnathidae	<i>Tetragnatha</i>	100	<i>Tetragnatha_sp.</i>
AAU5366	Therididae	<i>Latrodectus tredecimguttatus</i>	99.35	<i>Latrodectus_tredecimguttatus</i>
ACH4419	Theridiidae	<i>Steatoda</i>	97.22	<i>Steatoda_sp.</i>
ACB3563	Thomisidae	<i>Oxytate argenteooculata</i>	100	<i>Oxytate_argenteooculata</i>
AAQ0108	Thomisidae	<i>Runcinia affinis</i>	99.68	<i>Runcinia_affinis</i>
ABX9609	Thomisidae	<i>Tmarus</i>	99.69	<i>Tmarus_sp.</i>
ABY0832	Thomisidae	<i>Monaeses</i>	100	<i>Monaeses_sp.</i>
ABY1690	Zodariidae	<i>Chariobas</i>	100	<i>Chariobas_sp.</i>
ABY1690	Zodariidae	<i>Thaumatochilus</i>	99.23	<i>Thaumatochilus_sp.</i>
ACG1142	Salticidae	<i>Thyene pulchra</i>	100	<i>Thyene_pulchra</i>
ACH1375	Salticidae	<i>Thyenula</i>	95.99	<i>Thyenula_sp.</i>
ACB3613	Salticidae	<i>Thyene natalii</i>	98.61	<i>Thyene_natalli</i>
ABX9477	Salticidae	<i>Thyene</i>	100	<i>Thyene_tsp.</i>
ACB2152	Salticidae	<i>Thyene thyenioides</i>	100	<i>Thyene_thyenioides</i>
ABY0835	Salticidae	<i>Massagris natalensis</i>	100	<i>Massagris_natalensis</i>
ABY0986	Salticidae	<i>Maeota</i>	100	<i>Maeota_sp.</i>

Sequences with greater than 95% sequence similarity when searched against BOLD were assigned genus name and species names where applicable. Only 23 BINs were matched to species and 33 BINs matched to genus level (Table 2.1). The families, Salticidae (15 genera), Araneidae (7 genera) and Thomisidae (4 genera) were well represented in my eThekwini data.

### 2.3.1 Threshold optimization and measures of identification accuracy

The Nearest Neighbour criterion identified 279 'TRUE' matches (nearest species has name the same as the query) and 0 'FALSE' identifications (no species names match the query). The least cumulative error using the threshold optimization function was obtained at the 3% threshold level, at this level the number of true positives was the highest. The cumulative

error was seen to increase when the threshold was adjusted beyond the 3% level and the highest cumulative error was observed at the 5% threshold. The results are shown in Table 2.3. A similar threshold of 3.17% was reported by the localMinima function.

Table 2.2 The results of the threshold optimization analysis conducted on SPIDER. Only thresholds ranging from 1.0 -5.0% were tested.

Threshold (%)	True negative	True positive	False negative	False positive	Cumulative error
1	0	253	0	26	26
2	0	274	0	5	5
3	0	279	0	0	0
4	0	269	10	0	10
5	0	249	30	0	30

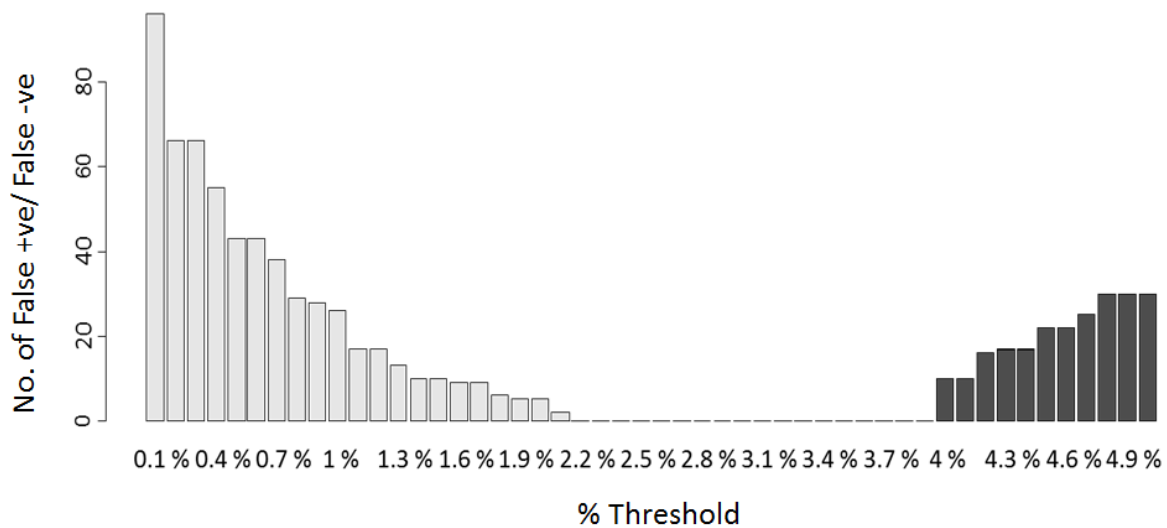


Figure 2.4 Barplot of false positives (light grey) and false negative (dark grey) rate of identification for the Araneae of eThekwini data set using a range of threshold values from 1.0 to 5.0%. As the percentage threshold for species delimitation increases, the number of false negatives increases.

The Best Close Match and ThreshID criteria (Figure 2.5) yielded similar results and were consistent with the results obtained by the least cumulative error method (Figure 2.4 and Table 2.2), the number of correct identifications increased with an increase in threshold value until 3% threshold was exceeded, thereafter the number of correct identifications decreased as ambiguous identifications increased. Consistency of findings across the different methods suggest that for the eThekwinini data a 3% sequence divergence value is the optimal threshold value for accurate delimitation of species.

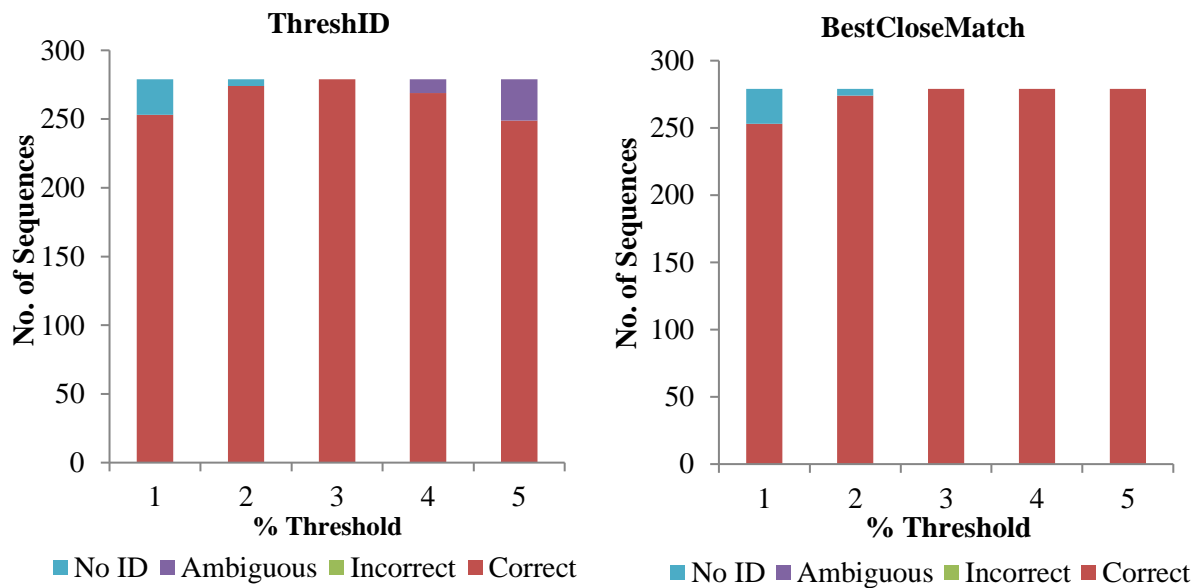


Figure 2.5 Barplot of the success rate using ThreshID identification criterion and the Best Close Match at 1-5% thresholds. The optimal threshold for identification for both criteria was 3%.

### 2.3.2 The Barcoding Gap

The computed K2P sequence divergence values for COI were greater between species (ranging from 0.035 - 0.354) than within species (ranging from 0.0 - 0.020). There was also no overlap in inter- and intraspecific divergence values, suggesting the presence of a barcoding gap. The barcode gap was observed between the 0.020 - 0.035 range (Figure 2.5). The clear separation in the distribution of intra- and interspecific distances suggests COI is a good marker for use in spider species delimitation within the eThekweni region.

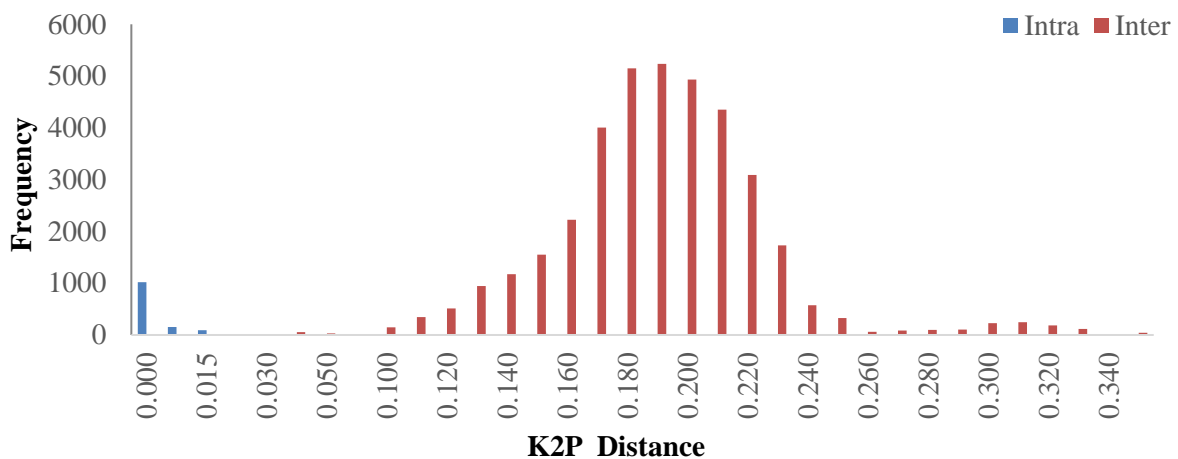


Figure 2.6 Distribution of pairwise divergence values calculated using the K2P model. There is no overlap between the intra- and interspecific distances. The gap occurred between the 0.020 - 0.035 range of K2P distance.

### 2.3.3 Phylogenetic analysis

Applying the best-fit model, GTR+G+I, the maximum likelihood (Figure 2.8) approach and the Bayesian Inference (Figure 2.7) recovered very similar phylogenies and differ only in placement of certain genera (for example in maximum likelihood *Scytodes* and *Mimetus* occur as sister taxa while they are not clustered together in Bayesian tree) are these significant (i.e. high bootstrap support or posterior probability support for alternative hypotheses) and level of branch support. The phylogenies produced by both methods recovered the families Salticidae (ML bootstrap: <75; Bayes PP: 0.76) an Araneidae (ML bootstrap: <75; Bayes PP: 0.99) as monophyletic. The Thomisidae and Therididae families were not monophyletic in both trees. In general, more clades (within families) in the maximum likelihood tree were well supported (bootstrap values  $\geq 75\%$ ) while in the Bayesian tree only clades within the Araneidae family were well supported (posterior probabilities  $\geq 0.95$ ).

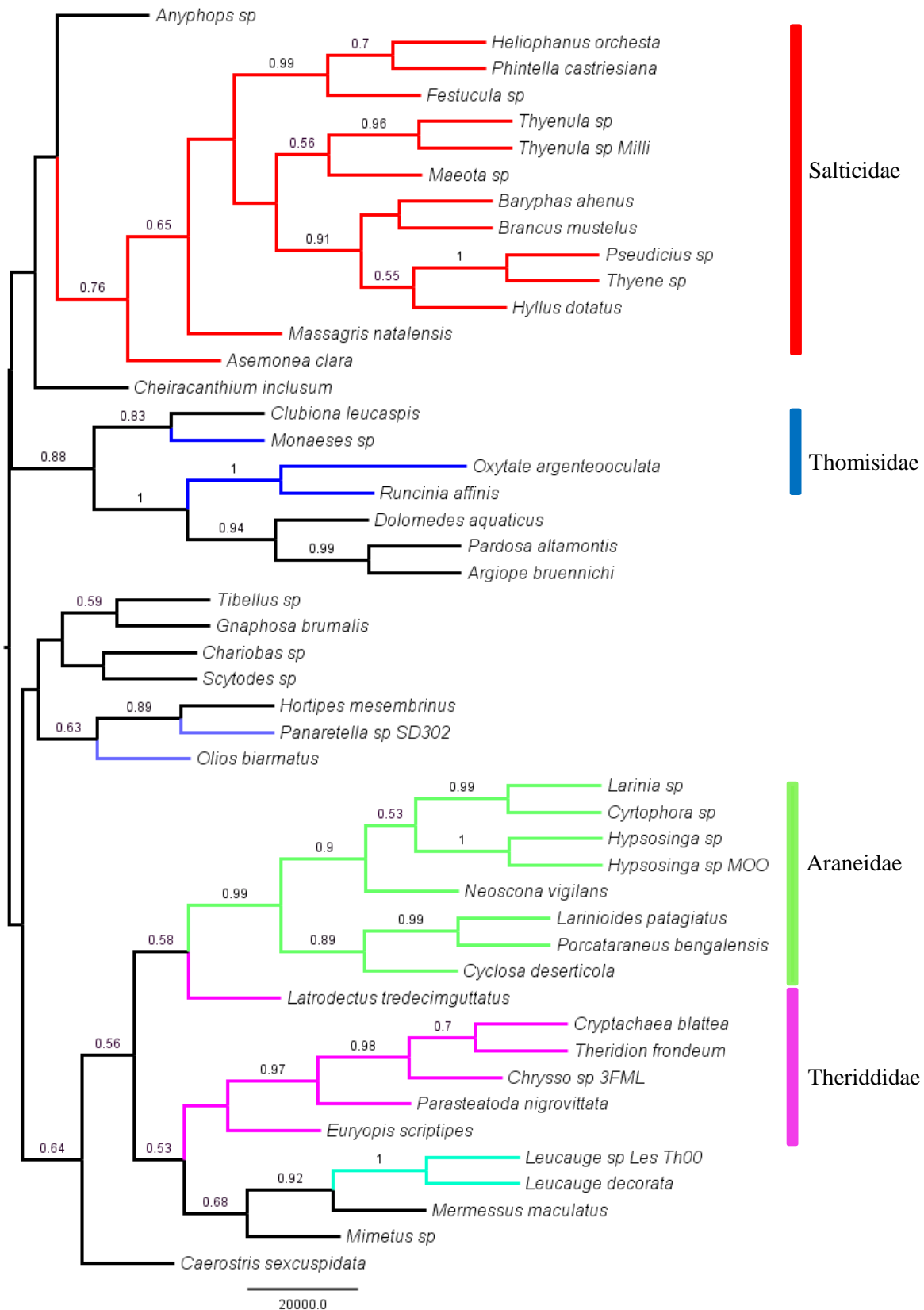


Figure 2.7 Bayesian tree of eThekweni spiders including 48 species from 45 genera. The tree is midpoint rooted and only posterior probabilities above 0.5 are presented in the tree. Only major family groups, represented by at least three genera, are highlighted.

