Investigating floral choice in bees (Megachilidae) using pollen metabarcoding

by

ANNEMARIE GOUS

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in Conservation Genetics

School of Life Science

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

January 2017

PREFACE

The research contained in this thesis 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 Campus, South Africa. The research was financially supported by the National Research Foundation (grant 95649) and the Belgian Directorate-General for Development Cooperation, partim GTI. The candidate was also funded by the Department of Science and Technology and National Research Foundation Professional Development Programme (grant 99781), the Biotechnology Platform at the Agricultural Research Council (ARC), as well as the Professional Development Programme of the ARC. Her postgraduate fees were remitted by the University of KwaZulu-Natal.

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. Sandi Willows-Munro

Date:

_____

Signed: Dr. Connal Eardley

Date:

_____

Signed: Dr. Dirk Swanevelder

Date:

DECLARATION 1: PLAGIARISM

I, Annemarie Gous, 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: Annemarie Gous

Date:

DECLARATION 2: PUBLICATIONS AND PRESENTATIONS

My role in each paper and presentation is indicated. Papers that are still in progress, and have a status other than accepted or published are indicated.

Chapter 2

1. Gous A, Willows-Munro S, Eardley CD, Swanevelder DZH. Pollination: Impact, role-players, interactions and study – a South African perspective. South African Journal of Science. [Submitted 2016, revisions submitted for approval January 2017]

This paper is a review on various aspects of pollination and plant-pollinator interactions, and focusses specifically on the interactions of bees with their plant hosts. I conceived the layout, conducted the literature review and wrote the paper.

Chapter 3

2. Gous A, Swanevelder ZH, Willows-Munro S and Eardley C. 2014. Using molecular barcodes to investigate pollinator-plant interactions in long-tongued bees. 7th International Symposium on Molecular Insect Science, Amsterdam, The Netherlands held from 13 – 16 July 2014. Poster presented by A Gous.

3. Gous A, Swanevelder ZH, Eardley C and Willows-Munro S. 2014. Investigating floral choice in long-tongued bees (Megachilidae) from a national collection using molecular barcodes and next generation sequencing. 2nd Joint Congress of the South African Society for Bioinformatics (SASBi) and the South African Genetics Society (SAGS), Pretoria, South Africa held from 23 – 26 September 2014. Poster presented by A Gous.

The aforementioned work was focused on the early stages of pollen metabarcoding in our laboratory and showcased the bioinformatic troubleshooting performed in order to obtain taxonomic identification of pollen samples. I conducted the laboratory work, design and compiled the posters and presented them at the conferences.

4. Gous A, de Bruin JJ, Willows-Munro S, Eardley C and Swanevelder ZH. 2015. The application of NGS barcoding in identifying mixed pollen samples from a historic bee collection. 6th International Barcode of Life Conference, Guelph, Canada held from 18 – 21 August 2015. Oral presented by A Gous.

5. Gous A, de Bruin JJ, Willows-Munro S, Eardley C and Swanevelder ZH. 2015. The application of NGS barcoding in identifying mixed pollen samples from a historic bee collection. Agricultural Research Council (ARC) Professional Development Programme (PDP) Conference, Pretoria, South Africa held 19 – 21 October 2015. Oral presented by A Gous.

6. Gous A, de Bruin JJ, Willows-Munro S, Eardley C, Swanevelder D. 2015. The application of next-generation sequencing barcoding in identifying mixed-pollen samples from a historic bee collection. Genome 58, 222. [Abstract published]

The oral presentation at the 6th International Barcode of Life Conference was accepted in the session on pollen barcoding. I obtained a travel grant from the conference organisers and the International Development Research Centre (IDRC) for Excellence in Socio-Economic Applications of DNA Barcoding. I discussed the metabarcoding of pollen from 22 *Megachile venusta* specimens from the National Insect Collection housed at the Agricultural Research Council (ARC), Pretoria. During this session, the Pollen-Barcode of Life (Pollen-BoL) initiative was formed and all speakers became inaugaral members. I was requested to give the same presentation at the ARC's Professional Development Programme Conference. The abstract submitted for acceptance to the 6th International Barcode of Life (iBoL) Conference was published in the journal Genome. I collected all data, performed analyses, and wrote the abstract.