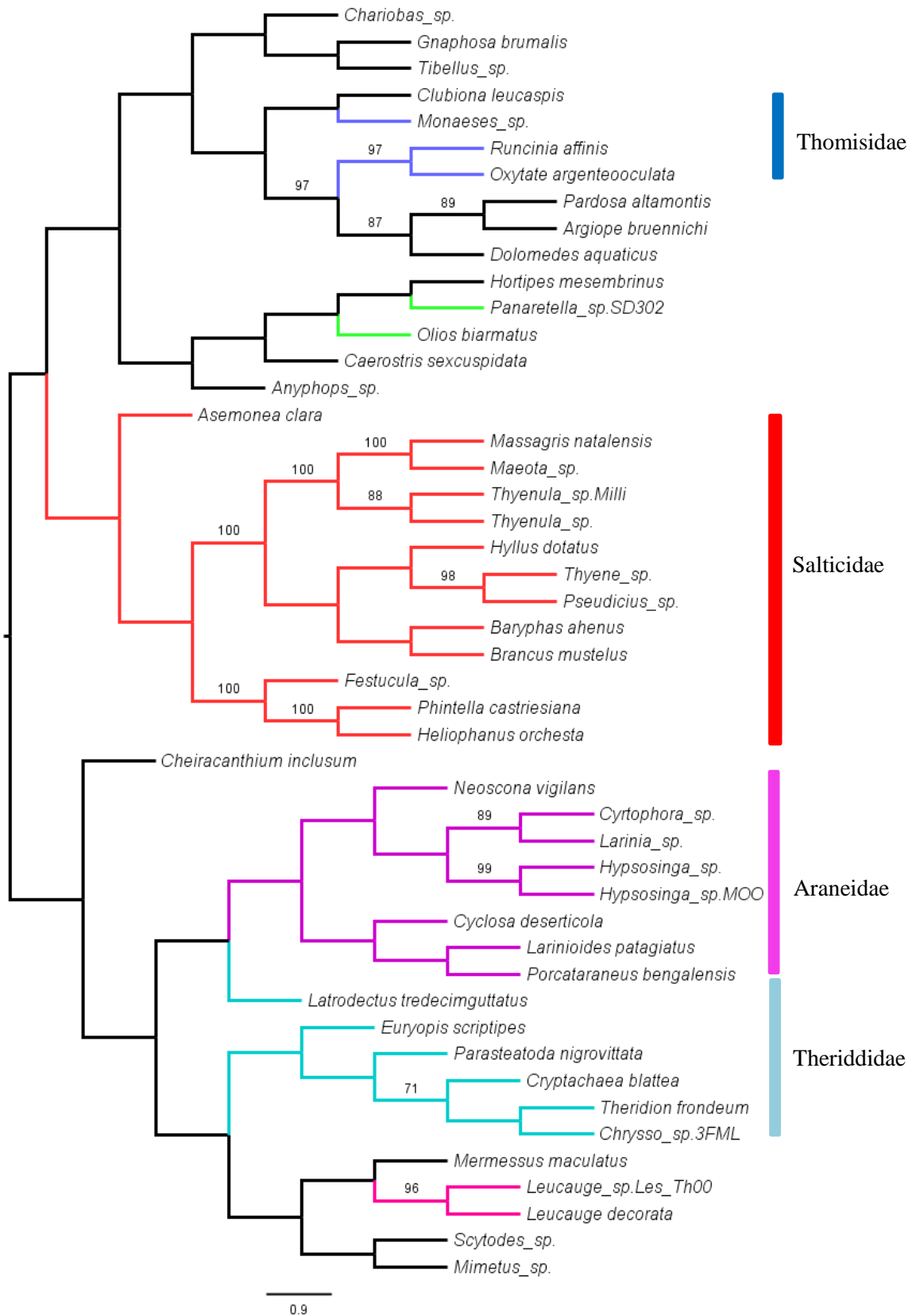


Figure 2.8 Maximum likelihood tree including 48 species of eThekwi spiders from 45 genera. The tree is midpoint rooted and bootstrap values greater than 70% are shown. Only major family groups, represented at least by three genera, are highlighted.



## 2.4 Discussion

DNA barcoding provides a rapid way of identifying unknown specimens to previously described species or families, by comparing an unknown query barcode sequence to barcodes of already identified species in the BOLD database. The successful application of DNA barcoding relies on creation of a good reference library that has good taxonomic and intraspecific coverage. Most studies that use barcoding to identify species use genetic distance to assign a query sequence to a species BIN (Hebert et al., 2003; Hebert et al., 2003b; Krishnamurthy and Francis, 2012). Although this works well in most cases problems can arise when there is an overlap in inter- and intraspecific divergence values.

When a specimen belongs to a species not yet represented in the DNA barcode reference library DNA barcoding will fail to identify that individual, instead the query will be assigned to higher taxon (Wilson et al., 2011). Identifications using DNA barcoding are carried out by comparing genetic distances of sequences; query against reference library, the query is then assigned the name of its highest match (sequence similarity). While the current study made an enormous contribution to BOLD (1 153 barcode records of at least 288 genetic clusters) only a fraction of these could be identified to species and genus level. This indicates the deficiency in the database in terms of the taxa already represented. However using only a fraction of the eThekwini data, results indicate a great potential for spider identification using COI.

The main aim of this chapter was to test for the utility of DNA barcoding for the accurate identification of spider species collected from the eThekwini region. In particular the presence of the DNA barcode gap was tested for and threshold optimization and measures of identification accuracy. The small scale of the region in the current study is an important consideration as many of the species are expected to be closely related (Bergsten et al., 2012) As such this study is an excellent test of the utility of the DNA barcode method. Using individual representatives of the 48 species from 45 genera, phylogeny-based analysis (maximum likelihood and Bayesian inference) indicated that the COI sequences are phylogenetically informative enough to provide taxonomic information at family and generic level as most genera were clustered together to their respective families with a few exceptions in the Thomisidae and Therididae families which were polyphyletic.

The presence of the barcode gap confirms the utility of DNA barcoding in spiders of the eThekweni region (Figure 2.6). There was no overlap between maximum intraspecific (mean = 0.020) and minimum interspecific divergence (mean = 0.035). This result is comparable to the findings of previous study testing COI on spiders by Barrett and Hebert (2005b), where congeneric pairs examined were found to possess at least 3% divergence. The range in which the barcode gap occurs may, however, differ across data sets and it is important that each study that uses barcoding include barcode gap analysis as a critical step. One of the criticisms of DNA barcoding is it tends to overestimate or underestimate the number of species, this can be corrected for by changing the threshold value. The threshOpt function indicated an optimal threshold value of 3% (yielded 100% identification success). This result was also confirmed by both ThreshID and Best Close Match criteria (Figure 2.5). There seems to be no universal threshold for the Araneae order because of very different ranges of genetic divergence from other studies, for example intraspecific divergences of 0.0 to 0.09, (Candek and Kuntner, 2015; Hebert et al., 2003b) and intraspecific divergence ranging from 0.0 to 0.031 (Robinson et al., 2009).

## **2.5 Conclusion**

DNA barcoding is a molecular tool for species identification for a wide range of taxonomic groups. In this study, DNA barcoding proved to be effective in delimiting spider species and be used in future studies within the area to potentially discover of instances of cryptic speciation and can be used to successfully assess spider diversity. This however was a preliminary study, barcoding was assessed based on a few species despite the amount of data this project has uploaded to BOLD. This indicates the deficiency in number of identified taxa in the database. Additional sampling and sequencing to ensure presence of barcode gap at large scale of sampling is required before full reference library will be available for use.

## **Chapter Three: Assessing the Spatial Distribution of Spider Species in eThekweni using DNA Barcoding**

### **Abstract**

**Quantifying species richness is useful in conservation planning and biodiversity assessment. Assessing species richness for a biodiverse region such as eThekweni can be time consuming as accurate taxonomic assignment of specimens requires a substantial input in terms of both expertise and time. In this chapter I use the DNA barcode data to assess spider diversity across open spaces within the eThekweni region. In total 1 153 specimens were collected from 16 localities, within and around eThekweni. These samples were divided into 290 morphospecies, and genetic analysis revealed 288 genetic clusters (BINs). The number of morphospecies was compared to the number of genetic (BIN) clusters per locality to determine how well the genetic data differentiates species. There was no significant difference between the numbers of morphologically and genetically delimited species and this result suggests that barcode data can be used successfully in biodiversity assessments in this region. On average, 18 BINs and 18 morphospecies were observed per locality, Springside, Iphithi and Palmiet nature reserves were the most diverse localities (haplotypes >80) Bartlett estate and Hazelmere dam were the least diverse (haplotypes <5). The haplotype accumulation curves indicate that much more diversity is present in these open areas than is captured by this study, and future studies would benefit from further sampling. The Michaelis-Menten estimator predicted at least 591 species could be encountered while Chao 2 measure reported 527 and the Jackknife 1 estimator 450 potential species. Patterns of species distribution were associated with vegetation type, with forest habitats sharing more species not found in grasslands and vice versa. There was a weak correlation ( $R^2 = 0.01$ ) between the number of species and open space area size. There was a weak correlation ( $R^2 = 0.03$ ) between number of species and degree of urbanisation surrounding open spaces. The use of DNA barcode data provides a great alternative and compliment to traditional morphology-based taxonomy for biodiversity assessment as it is quick and does not rely on extensive taxonomic knowledge.**

### **3.1 Introduction**

Transformation of natural ecosystems by urbanisation is known to have a negative impact on species diversity and richness (Magura et al., 2008). To effectively minimize biodiversity loss, urban planners and managers need to be able to consider areas of high conservation value off limits for development. Deciding which areas need protection should be based fundamentally on biodiversity data, and areas with high species richness and abundance should be prioritized for conservation (Brooks et al., 2006). Biodiversity assessment is, however, a difficult task especially in urbanised areas where viable taxa such as large native vertebrates are often lost as a result of habitat loss and conflict with humans (Lowe et al., 2014). Using information on invertebrate species richness could provide valuable data for ecologists and conservationists (Krishnamurthy and Francis, 2012; Kvist, 2013). Monitoring invertebrate biodiversity is difficult as it involves sampling the entire invertebrate assemblage. This inevitably involves vast numbers of specimens. A far more common alternative is to focus on one or more indicator groups that reflect broader patterns of invertebrate biological integrity (Andersen et al., 2004a). For example, ants in Australia (Andersen et al., 2004a), beetles of the Carabidae family in the northern hemisphere (York, 1999) and spiders in Belgium (Maelfait and Hendickx, 1998) and South Africa (Muelelwa et al., 2010) have been used. In this study the diversity and distribution of spider species across an urban transect through the eThekweni region is assessed. The sampling sites included in this study differ in terms of degree of urbanisation surrounding open space (for example, sites very close to the Durban harbour are surrounded by heavily transformed areas), vegetation type (forest or grassland) and size of the open space.

#### **3.1.1 Spiders as bioindicators**

In Durban, reliable and quantitative datasets for most groups of organisms are difficult to obtain because of the lack of long term monitoring initiatives (Govender, 2014). This lack of quantitative data makes attempts to study the effects of urbanisation on the region a difficult task. In this study spiders are evaluated for their practical use as bioindicators. Arthropods in general are often used as indicators of biodiversity because they are abundant and sensitive to environmental health (Yu et al., 2012). Another advantage of using arthropods as bioindicators is the high rate of spatial turnover (replacement of species) which can provide information on biodiversity at a scale which conservation decisions are typically made (Speight et al., 2008; Yu et al., 2012).

Spiders are useful taxa in which to study the effect of urbanisation as they can have large effects on food webs and ecosystems traits including herbivore abundance, plant community composition and nutrient cycling (Hodkinson et al., 2001; Schmitz et al., 2010). While spiders are good bioindicators because of their ability to thrive in a variety of terrestrial ecosystems, spider taxonomy can be challenging to non-specialist researchers. DNA barcoding could provide a useful tool that can be used by non-taxonomic experts for spider species identification. Previous studies concerning species richness and abundance of spiders in response to anthropogenic disturbance show varied responses. Some studies found no change in species richness in response to urbanisation (Alaruikka et al., 2002) while others found an increase in species diversity (Magura et al., 2010) or an increase in total abundance in a more urbanised environment (McIntyre, 2000). The latter response could be the result of certain species which are urban exploiters (Bolger et al., 2008; Shochat et al., 2004). Given the differences between published studies generalization about spider diversity along a urban transect is difficult as environmental variables such as presence of alien species (among others) may differ between cities (Niemela et al., 2002).

### **3.1.2 DNA barcode data to evaluate levels of biodiversity**

The majority of conservation programs and legislation focus on saving species, thus ‘species’ is the principal currency of biodiversity and is usually the focal taxonomic unit of conservation biology. Accurate, unambiguous and robust species identification are an essential component of conservation management decisions (Rubinoff, 2006). The DNA barcode data can circumvent the problem of having to morphologically assign specimens to species, as Molecular Operational Taxonomic Units (MOTUs) and BINS have in many cases been shown to be analogous to species (Blaxter et al., 2005; BOLDSYSTEMS, 2013). Estimates of genetic diversity represent a valuable resource for biodiversity assessments and are increasingly used to guide conservation and management programs (Goodall-Copestake et al., 2012). Estimation of biodiversity indices can be based on BINS detected using the barcoding approach, where the relative abundances of each BIN can be used to calculate biodiversity indices such as species richness, Shannon’s or Simpson’s indices (Blackwood et al., 2007; Wu et al., 2011).

### **3.1.3 Vegetation type and biodiversity as a function of urbanisation**

Urbanisation modifies landscapes at multiple scales, impacting the local climate (Lowe et al., 2014; Palmer et al., 2009) and changing the extent and quality of natural habitats (Alberti, 2005; McKinney, 2002). Urbanisation also leads to the fragmentation of habitats, and any increase in human housing and population densities can change surrounding land cover from predominantly vegetation to a matrix of hard surfaces, industry and parklands. These broad scale vegetation changes have been shown to alter species distribution (Magura et al., 2008; McKinney, 2008; Shochat et al., 2004). Biodiversity responds to urbanisation at multiple spatial scales (Alaruikka et al., 2002; Niemela et al., 2002; Shochat et al., 2004). In this study biodiversity is measured across an urban gradient from the heavily urbanised centre of the city to outlying natural areas. As such this study provides an interesting case study on the effect of urbanisation on biodiversity.

## **3.2 Aims**

The main aim of this chapter is to test if the DNA barcode data can be used as surrogate to traditional morphology-based taxonomy in assessing spider species diversity in the eThekweni region. The main objectives of this chapter are:

- (i) Using the barcode data the study aims to assess the diversity and distribution of spider species across the different study sites in the eThekweni region.
- (ii) Compare the diversity and distribution of spider species and relate that to the degree of urbanisation of area surrounding each study site and size of green space. The study sites follow an urbanisation gradient (from city hub which is almost completely transformed to less urbanised outlying areas near Pietermaritzburg)
- (iii) Determine if spider diversity is linked the two main vegetation types found in the region (grassland versus forest).

### **3.3 Materials and Methods**

In this chapter, data from the DNA barcode library for the Araneae of eThekwini was used (see Chapter 2). Each barcode record included the 654-bp COI sequence data, GPS coordinates of the collection site, collection site vegetation type and photograph of each specimen. All specimens were sorted into morphospecies using available taxonomic keys (Dippenaar-Schoeman et al., 2013; Dippenaar-Schoeman and Jocque, 1997; Foord et al., 2011). A simple classification method for different vegetation types was used; sites were identified as either grasslands or forests. Grasslands included grass patches, open spaces dominated by grass (with shrubs present or absent). Each site was sampled more than once, with some sites Palmiet NR, UKZN, and Giba Gorge sampled on several occasions (Table 3.1).

#### **3.3.1.1 Species accumulation curves**

Accumulation curves were generated for the complete data set and also for each site based on COI haplotype diversity of specimens barcoded using SPIDER R (Brown et al., 2012). Because the actual number of samples differed from one site to another, the number of sequences or samples was set to 400 per site. Sites like Bartlett Estates, Hazelmere dam and High Meadows grassland are not shown, because very few samples were collected from those sites. These accumulation curves also provide an estimate of sampling effort (Bolger et al., 2008). Estimates of species richness based on the actual number of observed species can represent underestimates of the true species richness as rare species could be missed during sampling (Colwell, 2013). To correct for this bias three extrapolation methods were also used to predict the number of species present within the eThekwini region. The Chao 2 (Leroy and Leroy, 2003), Jackknife 1 (Filmer, 1999; Leroy and Leroy, 2003) and Michaelis-Menten (Dippenaar-Schoeman et al., 2013; Soria-Carrasco et al., 2007), methods were used as they have been shown to be very useful in estimating species richness when species inventories are incomplete (Soria-Carrasco et al., 2007).

Table 3.1 Detail of sampling conducted at 16 sites through a period of 2011-2015. The co-ordinates of the exact locations are presented, the size of each open space, vegetation classification, desistance from city hub as well as the number of times each open space was sampled.

Sites	GPS coordinates		Size of open space (ha)	Vegetation classification	Distance from harbor (Km)	No. of times sampled	number of BINs	Number of morphospecies
	Lat DDS	Long DDE						
Bartlett Estates Grassland	-29,7663	30,62469	40.98	Grass	31.62	2	4	3
Bisley Nature Reserve	-29,6622	30,3902778	250	Grass	66.50	2	12	12
Hamilton Grassland	-29,7365	30,63099	15.03	Grass	35.35	5	24	24
Drummond Grassland	-29,7653	30,6830278	-	Grass	42.10	3	30	30
Giba Nature Reserve	-29,81	30,7779722	72	Forest; Grass	47.35	5	23	24
Hazelmere Dam	-29,6003	31,04167	350	Forest; Grass	27.58	3	3	3
High Meadows	-29,7824	30,7144444	4.01	Grass	21.22	4	19	18
Iphithi	-29,7886	30,803	12	Forest; Grass	25.1	4	77	77
Kenneth Stainbank Nature Reserve	-29,9099	30,9354	253	Forest; Grass	11.55	2	44	42
Msinsi Nature Reserve	-29,867	30,9755167	7	Forest; Grass	5.90	4	18	18
New Germany Nature Reserve	-29,8127	30,8911833	110	Forest; Grass	15.66	2	35	34
North Park Nature Reserve	-29,8742	30,88155	52	Forest; Grass	15.38	2	7	7
Palmiet Nature Reserve	-29,8233	30,9279167	90	Forest; Grass	9.26	7	66	67
Springside Nature Reserve	-29,7816	30,7762167	21	Forest; Grass	33.35	4	72	72
UKZN Pietermaritzburg	-29,6269	30,403883	4.64	Grass	63.48	6	52	51
Vernon Crookes Nature Reserve	-30,2681	30,6122667	2 189	Grass	64.50	3	25	27



### 3.3.2 Spatial distribution of spider diversity

To test whether the barcode data performs as well as morphology-based determination of species, the number of morphospecies per locality were compared to the number of BINs recovered by the molecular data. Given that the number of morphospecies was so similar to the number of BINs (see below) subsequent analyses were carried out on the barcode data treating BINs as species. Species richness is the most commonly used measure of biodiversity however it is not sufficient for exhaustive biodiversity assessment as it only takes into account the number of species present in the sample while ignoring the dominance or unevenness of species in relation to one another. The latter is important when one wants to compare different communities' structures, and different diversity indexes are used for this purpose.

Species richness and abundance data were used to compute the four most commonly used diversity indices, namely, Margalef's diversity index  $d$  (Dippenaar-Schoeman and Jocque, 1997), Simpson's diversity index  $D$  (Andersen et al., 2004b; Hofreiter, 2012), Shannon's diversity index  $H'$  (lahaye et al., 2008) and Fisher's  $\alpha$  (Alaruikka et al., 2002) to describe the spider assemblage at each site using the program Past v3.11 (Wiemers and Fiedler, 2007). The Margalef's diversity index is calculated from the total number of species and the total number of individuals present in the sampling area (Dippenaar-Schoeman and Jocque, 1997; Hofreiter, 2012). The Simpson's diversity index measures the evenness of the community; it is the probability that in an infinitely large community, any two individuals sampled at random will belong to the same species. The values calculated range from 0 (all taxa are equally present) to 1 (one taxon dominates the community completely) (Hofreiter, 2012). Like Simpson's index, Shannon's diversity index accounts for both abundance and evenness of the species present (Hajibabaei et al., 2011). Fisher's  $\alpha$  diversity index describes how individuals sampled are divided among species in the sampling area, it is a scale independent indicator of biodiversity which assumes that species abundance follows a log distribution (Alaruikka et al., 2002; Hajibabaei et al., 2011). These diversity indexes have been used previously on terrestrial invertebrate studies (Death, 2002; Hazarika, 2013; Human and Gordon, 1997).

Similarities between sampling sites based on species occurrence and composition and vegetation type were investigated, using Bray-Curtis cluster analysis performed using the program BioDiversity Pro v.1 (Magura et al., 2010).

To determine what effect degree of urbanisation has had on spider diversity within the eThekweni region, the relationship between distance from harbour (main urban hub) and spider diversity was examined by performing a correlation test in Microsoft Excel 2010. The distance of each sampling locality from the Durban harbour was estimated using Google Maps (<https://maps.google.com>). If urbanisation has a negative effect on South African spider species, there has to be a positive correlation between distance from urban hub and spider diversity. Alternatively if urbanisation has no effect then there should be no significant correlation. The effect of area size of sampling localities on species richness was also examined. Area sizes were obtained from the KZN Wildlife website and others measured on Google Maps. We expect the number of species to increase with area size of sampling locality.

### **3.4 Results**

Morphology-based species identification was compared to the number of BINs recovered from the genetic analysis for each locality, the 1 153 specimens were divided into 290 morphospecies and genetic analysis revealed 288 BINs. At most localities the number of morphospecies was very similar to the number of barcode clusters with a few exceptions e.g. Bartlett estate, High Meadows and Vernon Crookes nature reserve. This observation suggests DNA barcodes can be used in place of traditional taxonomy to provide rapid and reliable biodiversity estimates.

The haplotype accumulation curve constructed using all data from all localities (1 153 barcode sequences (Figure 3.1) has not reached a plateau. While the haplotype accumulation curves for individual sites (Figure 3.2) may seem to have reached plateau it is important to note the simulation was carried at a higher number of samples (400 per site) than were actually present as most sites had 50 samples on average. This suggests that species diversity is underestimated in our data and that sampling effort should be increased.

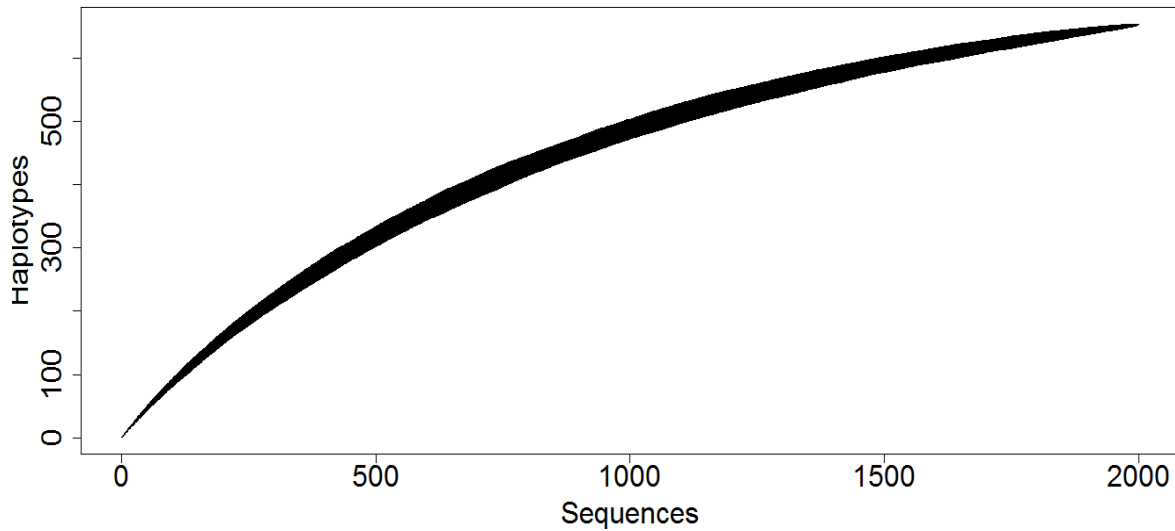


Figure 3.1 Haplotype accumulation curve for 1 153 DNA barcodes generated on SPIDER, using 1000 “random permutations” and the number of sequences set to 2000. The poor gradient indicates the need for additional sampling.

Palmiet, Springside and Iphithi nature reserves (>80 haplotypes) were the three most diverse sites included in the study in terms of haplotype diversity, while Drummond, Hamilton grassland, Msinsi and Vernon Crookes nature reserves were the least diverse (Figure 3.2 & 3.3).

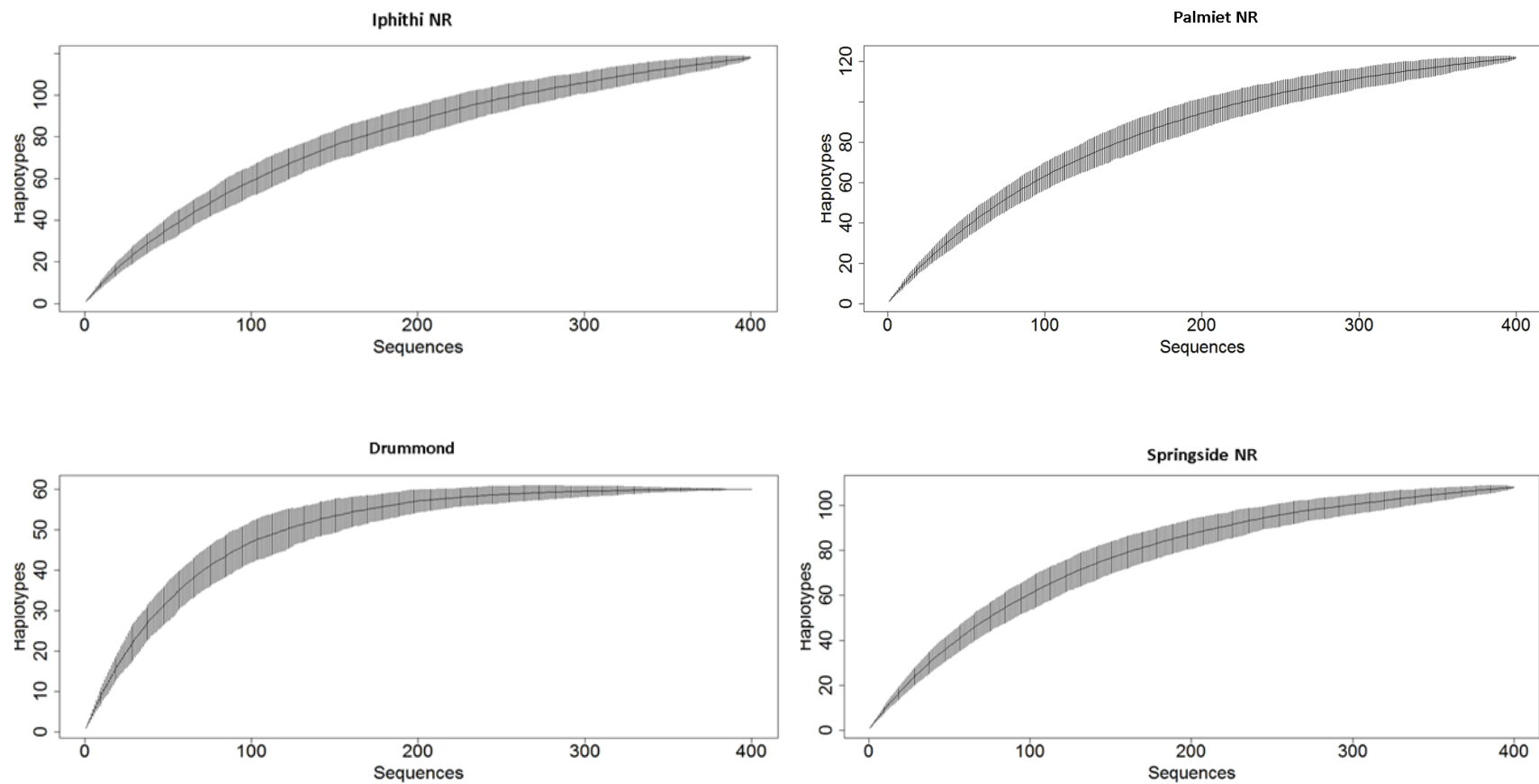


Figure 3.2 Haplotype accumulation curves for Iphithi, Palmiet, Springside NRs and Drummond grassland. The numbers on the y-axis represent the number of haplotypes observed and on the x-axis is the number of sequences barcoded from each locality. The species accumulation curve is the solid black line, which represents the average of 100 random draws, sampling with replacement, at each level of abundance. The shaded envelope represents a symmetric 95% bootstrap confidence interval, calculated from the estimated variance of the random draws.

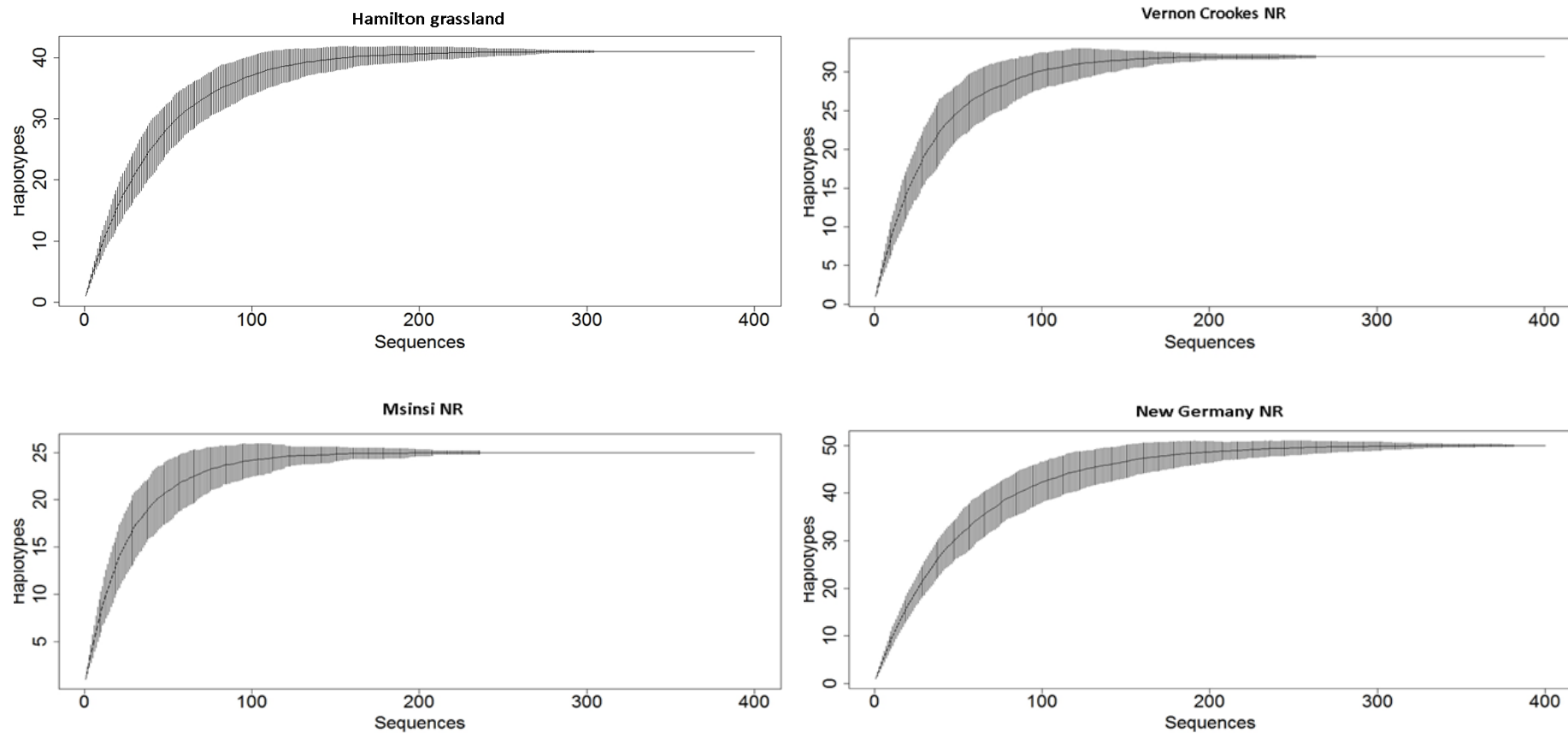


Figure 3.3 Haplotype accumulation curves for Hamilton grassland, Vernon Crookes, Msini and New Germany Nature Reserves. The numbers on the y-axis represent the number of haplotypes observed and on the x-axis is the number of sequences barcoded from each locality. The species accumulation curve is the solid black line, which represents the average of 100 random draws, sampling with replacement, at each level of abundance. The shaded envelope represents a symmetric 95% bootstrap confidence interval, calculated from the estimated variance of the random draws.