7. Gous A, de Bruin JJ, Eardley C, Willows-Munro S, Swanevelder D (2015) Metabarcoding of pollen from a historic bee collection. Barcode Bulletin 6(4), 12-13. [Published]

A newsletter article was published in the iBoL Newsletter, the Barcode Bulletin. I wrote the article and performed the editing after review.

8. Gous A, Willows-Munro S, Eardley CD, Swanevelder DZH. 2016. Pollen Metabarcoding from Bees in a Historic Collection Identifies Provenance and Possible Microbial

Associations. PLoS ONE. [Submitted 2016, revisions submitted for approval October 2016]

This paper is based on the data collected and analysed for Chapter 3. I designed the experiment, collected pollen samples from the ARC's bee collection, performed the laboratory work, analysed the data, wrote the paper, and performed editing post-review.

Chapter 4

9. Gous A, Hefer C, Willows-Munro S, Eardley C and Swanevelder ZH. 2016. Investigating bee-plant interactions in South Africa by metabarcoding the National Insect Collection. 1st Plant and Pollen Metabarcoding Workshop and Pollen-BoL Mini-Symposium, Würzburg, Germany held 9 – 13 May 2016. Oral presented by A Gous.

I was invited to attend the Plant and Pollen Metabarcoding Workshop during the 6th International Barcode of Life Conference. I obtained a scholarship from the German Academic Exchange Service (DAAD) to attend this workshop. The data presented in at the Pollen-BoL Mini-Symposium were based on pollen collected from four bee species, *Megachile felina, M. maxillosa*, *M. karooensis* and *M. murina*. I performed all the pollen collections, laboratory work and analyses on the sequence data for this presentation.

10. Gous A., Willows-Munro S., Johnson, S.D., Eardley C.D., Swanevelder D.Z.H. 2016. Floral choice in South African bees (*Megachile*: Megachilidae) as revealed by pollen DNA metabarcoding of historic specimens. [manuscript in preparation for submission to Molecular Ecology, January 2017].

This paper presents the work performed for Chapter 4. I designed the experiment, performed all the laboratory work, analysed data, and wrote and edited the paper.

Appendix C

11. Bell K, de Vere N, Keller A, Richardson R, Gous A, Burgess K, Brosi B (2016) Pollen DNA barcoding: current applications and future prospects. Genome 59, 629-640. [Published].

I was approached by Dr. Karen Bell and asked whether I would like to co-author the first review article on pollen barcoding with the rest of the presenters of the pollen barcoding session at the 6th International Barcode of Life Conference. I wrote the sections on plant-pollinator interactions over space and time, ancient pollen DNA barcoding, and airborne allergen monitoring. I gave input into the other sections and reviewed the manuscript editorially.

_____

Signed: Annemarie Gous

Date:

ABSTRACT

Interactions between plants and their pollinators are often poorly understood, specifically in a species-diverse country such as South Africa. Traditional methods of studying plant-pollinator interactions are time-consuming and imprecise. This study aimed to develop a technique that uses genetic analyses to identify pollen provenance directly from bees (Megachilidae) housed in a historic collection and to apply this technique to investigate floral choice differences in species of megachilid bees from three regionally important areas in South Africa: the Succulent Karoo, Savanna, and a widespread group, with bees occurring throughout the country. To develop the technique to accurately identify provenance, pollen was sampled from *Megachile venusta* specimens in the collection. Three DNA barcode regions were amplified and sequenced on an Illumina MiSeq instrument: the internal transcribed spacer 1 (ITS1) and internal transcribed spacer 2 (ITS2) regions, and the ribulose-1,5-biphosphate carboxylase (*rbcL*) gene. Sequenced reads were compared to sequence reference databases that were generated by extracting sequence and taxonomic data from GenBank. ITS2 reads were also compared to an established ITS2 database for Viridiplantae. More diverse plant classifications were obtained with ITS2 compared to ITS1. Amplification and sequencing of *rbcL* was inconsistent on pollen sampled from historic specimens. To study how floral choice differed in three South African regions, ITS2 was sequenced on Illumina HiSeq and MiSeq from pollen sampled from two different bee species from each region. Sequence reads were compared to the previously published ITS2 sequence reference database. Generalised linear models (GLM) indicated that the mean number of both plant families and species varied significantly between bee species. No significant effect of the time since bee collection was found. Taxon identifications were only confidently interpreted on family-level due to very limited local plant representation in sequence reference databases. DNA metabarcoding of mixed-origin pollen samples provided a faster, more accurate method of determining pollen provenance, without the need for expert palynologists. The use of historic collections to sample pollen directly from pollinators provided additional value to these collections. Sampling pollen from historic collections can also provide the spatial and temporal scales for investigations into changes in plant community structure or pollinator floral choice in the face of global climate change.

ACKNOWLEDGMENTS

I am sincerely grateful to a number of people and institutions that supported me in various ways throughout the duration of my PhD:

I would like to thank Dr. Connal Eardley at Biosystematics (ARC), Dr. Dirk Swanevelder at the Biotechnology Platform (ARC), and Dr. Sandi Willows-Munro (UKZN) for their valuable supervision, guidance and mentorship throughout my studies. Thank you for allowing me the space to grow as a research scientist.

Without the funding from various institutions, my PhD would not have been realised. A special thanks is due to the Department of Science and Technology and National Research Foundation's Professional Development Programme (PDP) for awarding me a PhD bursary. The University of KwaZulu-Natal and the Agricultural Research Council (ARC) PDP deserves thanks for paying my study fees. I am also thankful to the National Research Foundation, the Belgian Directorate-General for Development Cooperation, partim GTI, and the Biotechnology Platform at the ARC for covering research costs.

I am grateful to the anonymous reviewers of the manuscripts, emanating from this PhD, that we have submitted to peer-reviewed journals. The comments we have received helped to improve the quality of the work.

I would like to thank Jurgens de Bruin for the time he put into creating AvoGaDRo, and helping with bioinformatics problems.

Thank you to Dr. Charles Hefer for always being willing to help with bioinformatics-related issues, and spending so much time on getting this project running.

I would like to thank Prof. Steve Johnson for his extremely valuable help and guidance towards the end of my PhD studies.

To the other three members of the Golden Girls, Minique de Castro, Erika Viljoen and Stephanie Cornelissen, I would not have been able to do this without you. Thank you for all the stimulating discussions, help with troubleshooting, times you reviewed written works, and most of all, the fun and laughter.

I am also grateful to my other labmates and friends for making difficult times easier with laughter and good wine. A special thanks is necessary to Soné Ungerer Hendriks for putting me in touch with the right people at the right time.

Thank you to my wonderful parents, sisters, and in-laws for always understanding and supporting me, and not asking how the writing up is going too many times!

Lastly, to my husband Stefan de Jager, you loved and supported me through it all. Thank you. I dedicate this to you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research (nature and scope)

South Africa is well-known for its biodiversity, with three Conservation International global hotspots declared within its borders, namely the Succulent Karoo, the Cape Floristic Province and the Maputaland-Pondoland-Albany region (Mittermeier *et al*., 2011). Complex ecosystems are present across the country, and various different taxa have become interdependent on one another for survival. Pollination is an important ecosystem service provided to humans by animals from these ecosystems, and contributes directly to about a third of the food we consume (Klein *et al*., 2007). The most important pollinators of angiosperms are bees (Hymenoptera) represented by an estimated 25 000 species across the world. Many species are yet to be discovered and formally described (Michener, 2000).