The Michaelis-Menten measure estimated that at least 591 species could be encountered in the region as a whole, while Chao 2 measure reported 527 and the Jackknife 1 estimator 450 potential species. The three extrapolation methods predict an increase in species coverage if sampling is continued in future (Figure 3.4).

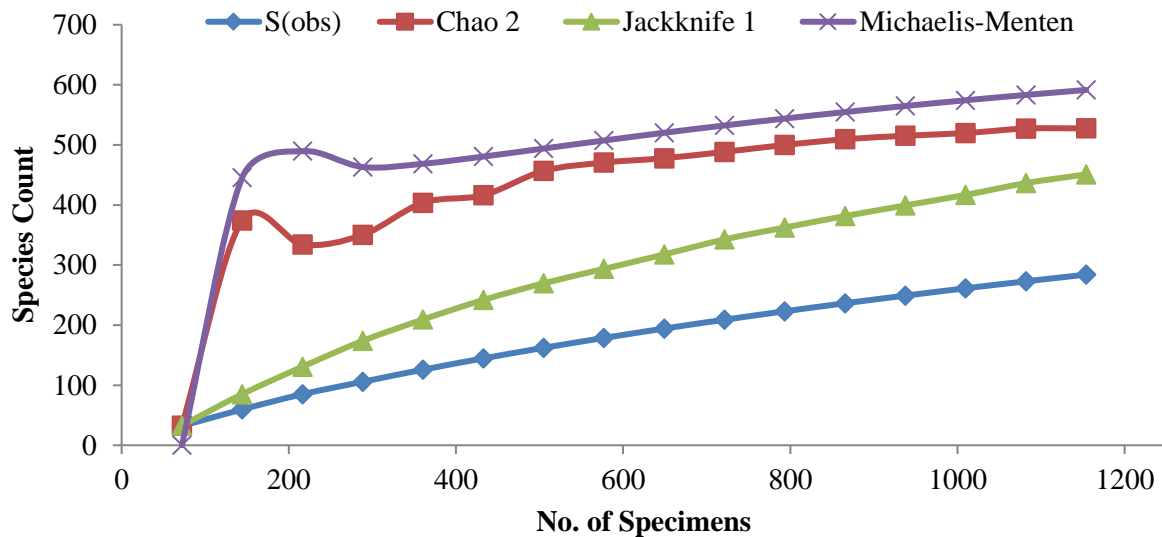


Figure 3.4 Species accumulation curves for 1 153 specimens using three extrapolation methods, where S (obs) is the number of observed species while Chao 2, Jack 1 and MM are estimates of true species richness.

### 3.4.1 Species richness and diversity

The Iphithi NR had the highest number of barcode clusters (77 BINs), indicating the highest species richness in my sample, followed by Springside NR with 72 genetic BINs. The least number of BINs were recorded in Hazelmere dam (3 BINs) and Bartlett estate (4 BINs). Very few specimens were collected from the latter two localities even though they were sampled multiple times. This could mean a true lack of diversity or just that intense sampling is required. While in most cases the number of morphospecies matched the number of BINs present in each locality, in three localities the number of morphospecies exceeded the number of BIN clusters (Palmiet: morphospecies = 67, BINs = 66; Vernon Crookes nature reserve: morphospecies = 27; BINs = 25 and Giba Gorge: morphospecies = 24, BINs = 23).

Only in two cases did the number of BINs exceed the number of morphospecies (Bartlett estate: morphospecies = 3; BINs = 4 and High-Meadows: morphospecies = 18; BINs = 19).

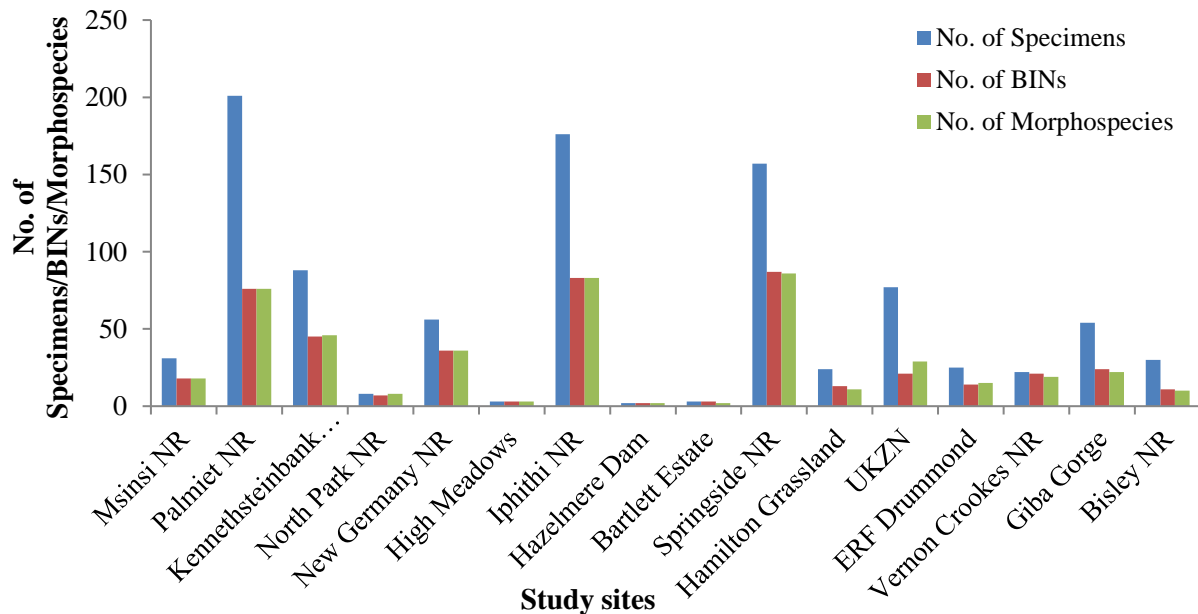


Figure 3.5 Graph showing the number of specimens, BIN clusters and morphospecies per sampling site. The sampling sites are sorted in an increasing geographic distance from the harbour; Msinsi nature reserve is closest to the urban hub and Bisley nature reserve the farthest.

At most localities the distribution of species appeared to be equally matched or balanced, as indicated by low Simpson's index values ( $D < 0.20$ ). The Simpson's index was lowest in Springside NR ( $D = 0.028$ ) followed by Palmiet NR ( $D = 0.037$ ) suggesting most species are present at similar abundances. At Bartlett estate and Hazelmere dam  $D > 0.20$  (0.277 and 0.333, respectively) suggesting one or more species dominated the sample. The  $D$  values observed for both Bartlett estate and Hazelmere dam are probably associated with under sampling at these two localities. The Shannon's diversity index takes into account the number of individuals and number of taxa present in a given area (Hajibabaei et al., 2011), with higher values indicating greater diversity and distribution of species occurring in a sample. This index was highest for Springside NR ( $H' = 3.949$ ) followed by Iphithi ( $H' = 3.799$ ), Palmiet NR ( $H' = 3.69$ ) and Kenneth Stainbank ( $H' = 3.481$ ), meaning these samples were

highly diverse and most species were present in equal abundances while Hazelmere dam had the lowest value ( $H' = 1.099$ ).

Table 3.2 Biodiversity estimation indices where  $S$  is the total number of COI BINs,  $N$  is the total number of individual,  $D$  is Simpson's index,  $H'$  Shannon's index,  $d$  Margalef's index and  $\alpha$  Fisher's  $\alpha$

Sites	$S$	$N$	$D$	$H'$	$d$	$\alpha$
Bartlett Estate	4	6	0,277	1,33	1,674	5,245
Bisley NR	12	21	0,106	2,351	3,613	11,64
Drummond	30	65	0,093	2,888	6,947	21,61
Hamilton Grassland	24	45	0,081	2,853	6,042	20,9
High-Meadows	19	28	0,081	2,748	5,402	25,99
Hazelmere Dam	3	3	0,333	1,099	1,82	0
Giba Gorge	23	51	0,121	2,598	5,595	16,13
Iphithi NR	77	203	0,039	3,799	14,3	45,22
Kenneth Stainbank NR	44	88	0,043	3,481	9,604	35,02
Palmiet NR	66	204	0,037	3,69	12,22	33,85
Springside NR	72	157	0,028	3,949	14,04	51,48
Msinsi NR	18	33	0,079	2,698	4,862	16,21
New Germany NR	35	57	0,047	3,319	8,409	38,58
Vernon Crookes NR	25	39	0,061	3,024	6,551	30,05
UKZN	52	146	0,048	3,458	10,23	28,87
North Park NR	7	8	0,156	1,906	2,885	26,78

The highest  $d$  value was observed for Iphithi NR ( $d = 14.3$ ) followed by Springside NR ( $d = 14.04$ ) then Palmiet NR ( $d = 12.22$ ) while the lowest value was observed for Bartlett estates ( $d = 1.674$ ) and Hazelmere dam ( $d = 1.82$ ). The high  $d$  value indicates presence of many different species in a sample. The highest value of Fisher's  $\alpha$  was seen for Springside NR ( $\alpha = 51.48$ ) followed by Iphithi NR ( $\alpha = 45.22$ ) and the lowest  $\alpha$  value was observed for Hazelmere dam grassland ( $\alpha = 0.00$ ). Taking into account the results of all the calculated diversity indices, Springside and Iphithi nature reserves represent the most diverse localities



in the study, while the least diverse sites were Bartlett estates, High Meadows grassland and Bisley nature reserve.

### 3.4.2 Spatial distribution and species richness

Dendograms to show similarities between localities in terms of species composition based on a resemblance matrix were created from the abundance data for each sampling site (Figure 3.6a) and based on the presence and absence data (Figure 3.6b). Hamilton grassland and Bisley nature reserve were seen to be the most similar in terms of species assemblage (% similarity = 45 %) (Figure 3.6a). Springside NR and Iphithi NR were approximately 40% similar (Figure 3.6a). The sample set collected from Hazelmere dam includes many rare species and the species assemblage collected from this locality is quite unique (Figure 3.6a).

The letters F and G in Figure 3.6b indicate vegetation type, forest and grassland, respectively. A few localities missing these letters could not be identified as forest or grasslands because of mixed vegetation. Vegetation type plays a significant role towards the distribution of species, species assemblages in forests (F) is quite unique to species assemblages in grasslands (G), as a most grassland (G) sites are drawn together in Figure 3.5(b) and so are the forest (F) sites. This finding suggests that most spider species are specific to vegetation type, with only a few species that thrive in both vegetation types. The same result can be seen in Figure 3.7, when looking at the distribution of genera across the study sites. At least 4 genera were found to occur in more than six localities, those were *Lecauge*, *Hypsosinga*, *Oxyopes*, and *Thyene*. Three genera *Lecauge*, *Hypsosinga* and *Oxyopes* are well represented at most localities while the genera *Caerostris*, *Evarcha* and *Festucula* are rare occurring only in one site.

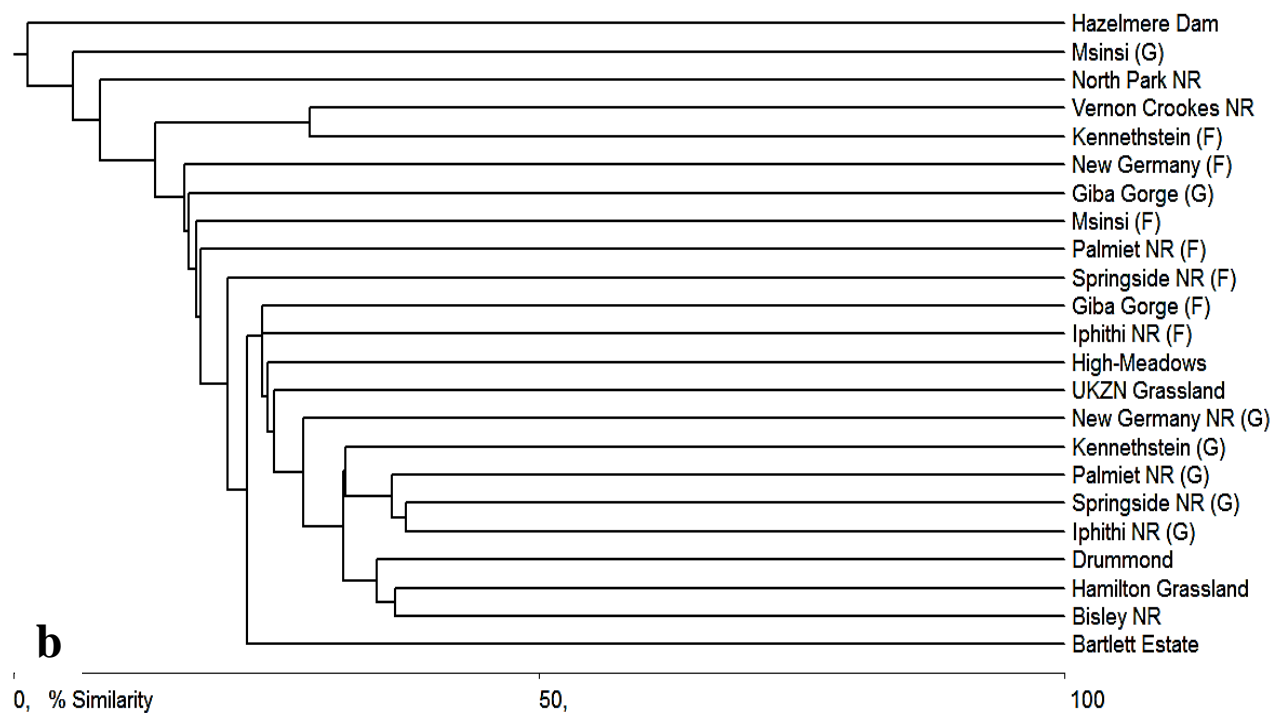
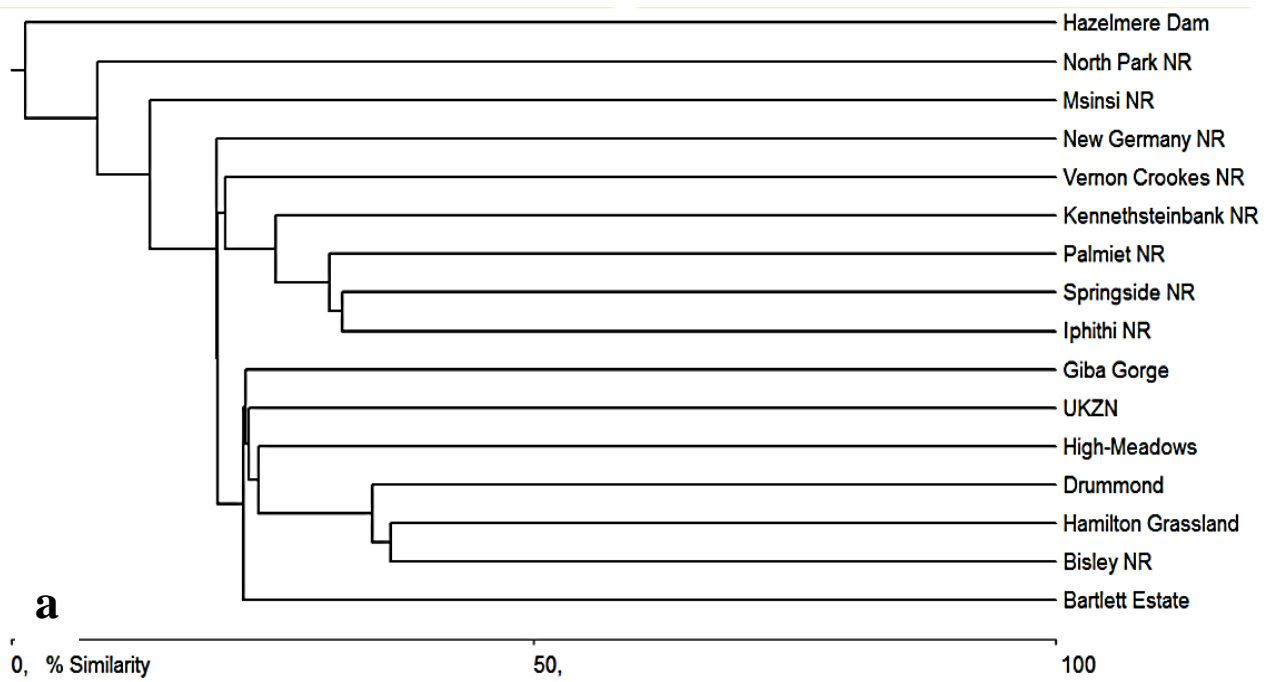


Figure 3.6 Dendograms based on the Bray-Curtis similarity, showing the similarity in spider assemblages between study sites based on (a) abundance data (to compare the distribution of spider species throughout the entire sampling region) and (b) presence/absence data (distribution of species as affected by vegetation type). Sites sharing more species are drawn together.

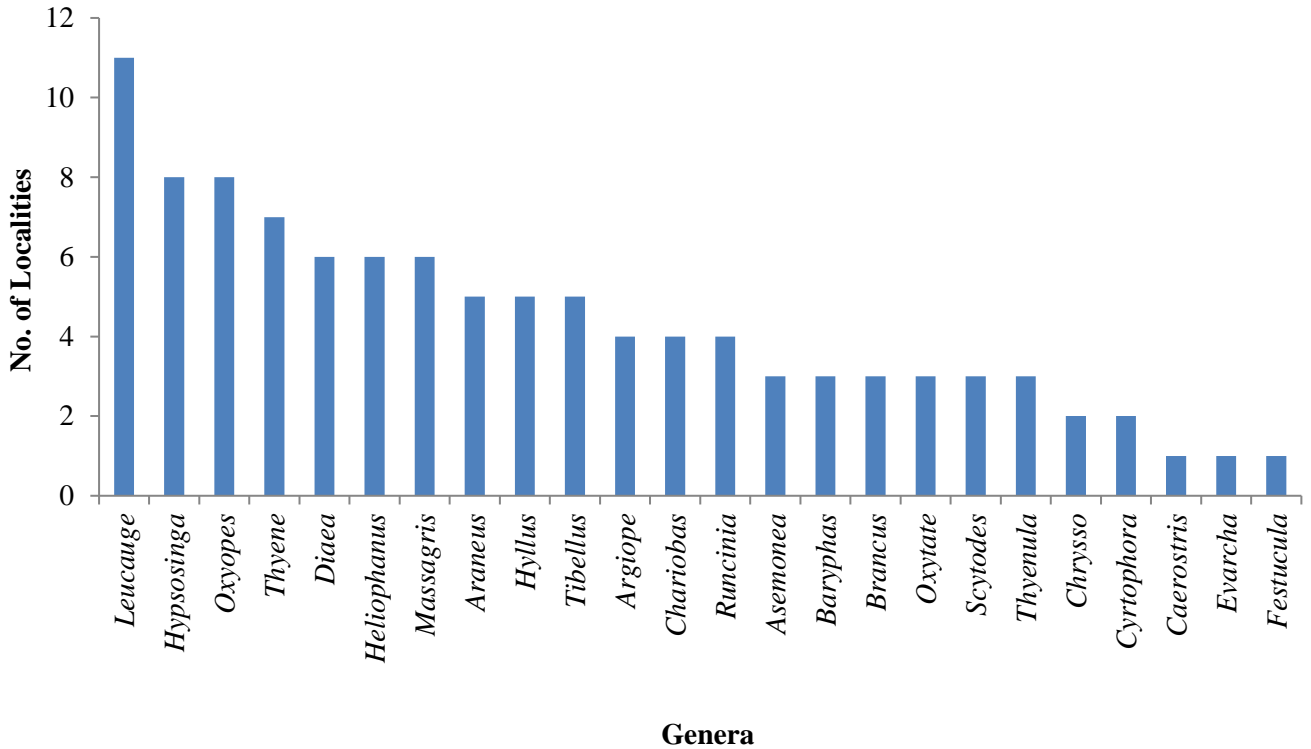


Figure 3.7 Bar graph showing the distribution of genera across the eThekweni study sites. Only three genera, *Leucauge*, *Hypsosinga* and *Oxyopes*, were found in at least 50% of the sampling sites whereas the rest of the genera occurred only in a few sites.

### 3.4.3 The effect of area size and distance from the city on species richness

The strength and direction of the two variables; area size ( $r = -0.105$ ) and distance from harbor ( $r = -0.174$ ), against species richness both showed a weak relationship (Figure 3.8). A positive slope in the regression model was observed for area size against species richness (Figure 3.8a), the coefficient of determination ( $R^2 = 0.01$ ) however was very low and not significant. Species richness decreased with increase in distance from city hub, Figure 3.8b, this result however is not reliable as only 3% of variation in species richness can be explained by distance from the harbor.

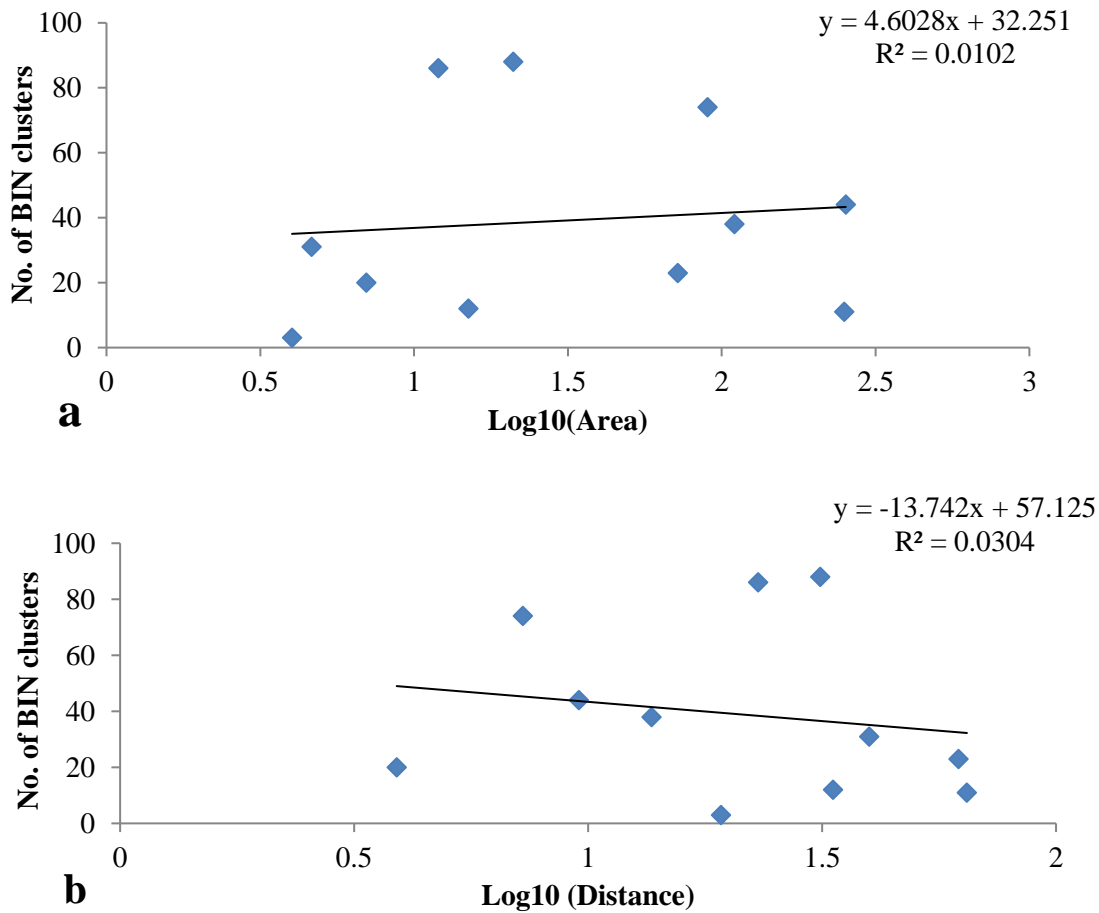


Figure 3.8 Regression analysis of species richness against (a) area size and (b) distance from the city hub.

### 3.5 Discussion

The number of species in a given area is the most basic and fundamental measure of diversity (Moorhead and Philpott, 2013) and the collection of quantitative data for any area of high conservation value is important. Assessing species richness using classic morphology-based methods, however, requires taxonomic knowledge and a substantial investment of time. In this study a complimentary approach for rapid biodiversity assessment using DNA barcodes of spiders was used. This method can never replace traditional taxonomic methods but can provide a useful tool that can be used by municipality managers to identify and prioritize areas of high biodiversity for conservation.

In this study 1 153 specimens from 16 localities were collected and barcoded. Although morphological analysis was not conducted by a taxonomic expert, using published taxonomic keys, 290 morphospecies were identified. Analysis of the molecular data suggests that there are 288 BINs or unique genetic clusters present in the collected material, with an average of 18 BINs collected from each locality. The number of species delimited by the molecular data was more conservative than the morphological estimate in this study, this is in contrast to other studies have shown that sequence-based methods tended to yield greater species richness (Blaxter et al., 2005; Smith et al., 2005). The difference in the number of barcode clusters and morphospecies could have risen due to sexual dimorphism (as such some of the morphospecies may actually belong to the same species) or cryptic speciation (individuals of different species can be identified as one species when they are actually different species) (Barrett and Hebert, 2005a). However, this difference between BINs and morphospecies was not large (Figure 3.5 and Table 3.1) and suggests that DNA barcodes can be used to provide rapid biodiversity assessment for this particular group with the eThekweni region.

Diversity within each site was tested as a measure of haplotype diversity; in this study Iphithi NR, UKZN grassland and Palmiet NR (greater than 80 haplotypes) were the most diverse. Accumulation curves indicate that no sampling site was thoroughly sampled and adequate sampling would be reached at approximately 400 samples, this highlights the difficulty involved in creating comprehensive species inventories. Using this preliminary data, however, species richness was estimated using three extrapolation methods (Figure 3.4). The Michaelis-Menten method predicted 506 species could exist within the eThekweni region, Chao 2 predicted 422 species and Jackknife 1 predicted 389. While these estimators can be subject to sampling method used (Cardoso et al., 2008), they all indicate that more spider diversity is actually present than that is reflected in this study. Cardoso et al. (2008) found the Chao estimators were most reliable in assessing spider species richness.

While species richness is the primary measure of biodiversity it does not provide enough information to make informed good decisions on conservation practices. Ecologists and conservation biologists are more interested on community assemblages rather than ‘simple’ species richness estimates (Bolger et al., 2008; Robinson and Foulds, 1981; Speight et al., 2008). To describe community assemblages, four diversity indices were used. These

descriptors of community assemblages allow for species richness assessment as well as providing information on the evenness of species in the community and are thus useful in comparing two or more sites. Springside and Iphithi nature reserves are not only the most diverse sites, but also the taxa in the two sites are equally present indicated by a fairly low  $D$  values, also supported by the high values of  $H'$  (Table 3.2). Given the number of times these two localities were sampled (4 times) and the differences in locality size (Springside = 21ha; Iphithi = 12ha), even Palmiet nature reserve which was sampled the most (7 times) and had the greater in size (90ha) showed less diversity compared to Springside and Iphithi nature reserves, this suggests these areas are truly rich in diversity. The positive correlation between species richness and open space size however indicates that area size does affect species richness.

Biotic similarity quantifies the extent to which two or more sites are similar in species composition and distribution of relative abundance. The concept is very important at large spatial scales for the designation of biogeographic provinces that harbor distinctive species assemblages with both endemic and shared species (Gotelli and Chao, 2013). But even at smaller spatial scales this analysis can provide valuable information. The Bray-Curtis cluster analysis was implemented to find most similar sites in terms of species composition and abundance. An important finding from this study is that most spider species collected were not widely distributed among all collection sites. This finding is reflected in the low percentage similarity among the sampled sites. The grouping with the highest similarity was Bartlett estate together with Hamilton grassland at approximately 45% (Figure 3.6a). Vegetation type (grassland or forest) remains the largest contributing factor in shaping the community assemblage of spiders (Meyer and Paulay, 2005) and this was confirmed by the eThekwini data (Figure 3.6b). When the presence/abundance data were used, sites that share vegetation type were drawn together. This indicates the presence of spiders that are habitat specialist, however the percentage similarity was less than 50% meaning most of the spiders are generalists or could also do with the fact that the system used to classify vegetation types was not very strict. This however is not surprising given that spiders tend to be habitat specific as a result of different feeding behaviors and different vegetation complexes (Strong and Sherry, 2000; Tews et al., 2004).

Previous studies have highlighted the role played by urbanisation on species distribution including, habitat fragment size and level of habitat degradation (Cardoso et al., 2008; McKinney, 2002; Wilson et al., 2011), Purvis and Hector (2000) found that human land-use negatively affected both species richness and spider abundance. In the current study, the level of urbanisation (distance from main urban hub used as proxy), did not seem to affect species richness, instead, areas closer to the city were more diverse. The same result was obtained by McDonnell and Hahs (2008) and Magura et al. (2008) when they performed rural-urban gradient analysis on ground dwelling invertebrates. However, in Figure 3.6a, two major groupings can be seen despite the fair amount of species shared, localities in the outer KZN (closer to the city) grouped together while localities in the inner part of KZN grouped together. These groupings could mean many factors are involved in shaping species distribution than just vegetation type, factors like climatic conditions i.e. temperature and humidity. Spiders have been reported to exhibit thermal preferences and tolerance to certain temperature (Sepulveda et al., 2014; Sevacherian and Lowrie, 1972).

### **3.6 Conclusion**

In this study the number of morphospecies closely matches the number of BINs estimated from the barcode data and as such is strong evidence that DNA barcoding can be used to provide a surrogate for traditional morphology-based species assignments and can be used to describe species diversity patterns even at a small regional scale. The study set out to test if spiders can be used as bioindicators. The difference in community assemblages and habitat specificity indicates spiders provide a good model as bioindicators in this region. While there were generalist spider genera that occurred in most localities (e.g. *Leucauge*, *Hypsosinga* and *Oxyopes*), most genera occurred only in a few sampling sites which may indicate different environmental conditions i.e. habitat and climatic conditions. Based on the haplotype accumulation curves, the study could benefit from additional sampling, also this would increase our confidence on biodiversity assessment results as it is not clear if the low diversity in most sites is the result of few sample being collected at each site or if the data is a true reflection of limited biodiversity. As such this study highlights the challenges involved in the construction of comprehensive regional species inventories.

## **Chapter Four: Exploring the Utility of Mini-Barcodes in the Identification of Spider Species in South Africa**

### **Abstract**

DNA barcoding has been applied with success across a wide range of eukaryotes. The standard animal barcode marker is 658-bp of the COI gene from the mitochondrial genome. Most DNA barcoding studies use samples that are specifically collected for the study and which contain good quality source of DNA. Archival material may offer a potential source of expertly identified specimens. But extracting good quality DNA and amplifying the entire DNA barcode region in a single PCR reaction from old museum samples poses several challenges. In such cases the application of mini-barcodes may be very useful. The use of smaller section of the COI gene may however reduce the number of characters available for species delimitation and would limit the success of the barcoding approach. In this chapter, the utility of mini-barcodes in identification of South African spider specimens is tested and two methods of designing mini-barcodes are compared; manual and sliding windows analysis. Mini-barcodes of 109 and 218-bp were generated and phylogenetic trees constructed. The resulting trees were compared statistically using K-tree distance and Robinson-Foulds symmetric difference. The number of variable characters in a fragment did not affect species delimitation power of the mini-barcode fragments in any way. Effective mini-barcode fragments were found in the middle part of the COI gene around 271-bp from the 5' end. Mini-barcodes generated from sliding windows analysis were more effective compared to mini-barcodes designed manually. Identification accuracy increased with mini-barcode size or length, 218-bp had a greater species resolving power than 109-bp fragment. Despite the success reported by other authors who explored the use of mini-barcodes, my study shows they can be less effective, however, they may still be used to assign specimens to their higher levels of taxonomy (genus or family).



#### **4.1 Introduction**

DNA barcoding is a molecular tool that can be used to facilitate species identification and discovery through the use of a standardized genomic fragment. In animals, the 650-bp fragment of the mitochondria's COI gene has been routinely used as a species tag (Hajibabaei et al., 2006b). Ideally, DNA barcoding studies use tissue from specimens that were specifically collected for the study or tissues that were recently collected and preserved specifically for DNA-based analysis (Hajibabaei et al., 2005; Meusnier et al., 2008). The use of fresh material means that extraction and amplification of the 650-bp COI fragment is routine and the sequence data is easily obtained using universal primers (Miller et al., 2013). In some cases, however, samples with degraded DNA must be used instead (Boyer et al., 2012). For example, identification of specimens from environmental DNA samples or from samples that have not been adequately preserved (Boyer et al., 2012; Hajibabaei et al., 2006b), and in cases where modern barcode records need to be validated by comparing them to the barcode records from their holotype specimens of a species (Hajibabaei et al., 2006b).

Degraded DNA is particularly problematic in old specimens (as the DNA molecule is unstable) (Meusnier et al., 2008) and in tissues that have low concentrations of DNA (hair, feathers and small fragments of material) (Hatley and MacMahon, 1980; Horvath et al., 2009). Given enough time and exposure to the environment all sources of DNA will degrade (Boyer et al., 2012). While the standard barcode for species identification in animals is approximately 650-bp of the COI mitochondrial gene, studies show that even short fragments of the standardized barcoding gene are sufficiently informative to aid accurate species identification (Bhattacharjee and Ghosh, 2013; Dubey et al., 2011). These short fragments are 'mini-barcodes' (Fan et al., 2009; Hajibabaei et al., 2006b; Meusnier et al., 2008). A number of studies have successfully designed and tested the potential mini-barcodes in identification of specimens from a wide taxonomic range. For example, mini-barcodes have been designed for Indian snakes (Dubey et al., 2011); fruit flies (Fan et al., 2009); moths and wasps (Hajibabaei et al., 2006b); fish, amphibians and flies (Meusnier et al., 2008) but mini-barcodes have yet to be tested on spiders.

While exploring the potential use of mini-barcodes in species identification, Hajibabaei et al. (2006b) demonstrated that the length and position of mini-barcodes plays an important role in their ability to discriminate among species. Meusnier et al. (2008) also demonstrated that the ability of smaller fragments (less than 150-bp) to correctly identify species decreased with a decrease in barcode length and the full-length barcodes always performed best.

The use of mini-barcodes is expected to significantly broaden the application of DNA barcoding in biodiversity studies (Meusnier et al., 2008). Meusnier et al. (2008) demonstrated that sequence information can be reliably obtained from archival specimens or those with degraded DNA and the universality of the primers enables the recovery of comprehensive barcode information from environmental samples. The latter is critical especially in the new emerging field of meta-barcoding, a relatively new tool for biodiversity assessment that involves amplification of barcodes from mixed environmental samples (Deagle et al., 2014). In this chapter, mini-barcodes were designed and tested for potential use in identification of South African spider species. The performance of different fragments was assessed to find the optimal mini-barcode and the two methods for creating mini-barcodes were compared.

## **4.2 Materials and Methods**

For this chapter, a subset of the data from Chapter two, the eThekweni spider reference database, was used. A 654-bp COI alignment of 453 individuals consisting 117 species belonging to 87 genera, excluding singletons (species with only one representative) was used. The final alignment was created using ClustalX 2.1 and optimized manually to ensure homology using BioEdit 7.2.5.

### **4.2.1 Designing of mini-barcodes, manually**

Two sets of mini-barcodes were created. First, the full-length barcode alignment (654-bp) was divided into 6 fragments of approximately 109-bp each. Second, the full-length alignment was divided into 3 fragments of 218-bp length each. Each mini-barcode alignment and the full length alignment was analysed separately. Summary statistics of each alignment including the number of variable and parsimony informative characters were estimated using MEGA version 5.

#### **4.2.1.1 Phylogenetic analysis of mini-barcodes generated manually**

To test the performance of mini-barcodes, phylogenetic trees were created for each alignment. Maximum likelihood trees were constructed using RAxML. In each case the GTR+G+I substitution model was used and the ML search method was used. The maximum likelihood trees for each mini-barcode fragment were compared to the reference tree (maximum likelihood analysis of the full barcode length alignment) statistically using K-tree scores and Robinson-Foulds symmetric differences calculated using the program Ktreedist 1.0 (Soria-Carrasco et al., 2007). K-tree scores compare the topology and branch lengths of trees, two trees with very different relative branch lengths get a high K-score whereas two trees that follow a similar among-lineage rate variation get a low score (Soria-Carrasco et al., 2007), while the Robinson-Foulds metric compares tree topologies, a high score means trees are less similar while least scores indicate trees are more similar (Robinson and Foulds, 1981).

#### **4.2.2 Sliding Windows Analysis on R**

Sliding window analysis was conducted using SPIDER package (Species Identify and Evolution in R). The sliding windows analysis allows for setting of the window length and the interval between starting position of the next window, in this way the best mini-barcode per window length will be picked from a large number of possible barcodes for the particular window length, for example, in a 654-bp alignment; a window of 50-bp and codon intervals will give 604 different mini-barcodes and a window of 100-bp will give 554 mini-barcodes. In this analysis, the 654-bp DNA alignment was broken into windows of two different sizes, 109 and 218-bp. The position of the optimal or best barcode segment is judged by the occurrence of larger K2P mean distances, least proportion of zero non-conspecific and zero cells in K2P distance matrix and a highest proportion of congruence of neighbor-joining trees (Clade Comp + Clade Comp Shallow). The same statistical methods (K-score and Robinson Foulds symmetric difference) were used to compare trees from best mini-barcodes picked by the sliding windows analysis to the full length barcode tree.

### 4.3 Results

#### 4.3.1 Manually designed mini-barcodes

As expected the number of parsimony informative characters decreased with the length of the alignment; the full-length barcode alignment had the highest number of parsimony informative characters while the mini109 category had the lowest (Table 4.1). The three 109-bp mini-barcodes (mini109-3 to mini109-5) had similar number of parsimony informative characters and mini109-1 had the highest number of both variable and parsimony informative characters. The nucleotide composition of all mini barcodes was similar.

Table 4.1 Summary statistics for all the mini-barcode and full-length barcode alignments computed on MEGA5. The names of the mini-barcode alignments indicate their size (length in bp) and their position relative to the 5' end of the mitochondrial COI gene.

dataset	#characters	#variable characters	#parsim.info. characters	Nucleotide Composition			
				T/U	C	A	G
full-length	654	395	371	42,8	12,7	25,5	19,0
mini109-1	109	95	87	37,1	9,5	29,1	24,3
mini109-2	109	53	47	46,5	9,0	27,0	17,5
mini109-3	109	66	64	43,6	14,8	24,3	17,3
mini109-4	109	60	60	41,1	17,2	20,6	21,2
mini109-5	109	68	65	47,1	10,1	28,8	14,0
mini109-6	109	55	49	41,4	15,9	24,9	17,8
mini218-1	218	146	133	41,7	9,2	27,3	21,8
mini218-2	218	126	124	42,3	16,0	22,5	19,2
mini218-3	218	123	114	44,3	13,0	26,9	15,9

Of the 109bp mini-barcode fragments, the lowest K-score value (0.665) was obtained for mini19-4 indicating that the maximum likelihood tree generated by the fragment mini109-4 was the most similar to the tree produced by the full length fragment (Table 4.2). The same result was obtained when tree topologies were compared as the least symmetric difference (R-F = 260) was obtained when mini109-4 was compared to the full length barcode tree. Of the 218bp mini-barcode fragments, mini219-2 was found to represent the optimal mini-barcode fragment with lowest K-score (0.565) and lowest R-F value (215).

Table 4.2 Comparison of maximum likelihood trees generated from mini-barcodes against the full length barcode tree.

Comparison Trees	K-score	Scale factor	Robinson-Foulds (R-F)	No. of Partitions
mini109-1	0.933	0.346	301	903
mini109-2	0.799	1.007	264	903
mini109-3	0.684	0.812	275	903
mini109-4	0.665	0.861	260	903
mini109-5	1.061	0.263	326	903
mini109-6	0.765	1.187	313	903
mini218-1	0.651	0.747	218	903
mini218-2	0.565	0.839	215	903
mini218-3	0.619	0.750	239	903

### 4.3.2 Sliding Windows Analysis

For the 109-bp mini-barcode, the highest mean K2P distance was observed at position 274-bp of the full dataset. Also around this position the least proportion of zero non-conspecific K2P distance together with the proportion of zero cells in K2P distance occurred while the highest proportion of clades shared was observed at around this position (congruence of neighbor-joining trees). The results of the top six mini-barcode fragments for each category, 109 and 218-bp, are shown in Table 4.3. While their values are different, especially in the 109-bp mini-barcodes, they all satisfy the characteristics of a good mini-barcode; high values on means of the distance matrix, the proportion of zero non-conspecific distances are very close to zero.

For the 218-bp mini-barcode, the highest mean K2P was observed at position 277-bp of the full dataset. Around this position five other mini-barcodes were obtained with very similar mean K2P distances, proportion of zero non-conspecific K2P distance, proportion of zero cells in K2P distance together with the proportion of shared clades. The result can be represented in graphic form, Figure 4.1.

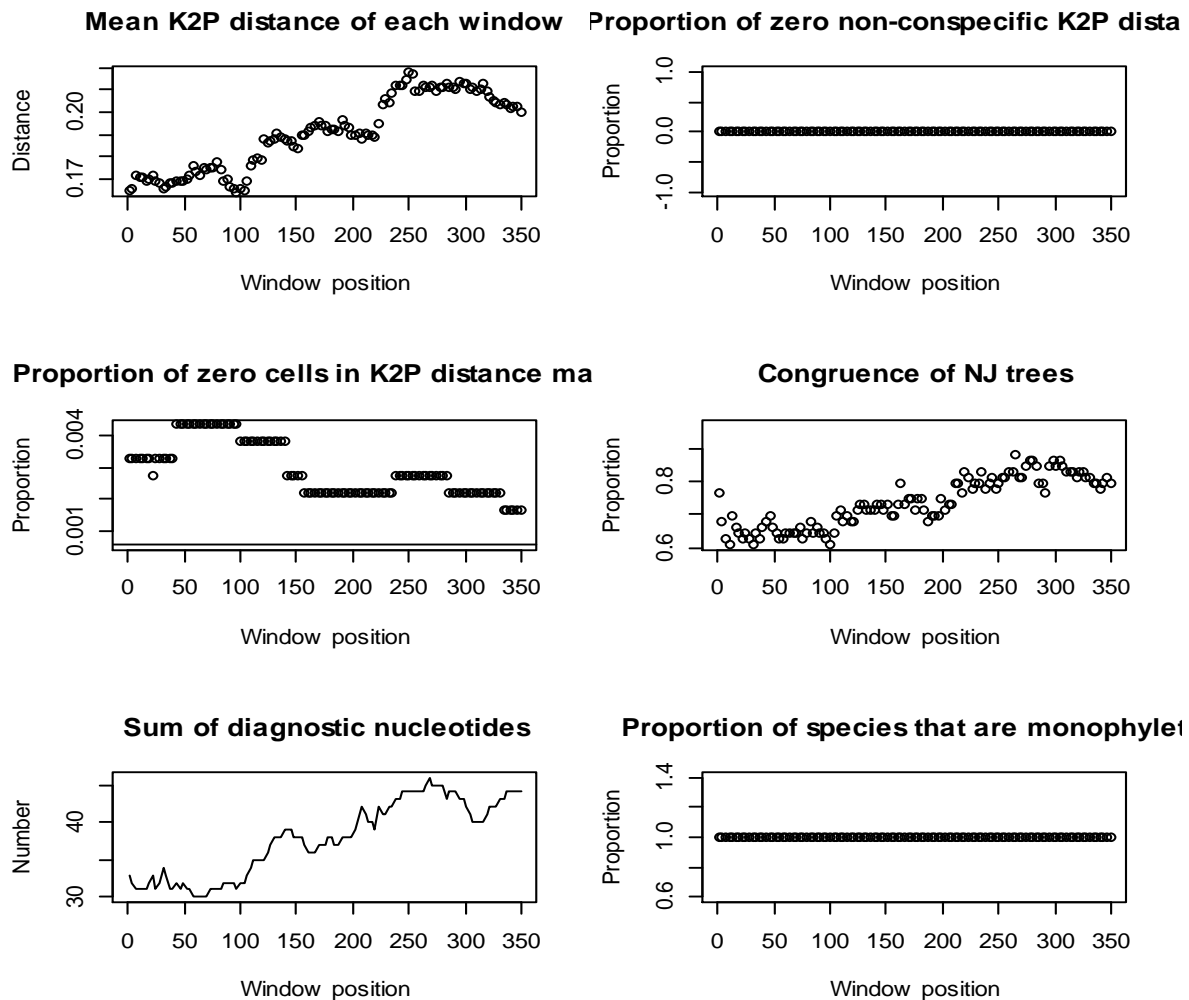


Figure 4.1 Sliding Windows Analysis carried on R environment using SPIDER package; window length = 218-bp; interval = ‘codons’. The top graph, first column shows the mean distances of the sequences. The third graph, column one shows the proportion of summary diagnostic nucleotides, characters informative for species identification. The second graph on the second column shows the proportion of congruent trees between mini-barcodes and the full length dataset.

Table 4.3 The results of the Sliding Windows Analyses summarized, showing potential mini-barcode segments and their positions in the full-length alignment.

Win length (bp)	Position	Mean K2P Dist.	Zero Non-Con Dist.	Zero Dist.	Clade Comp	Clade Comp Shallow
109-1	274	0.273	0.192	0.006	0.651	0.795
109-2	277	0.269	0.192	0.006	0.651	0.795
109-3	271	0.269	0.192	0.006	0.658	0.843
109-4	262	0.267	0.181	0.005	0.669	0.821
109-5	265	0.264	0.183	0.005	0.680	0.832
109-6	268	0.262	0.183	0.006	0.689	0.828
218-1	277	0.227	0.119	0.004	0.729	0.840
218-2	280	0.227	0.119	0.004	0.733	0.855
218-3	346	0.226	0.119	0.004	0.753	0.869
218-4	349	0.225	0.119	0.004	0.753	0.884
218-5	340	0.225	0.121	0.004	0.742	0.873
218-6	274	0.224	0.119	0.004	0.738	0.851

When the trees generated from optimal mini-barcodes (109-1 and 218-1) obtained from the sliding windows analyses were compared to the reference tree there was an improvement in the results, the K-score values decreased from 0.933 to 0.658 (109-1) and from 0.651 to 0.535 (218-1) as well as symmetric differences 109-1 (from 301 to 246) and 218-1 (from 218 to 178). The results are shown in Table 4.4.

Table 4.4 Comparison of maximum likelihood trees generated from mini-barcodes generated from Spider's sliding windows analyses against the full length barcode tree based topology and branch lengths.

Comparison Trees	K-score	Scale factor	Robinson-Foulds (R-F)	No. of Partitions
109-1	0.658	0.603	246	903
218-1	0.535	0.831	178	903

#### **4.4 Discussion**

The phylogenetic information sufficient to discriminate among species is not distributed evenly across the 648-bp COI barcode and as such mini-barcodes can be created and used to identify species. In this study I tested the utility of mini-barcodes in spider species identification. The mini-barcodes all contained different number of parsimony informative characters indicating the difference in resolving power is affected by both the position and size of the mini-barcode (Hajibabaei et al., 2006b). The 5' end of the COI fragment was more variable than the 3' end.

##### **4.4.1 Manually created mini-barcodes**

Mini-barcode fragments which contained more parsimony informative characters were expected to perform better in species identification and be more accurate, instead the mini-barcode fragments with the highest number of parsimony informative characters (mini109-1 and mini219-1) showed less identification success. These results were not expected as mini-barcodes with the highest number of parsimony informative characters should contain characters necessary to distinguish between individuals of different species or same genera. The analysis shows that choosing a mini-barcode based on number of variable characters alone may not be useful.

Trees were compared using K-score and the Robinson-Foulds symmetric difference method. This method only takes into account the topologies of the compared trees (Soria-Carrasco et al., 2007). Using the Robinson-Foulds measure to predict optimal mini-barcode, the tree recovered by ML analysis of mini109-4 shared the highest number of partitions (643 of 903 partitions) with the reference tree (Table 4.2). These results were congruent; both measures indicate mini109-4 is the best mini-barcode in this category. Of the 218bp mini-barcode fragments, mini218-2 was the optimal mini-barcode fragment with both a low K-score (0.565) and highest number of patterns shared between the trees (comparison and reference tree). These results show that the length of the mini-barcode does affect its species resolving power as the K-score of mini109-4 was higher than the K-score of mini218-2, also the number of partitions shared was highest for mini218-2, meaning mini218-2 would be the most effective fragment in species identification. This mini-barcode however may not be



sufficient for species identification may be useful in assigning specimens into their higher rankings, i.e. family or genus level (Appendix 2).

#### **4.4.2 Sliding Windows Analyses**

The sliding windows analysis allows for identification of the shortest, most effective mini-barcodes based on location of fragment along the COI gene (Boyer et al., 2012; Brown et al., 2012). Success of species identification using DNA barcoding is dependent on there being enough genetic divergence or distances between individual species, such that individuals belonging to the same species are expected to have minimal genetic divergence (intraspecific) while individuals of different species have greater genetic divergences (interspecific) (Elias et al., 2007; Hebert et al., 2003; Hebert et al., 2003b; Meier et al., 2006). For this reason most DNA barcoding analyses are based on genetic distance. When the trees generated from the predicted mini-barcodes were compared to the reference tree produced using the full 654-bp alignment, the optimal 109 and 218-bp mini-barcode fragments predicted on SPIDER were more effective than the ones created manually (K-scores of predicted min-barcode fragments were lower than that of manually created, also the number of shared partitions were greater in the predicted mini-barcode). These optimal mini-barcode fragments are situated at approximately 270-bp position in the alignment (highest mean K2P distances, least proportion of zero non-conspecific distances and greater number of summary diagnostic nucleotides and highest proportion in congruence of neighbor joining trees, full length dataset and mini-barcode). The 218-bp fragment performed better than the 109-bp fragment and recovered more similar tree to the reference tree.

#### **4.5 Conclusion**

The results suggest choosing mini-barcodes based solely on the number of parsimony informative characters present is not sufficient and can be misleading, however a character based approach in SPIDER can still be used successfully. Mini-barcode fragments obtained using sliding windows analyses performed the best. In agreement with previous studies (Meusnier et al., 2008), the study shows the position and length or size of the mini-barcode fragment affects the species resolving power, with longer fragments outperforming shorter fragments. While previous studies on mini-barcodes have reported high success rates on

species identification (greater than 90% identification accuracy) our study shows mini-barcodes can be less effective in species identification, however, can still be used to identify specimens to their higher ranking, for example the family or genus.

## Chapter Five: General Discussion

This MSc study set out to investigate the usefulness of the barcoding marker COI in identifying spider species collected from the eThekweni region. In Chapter One, the DNA barcoding method is reviewed and special mention is made of the challenges and current progress in the field.

In Chapter Two, the utility of DNA barcoding in identifying species of spiders was demonstrated. A DNA barcode library was assembled for spider specimens collected from eThekweni and surrounding areas which currently has 1 153 specimens from 288 putative species sampled from 16 geographic localities within the eThekweni region and surrounding areas (UKZN grassland and Bisley nature reserve). The ability of COI to accurately delineate species was evaluated using statistical methods (i.e. Best Close Match, ThreshID, threshold optimization and barcode gap analysis) implemented in the R package SPIDER (Brown et al., 2012). The DNA barcoding method proved effective in species identification indicated by high genetic divergence between species, and the presence of the barcode gap. Analyses also indicate that the 3% threshold obtained in this study was similar to the one obtained by previous authors Barrett and Hebert (2005a).

In Chapter Three I test if the DNA barcode data can be used to assess spider species diversity across the region. The number of morphospecies to the number of BIN clusters per locality was compared. In general, the number of morphospecies corresponded well to the number of barcode clusters. This means the barcode data can be used to accurately assess biodiversity. Species richness and area size were correlated. The difference in degrees of urbanisation as measured by distance from city hub did not seem to affect species richness; however, the distance from city hub appeared to have a significant impact towards shaping community assemblages (could be different climatic conditions, different temperatures and humidity of the inner part and outer part of KwaZulu Natal). The distribution of spider species was influenced by vegetation type (forest or grassland), with species composition more similar in areas of same vegetation. The haplotype accumulation curves and extrapolation measures of diversity suggest that there may be many more species within eThekweni. In order for a

comprehensive reference library to be created, it is recommended that more sampling is required.

In chapter four, mini-barcodes were designed and tested for their usefulness in identification of spider species. Using an alignment of 117 species from 87 genera, mini-barcodes of different lengths (109 and 219-bp) were designed and tested. The mini-barcodes designed manually were less effective compared to the mini-barcode fragments generated through sliding windows analysis. This study suggests that the phylogenetic information necessary to identify species is not evenly distributed across the COI gene, the best performing mini-barcode fragments for South African spiders is situated around the 270-bp position in the alignment. While mini-barcodes were less successful compared to the standard full length barcode, they may be used to assign unknown specimens to family level and even genus level to simplify taxonomy on degraded sample lacking sufficient morphological characters.

In conclusion, this study has shown that DNA barcoding is a useful compliment to traditional morphology-based taxonomy and that DNA barcode data has the potential to provide useful information for biodiversity monitoring.

**Appendix1. A neighbor-joining tree of the 1 153 specimens of spiders collected from the eThekwini region and surrounding areas.**

<https://www.dropbox.com/l/scl/AAB72CZi8Q8U3hfFWI0IRUSPAGN6ylBX-LM>

**Appendix2. A neighbor-joining tree of the optimal 219-bp mini-barcode as predicted by the sliding windows analysis.**

<https://www.dropbox.com/l/scl/AAARu26JOYUECP3CEqxp7oiy6Mo4EDBEGwg>

## 6. References

- Alarukka, D., Kotze, D. J., Matveinen, K., and Niemela, J. (2002). Carabid Beetle and Spider Assemblages along a Forested Urban-Rural Gradient in Southern Finland. *Journal of Insect Conservation* 6, 195-206.
- Alberti, M. (2005). THE EFFECTS OF URBAN PATTERNS ON ECOSYSTEM FUNCTION. *INTERNATIONAL REGIONAL SCIENCE REVIEW* 28, 168-192
- Andersen, A. A., Fisher, A., Hoffmann, B. D., Read, J. L., and Richards, R. (2004a). Use of terrestrial invertebrates for biodiversity monitoring in Australian rangelands, with particular reference to ants. *Austral Ecology* 29, 87-92.
- Andersen, A. N., Fisher, A., Hoffmann, B. D., Read, J. L., and Richards, R. (2004b). Use of terrestrial invertebrates for biodiversity monitoring in Australian rangelands, within particular reference to ants. *Austral Ecology* 29, 87-92.
- Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Veuille, M., and Laredo, C. (2009). DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics* 10, 1-13.
- Barrett, R. D. H., and Hebert, P. D. N. (2005a). Identifying spiders through DNA barcodes. *Can. J. Zool.* 83, 481-491.
- Barrett, R. D. H., and Hebert, P. D. N. (2005b). Identifying spiders through DNA barcoding. *Can. J. Zool.* 83, 481-489.
- 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, 851-869.
- Bhattacharjee, M. J., and Ghosh, S. K. (2013). Design of Mini-barcode for Catfishes for assessment of archival biodiversity. *Molecular Ecology Resources* 14, 469-477.
- Blackwood, C. B., Hudleston, D., Zak, D. R., and Buyer, J. S. (2007). Interpreting Ecological Diversity Indices Applied to Terminal Restriction Fragment Length Polymorphism Data: Insights from Simulated Microbial Communities. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* 73, 5276-5283.
- Blagoev, G. A., Nikolova, N. I., Sobel, C. N., Hebert, P. D., and Adamowicz, S. J. (2013). Spiders (Araneae) of Churchill, Manitoba: DNA barcodes and morphology reveal high species diversity and new Canadian records. *BMC Ecology* 13, 1-17.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., and Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Phil Trans R Soc B* 360, 1935-1943.

- Blaxter, M. L. (2003). Counting angels with DNA. *Nature* 421, 122-124.
- BOLDSYSTEMS (2013). The Barcode of Life Data systems.
- BOLDSYSTEMS (2015). The Barcode of Life Data systems: Databases, Araneae.
- Bolger, D. T., Beard, K. H., Suarez, A. V., and Case, T. J. (2008). Increased abundance of native and non-native spiders with habitat fragmentation. *Diversity and Distributions* 14, 655-665.
- Boyer, S., Brown, S. D. J., Collins, R. A., Cruickshank, R. H., Lefort, M. C., Malumbres-Olarte, J., and Wratten, S. D. (2012). Sliding Windows Analyses for Optimal Selection of Mini-Barcodes, and Application to 454-Pyrosequencing for Specimen Identification from Degraded DNA. *PLoS ONE* 7, e38215.
- Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D., and Rodrigues, A. S. (2006). Global biodiversity conservation priorities. *science* 313, 58-61.
- Brown, S. D. J., and Collins, R. A. (2011). SPIDER: SPecies Identity and Evolution in R. A Tutorial.
- Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M. C., Malumbres-Olarte, J., Vink, C. J., and Cruickshank, R. H. (2012). SPIDER: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12, 562-565.
- Burns, J. M., Janzen, D. H., Hajibabaei, M., Hallwachs, W., and Hebert, P. D. N. (2007). DNA BARCODES OF CLOSELY RELATED (BUT MORPHOLOGICALLY AND ECOLOGICALLY DISTINCT) SPECIES OF SKIPPER BUTTERFLIES (HESPERIIDAE) CAN DIFFER BY ONLY ONE TO THREE NUCLEOTIDES. *Journal of the Lepidopterists' Society* 61, 138-153.
- Candek, K., and Kuntner, M. (2015). DNA barcoding gap: reliable species identification over morphological and geographic scales. *Molecular Ecology* 15, 268-277.
- Cardoso, P., Gaspar, C., Pereira, L. C., Silva, I., Henriques, S. S., da Silva, R. R., and Sousa, P. (2008). Assessing spider species richness and composition in Mediterranean cork oak forests. *ACTA OECOLOGICA* 33, 114-127.
- CBOL-PWG (2009). A DNA Barcode for land plants. *Proceedings of the National Academy of Sciences* 106, 12794-12797.
- Chapple, D. G., and Ritchie, P. A. (2013). A Retrospective Approach to Testing the DNA Barcoding Method. *PLoS ONE* 8, e77882.
- Coddington, J. A., and Levi, H. W. (1991). Systematics and evolution of spiders (Araneae). *Annu. Rev. Ecol. Syst.* 22.

- Collins, R. A., and Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* 13, 969-975.
- Colwell, R. K. (2013). *EstimateS*: Statistical estimation of species richness and shared species from samples. Version 9.
- Costion, C., Ford, A., Cross, H., Crayn, D., Harrington, M., and Lowe, A. (2011). Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS ONE* 6, e26841.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). "jModelTest 2: more models, new heuristics and parallel computing". *Nature Methods* 9, 772.
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., and Taberlet, P. (2014). DNA metabarcoding and cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* 10.
- Death, R. G. (2002). Predicting invertebrate diversity from disturbance regimes in forest streams. *Oikos* 97, 18-30.
- Dentinger, B. T. M., Didukh, M. Y., and Moncalvo, J. M. (2011). Comparing COI and ITS as DNA barcode markers for mushrooms and allies (Agaricomycotina). *PLoS One* 6, e25081.
- DeSalle, R., Egan, M. G., and Siddall, M. (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1905-1916.
- DeWalt, R. E., Cao, Y., Tweddale, T., Grubbs, S. A., Hinz, L., Pessino, M., and Robinson, J. L. (2012). Ohio USA stoneflies (Insecta, Pecopectera): species richness estimation, distribution of functional niche traits, drainage affiliations, and relationships to other states. *ZooKeys* 178, 1-26.
- Dippenaar-Schoeman, A. S., Foord, S. H., and Haddad, C. R. (2013). "Spiders of the Savanna Biome," University of Venda, Thohoyandou, and the Agricultural Research Council, Pretoria, South Africa.
- Dippenaar-Schoeman, A. S., Haddad, C. R., Foord, S. H., Lyle, R., Lotz, L. N., and Marais, P. (2015). South African National Survey of Arachnida (SANSA): review of current knowledge, constraints and future needs for documenting spider diversity (Arachnida: Araneae). *Transactions of the Royal Society of South Africa* 0, 1-31.
- Dippenaar-Schoeman, A. S., and Jocque, R. (1997). "African Spiders: An Identification Manual," Agricultural Research Council, South Africa.
- Dubey, B., Meganathan, P. R., and Haque, I. (2011). DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species. *Forensic Sci Int Genet* 5, 181-184.



Elias, M., Hill, R. I., Willmott, K. R., Dasmahapatra, K. K., Brower, A., Mallet, J., and Jiggins, C. D. (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proc. R. Soc. B* 274, 2881-2889.

Fan, J., Gu, H., Chen, S., Mo, B., Wen, Y., He, W., Liu, W., and Zeng, X. (2009). Species Identification of 36 Kinds of Fruit Flies Based on Minimalist-barcode

*Chin. J. Appl. Environ. Biol.* 15, 215-219.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. *Department of Genome Science, University of Washington, Seattle.*

Ferrier, S., Powell, G. V. N., Richardson, K. S., Manion, G., Overton, J. M., Allnutt, T. F., Cameron, S. E., Mantle, K., Burgess, N. D., Faith, D. P., Lamoreux, J. F., Kier, G., Hijmans, R. J., Funk, V. A., Cassis, G. A., Fisher, B. L., Flemons, P., Lees, D., Lovett, J. C., and Van Rompaey, R. S. A. R. (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *BioScience* 54, 1101-1109.

Filmer, M. R. (1999). "Southern African Spiders: An Identification Guide," Struik Publishers, South Africa, Cape Town, 8001.

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. *Mol. Mar. Biol. Biotechnol.* 3, 249-299.

Foord, S. H., Dippenaar-Schoeman, A. S., and Haddad, C. R. (2011). South African Spider Diversity: African Perspectives on the Conservation of a Mega-Diverse Group, Changing Diversity in Changing Environment

*PhD. Oscar Grillo (Ed.).*

Gaikwad, S., Warudkar, A., and Shouche, Y. (2016). Efficacy of DNA barcoding for the species identification of spiders from Western Ghats of India. *Mitochondrial DNA Part A*, 1-7.

Goodall-Copestake, W. P., Tarling, G. A., and Murphy, E. J. (2012). On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene *cox1* in animals. *Heredity* 109, 50-56.

Gordon, A., Simondson, D., White, M., Moilanen, A., and Bekessy, S. A. (2009). Intergrating conservation planning and landuse planning in urban landscapes. *Landscape and Urban Planning* 91, 183-194.

Gotelli, N. J., and Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data.

Govender, N. (2014). DURBAN: STATE OF BIODIVERSITY REPORT 2013/2014.

Gunnarsson, B. (1990). Vegetation structure and the abundance size distribution of spruce-living spiders. *J. Anim. Ecol.* 59, 743-752.

Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., Hebert, P. D., and Janzen, D. H. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1835-1845.

Hajibabaei, M., Janzen, D. H., Hallwachs, W., and Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* 103, 968-971.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., and Baird, D. J. (2011). Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLoS ONE* 6.

Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N., and Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics* 23.

Hajibabaei, M., Smith, A. M., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., and Hebert, P. D. N. (2006b). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6, 959-964.

Hatley, C. L., and MacMahon, J. A. (1980). Spider community organisation: Seasonal variation and the role of vegetation architecture. *Environmental Entomology* 9, 632-639.

Hazarika, L. P. (2013). Diversity indices of macro invertebrates in the Satajan Wetland of Lakhimpur District, Assam. *Annals of Biological Research* 4, 68-72.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270, 313-321.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proc. R Soc Lond B* 270, 313-322.

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*. *Proc. Natl. Acad. Sci. USA* 101.

Hebert, P. D. N., Ratnasingham, S., and DeWaard, J. R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R Soc Lond B* 27, 1-4.

Hebert, P. D. N., Stoeckle, M., Zemaik, T. S., and Francis, C. M. (2004b). Identification of Birds through DNA barcodes. *PLoS Biology* 2, 1657-1663.

- Hebert, P. D. N., Stoeckle, M. Y., Zemplak, T. S., and Francis, C. M. (2004b). Identification of Birds through DNA Barcodes. *PLoS Biology* 2.
- Hodkinson, I. D., Coulson, S. J., Harrison, J., and Webb, N. R. (2001). What a wonderful web they weave: spiders, nutrient capture and early ecosystem development in the high Arctic - some counter-intuitive ideas on community assembly. *Oikos* 95, 349-352.
- Hofreiter, M. (2012). Nondestructive DNA extraction from Museum Specimens. *Methods in Molecular Biology* 840, 93-100.
- Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLoS ONE* 6, e19254.
- Horvath, R., Magura, T., Szinetar, C., and Tothmeresz, B. (2009). Spiders are not less diverse in small and isolated grasslands, but less diverse in overgrazed grasslands: A field study (East Hungary, Nyiriseg). *Agriculture, Ecosystems and Environment* 130, 16-22.
- Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., BurrIDGE, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., and Bernatchez, L. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* 3.
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.
- Human, K. G., and Gordon, D. M. (1997). Effects of Argentine Ants on Invertebrate Biodiversity in Northern California. *Conservation Biology* 11, 1242-1248.
- Ibrahim, A., Mohammad, A., and Khan, H. A. (2012). Inferring the Phylogeny of Bovidae Using Mitochondrial DNA Sequences: Resolving Power of Individual Genes Relative to Complete Genomes. *Evolutionary Bioinformatics* 8, 139-150.
- Jinbo, U., Kato, T., and Ito, M. (2011). Current progress in DNA barcoding and future implications for entomology. *Entomological Science* 14, 107-124.
- Jones, E. L., and Leather, S. R. (2012). Invertebrates in Urban areas: A review. *Eur. J. Entomol.* 109, 463-478.
- Koch, H. (2010). Combining morphology and DNA barcoding resolves the taxonomy of Western Malagasy *Liotrigona* Moure, 1961 (Hymenoptera: Apidae: Meliponini). *African Invertebrates* 51, 413-421.
- Kremen, C., Colwell, R. K., Erwin, T. L., Murphy, D. D., Noss, R. F., and Sanjayan, M. A. (1993). Terrestrial Arthropod Assemblages: Their Use in Conservation Planning. *Conservation Biology* 7, 796-808.
- Kress, W. J., Garcia-Robledo, G., Uriarte, M., and Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution* 30, 25-35.

- Krishnamurthy, P. K., and Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodivers. Conserv.* 21, 1901-1919.
- Kvist, S. (2013). Barcoding in the dark?; A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution* 69, 39-45.
- Iahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Dutoit, S., Barraclough, T. G., and Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *PNAS* 105, 2923-2928.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Leroy, A., and Leroy, J. (2003). "Spiders Of Southern Africa," Struik Publishers, South Africa, Cape Town, 8001.
- Little, D. P., and Stevenson, D. W. (2007). A comparison of algorithms for identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 23, 1-21.
- Lowe, E. C., Wilder, S. M., and Hochuli, D. F. (2014). Urbanisation at Multiple Scales is Associated with Larger Size and Higher Fecundity of an Orb-Weaving Spider. *PLoS ONE* 9.
- Lowenstein, J. H., Burger, J., Jeitner, C. W., Amato, G., Kolokotronis, S. O., and Gotchfeld, M. (2010). DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. *Biology Letters* 6, 692-695.
- Maelfait, J. P., and Hendickx, F. (1998). Spiders as bio-indicators of anthropogenic stress in natural and semi-natural habitats in Flanders (Belgium): some recent developments. *Proceedings of the 17th European Colloquium of Arachnology, Edinburgh 1997*.
- Magura, T., Horvath, R., and Tothmeresz, B. (2010). Effects of Urbanization on ground-dwelling spiders in forest patches, in Hungary. *Landscape Ecology* 25, 621-629.
- Magura, T., Tothmeresz, B., Hornung, E., and Horvath, R. (2008). Urbanisation and Ground-Dwelling Invertebrates. In "Urbanization: 21st Century Issues and Challenges" (L. N. Wagner, ed.). Nova Science Publishers, Inc.
- McDonnell, M., and Hahs, A. (2008). The use of gradient analysis studies in advancing our understanding of the ecology of urbanizing landscapes: current status and future directions. *Landscape Ecology* 23, 1143-1155.
- McGeoch, M. A., Sithole, H., Samways, M. J., Simaika, J. P., Pryke, J. S., Picker, M., Uys, C., Armstrong, A. J., Dippenaar-Schoeman, A. S., Engelbrecht, I. A., Brascler, B., and Hamer, M. (2011). Conservation and monitoring of invertebrates in terrestrial protected areas. *Koedoe* 53, 13.

- McIntyre, N. E. (2000). Ecology of Urban Arthropods: a Review and a Call to Action. *Annals of the Entomological Society of America* 93, 825-835.
- McKinney, M. L. (2002). Urbanization, Biodiversity, and Conservation. *BioScience* 52, 883-890.
- McKinney, M. L. (2008). Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosyst* 11, 161-176.
- 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. *Syst. Biol* 55, 715-728.
- 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.
- Meyer, C. P., and Paulay, G. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol* 3.
- Miller, J. A., Beentjies, K. K., van Helsdingen, P., and IJland, S. (2013). Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. *ZooKeys* 365, 245-261.
- Milton, M., Pierossi, P., and Ratnasingham, S. (2013). Barcode of Life Data Systems Handbook. Guelph, Ontario, Canada.
- Moorhead, L. C., and Philpott, S. M. (2013). Richness and composition of spiders in urban green spaces in Toledo, Ohio. *The Journal of Arachnology* 41, 356-363.
- Moritz, C., and Cicero, C. (2004). DNA Barcoding: Promise and Pitfalls. *PLoS Biol* 2.
- Muelelwa, M. I., Foord, S. H., and Dippenaar-Schoeman, A. S. (2010). Towards a Standardized and Optimized Protocol for Rapid Biodiversity Assessments: Spider Species Richness and Assemblage Composition in Two Savanna Vegetation Types. *African Zoology* 45, 273-290.
- Niemela, J., Kotze, D. J., Venn, S., Penev, L., Stoyanov, I., Spence, J., Hartley, D., and Montes de Oca, E. (2002). Carabid beetle assemblages (Coleoptera, Carabidae) across urban-rural gradients: an international comparison. *Landscape Ecology* 17, 387-401.
- Otter, R. R., Hayden, M., Mathews, T., Fortner, A., and Bailey, F. C. (2012). The use of tetragnathid spiders as bioindicators of metal exposure at a coal ASH spill site. *Environmental Toxic and Chemistry* 32, 2065-2068.
- Palmer, M. A., Lettenmaier, D. P., Poff, N. L., Postel, S. L., Richter, B., and Warner, R. (2009). Climate change and river ecosystems: protection and adaptation options. *Environmental Management* 44, 1053-1068.

- Paoletti, M. G. (1999). Using bioindicators based on biodiversity to assess landscape sustainability. *Agriculture, Ecosystems and Environment* 74, 1-18.
- Paquin, P., and Hedin, M. (2004). The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Circurina* (Araneae: Dictynidae) from Texas caves. *Molecular Ecology* 13, 3239-3255.
- Pearce, J. L., and Venier, L. A. (2006). The use of ground beetles (Coleoptera: Carabidae) and spiders (Araneae) as bioindicators of sustainable forest management: A review. *Ecological Indicators* 6, 780-793.
- Purvis, A., and Hector, A. (2000). Getting the measure of biodiversity. *Nature* 405, 212-219.
- Rach, J., DeSalle, R., Sarkar, I. N., Schierwater, B., and Hadrys, H. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. R. Soc. B* 275, 237-247.
- Rambaut, A. (2009). FigTree v.1.3.1 Computer program and documentation distributed by the author at <http://tree.bio.ed.ac.uk/software/>.
- Rambaut, A., and Drummond, A. J. (2007). Tracer v. 1.5 Computer program and documentation distributed by the authors at <http://beast.bio.ed.ac.uk/Tracer>.
- Ratnasingham, S., and Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data Systems ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes*.
- 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.
- Robideau, G. P., De Cock, A. W. A. M., Coffey, M. D., Voglamayr, H., Brouwer, H., Bala, K., Chitty, D. W., Desaulniers, N., Eggertson, Q. A., Gachon, C. M. M., Hu, C., Kupper, F. C., Rintoul, T. L., Sarhan, E., Verstappen, E. C. P., Zhang, Y., Bonants, P. J. M., Ristaino, J. B., and Levesque, C. A. (2011). DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Mol Ecol Resour* 11, 1002-1011.
- Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees *Mathematical Biosciences* 53, 131-147.
- Robinson, E. A., Blagoev, G. A., Hebert, P. D. N., and Adamowicz, S. J. (2009). Prospects for using DNA barcoding to identify spiders in species-rich genera. *ZooKeys* 16, 27-46.
- Rubinoff, D. (2006). Utility of Mitochondrial DNA barcodes in Species Conservation. *Conservation Biology* 20, 1026-1033.
- Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., and P.D.N., H. (2015). DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Molecular Ecology Resources*

- Schmitz, O. J., Hawlena, D., and Trussell, G. C. (2010). Predator control of ecosystem nutrient dynamics. *Ecology Letters* 13, 1199-1209.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., and Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *PNAS* 109, 6241-6246.
- Sepulveda, R., Taucare-Rios, A., Veloso, C., and Canals, M. (2014). Thermal preference of *Dysdera crocata* C.L. Koch 1838 (Araneae: Dysderidae). *The Journal of Arachnology* 42, 000-000.
- Sevacherian, V., and Lowrie, D. C. (1972). Preferred Temperatures of Two Species of Lycosid Spiders *Pardosa sierra* and *P. ramulosa*. *Annals of the Entomological Society of America* 65.
- Shochat, E., Lerman, S. B., Anderies, J. M., Warren, P. S., Faeth, S. H., and Nilon, C. H. (2010). Invasion, Competition, and Biodiversity Loss in Urban Ecosystems *BioScience* 60.
- Shochat, E., Stefanov, W. L., Whitehouse, M. E. A., and Faeth, S. H. (2004). Urbanization and spider diversity: influences of human modification of habitat structure and productivity. *Ecological Applications* 14, 268-280.
- Smith, A. M., Fisher, B. L., and Hebert, P. D. N. (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Phil. Trans. R. Soc. B* 360, 1825-1834.
- Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Trina, J., Fisher, B. L., Gibbs, J., Hajibabaei, M., Hallwachs, W., Hind, K., Hreck, J., Huang, D. W., Janda, M., Janzen, D. H., Li, Y., Miller, S. E., Packer, L., Quicke, D., Ratnasingham, S., Rodriguez, J. J., Rougerie, R., Shaw, M. R., Sheffield, C., Stahlhut, J. K., Steinke, D., Whitfield, J., Wood, M., and Zhou, X. (2012). *Wolbachia* and DNA barcoding insects: patterns, potential, and problems. *PLoS ONE* 7.
- Sonet, G., Jordaens, K., Nagy, Z. T., Breman, F. C., De Meyer, M., Backeljau, T., and Virgilio, M. (2013). *Adhoc*: an R package to calculate *ad hoc* distance thresholds for DNA barcoding identification. *ZooKeys* 365, 329-336.
- Soria-Carrasco, V., Talavera, G., Igea, J., and Castresana, J. (2007). The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics* 23, 2954-2956.
- Speight, M. R., Hunter, M. D., and Watt, A. D. (2008). "Ecology of Insects: Concepts and Applications," 2/Ed. Wiley-Blackwell, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK.
- Stoeckle, M. (2003). Taxonomy, DNA, and the Bar Code of Life. *BioScience* 53, 2.

- Strong, A. M., and Sherry, T. W. (2000). Habitat-specific effects of food abundance on the condition of ovenbirds wintering in Jamaica. *Journal of Animal Ecology* 69, 883-895.
- Tamura, K., Peterson, D., Peterson, N., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Method. *Molecular Biology and Evolution* 28, 2731-2739.
- Tavares, E. S., and Baker, A. J. (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evolutionary Biology* 8.
- Taylor, R. J., and Doran, N. (2001). Use of terrestrial invertebrates as indicators of the ecological sustainability of forest management under the Montreal Process. *Journal of Insect Conservation* 5, 221-231.
- Tews, J., Brose, U., Grimm, V. k., Tielbörger, K., Whichmann, M. C., Schager, M., and Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography* 31, 79-92.
- Trewick, S. A. (2007). DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* 24, 240-254.
- Valentini, A., Pomanon, F., and Taberlet, P. (2008). DNA barcoding for ecologists. *Trends in Ecology and Evolution* 24.
- Virgilio, M., Beckeljau, T., Nevado, B., and De Meyer, M. (2010). Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* 11.
- Ward, R. D., Zemplak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. Lond.* 360, 1847-1857.
- Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays* 29, 188-197.
- 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). *Proc. R. Soc. B* 274, 1731-1739.
- WHO (2014). Urban population growth situation. Vol. 2015.
- Wiemers, M., and Fiedler, K. (2007). Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* 4.
- Wilson, E. O. (1988). The current state of biological diversity. *Biodiversity* 521, 3-18.
- Wilson, J. J., Rougerie, R., Schonfeld, J., Janzen, D. H., Hallwachs, W., Hajibabaei, M., Kitching, I. J., Haxaire, J., and Hebert, P. D. N. (2011). When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology* 11.



- Witt, J. S. D., Threlloff, D. L., and Hebert, P. D. N. (2006). DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology* 15, 3073-3082.
- Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H., and Garey, J. R. (2011). Molecular study of worldwide distribution and diversity of soil animals. *PNAS* 108, 17720-17725.
- 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.
- York, A. (1999). "Ecologically sustainable management: the utility of habitat surrogates for assessing terrestrial invertebrate diversity in temperate forests. In: *The Other 99%*," Transactions of the Royal Society of New South Wales, Mosman.
- Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., and Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3, 613-623.
- Zhang, A.-B., Muster, C., Liang, H.-B., Zhu, C.-D., Crozier, R., Wan, P., and Feng, J. (2012). A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular Ecology* 21, 1848-1863.