Bees are excellent pollinators of many important crops, such as tomatoes, grapefruit, sunflowers, coffee and many more (Klein *et al*., 2007). Some species of bees are managed as pollinators for agricultural use, with the most well-known of these being the generalist honeybee, *Apis mellifera* (vanEngelsdorp and Meixner, 2010). The leafcutter bee, *Megachile rotundata*, is commonly managed in North America for alfalfa (lucerne) pollination (Pitts-Singer and Cane, 2011). However, in South Africa, the honeybee is the only managed pollinator and little is known about other native bee species and their plant interactions.

Plant and bee diversity and endemism are particularly high in the western areas of South Africa. Bee endemism, not diversity, is lower towards the east of the country but remains at noteworthy levels when compared to the western part of the country (Eardley, 1989; Pauw and Stanway, 2015; Van Wyk and Smith, 2001). Interactions between plants and bees appear to be more specialised where endemism is higher, and these associations are believed to be particularly vulnerable to changes in the environment (Packer *et al*., 2005; Zayed *et al*., 2006). High levels of interaction specialisation between plants and pollinators occur in hyper-diverse areas (Pauw and Stanway, 2015). Due to the high levels of plant and pollinator diversity in the western region of South Africa, plant-pollinator interactions in these regions are expected to have a similarly high level of specialisation. Disruptions in these interactions could lead to the loss of both bee and plant diversity. These interactions are greatly understudied, including in South Africa, due to the large number of both plant and pollinator species involved, as well as difficulties associated with the technicalities associated with its study. This study focused on bee-plant associations that could result in pollination.

Plant-pollinator studies are traditionally performed through observation in the field, with researchers spending many hours observing which plants are visited by pollinators (Johnson, 1997). This is incredibly time-consuming. Additionally, both plants and pollinators need accurate identification as well – another time-consuming activity requiring different specialists. Pollen loads from pollinators can also be studied microscopically, but this method is also riddled with difficulties (Rahl, 2008). Advances in identifying plant taxa using genetic methods have made it possible to address this problem in an alternative manner, saving time and costs.

DNA barcoding is the process of identifying an organism based on a short segment of DNA, and has been used successfully since the early 2000's to taxonomically identify both animals and plants (Hebert *et al*., 2003). This process has advanced significantly over the years and is now used in conjunction with next-generation sequencing (NGS) on samples that are of mixed origin, such as samples taken from the environment consisting of many different types of organisms that cannot be separated easily (Hajibabaei *et al*., 2011). Pollen loads taken from pollinators are often of mixed origin, as pollinators appear to visit more than one taxon of plant during a day, and may carry pollen from a number of different plant species.

Studies on bees and pollinators undertaken by taxonomists and other researchers lead to many specimens being caught for identification purposes. These specimens are caught across the country and are usually kept in well-maintained insect reference collections (Pennisi, 2000). Pollen loads on these specimens are seldom used for plant identifications and is an untapped resource for studying plant-pollinator interactions. The ability to use genetic methods that result in faster identification of pollen origins allows interactions to be studied without intensive sampling and without the necessity for pollen morphology expertise. This access to samples broadens the scope of research that can be done.

1.2 Justification

Few studies have been published that utilise DNA analyses to identify the plant origins of pollen loads from pollinators, with no genetic studies investigating the relationships of South African bee species with the indigenous flora available. One of the first studies sequenced pollen DNA from Hawaiian bees using Sanger sequencing (Wilson *et al.*, 2010), which (without initial cloning) is not optimal for distinguishing mixed-origin samples. A single study using NGS to determine the taxonomic composition of mixed pollen was published at the time this PhD was conceptualised (Valentini *et al*., 2010). At this stage, the only other way of studying plant-pollinator interactions was through observation, and microscopic

identification of pollen mixtures. There are many different pollen exine structures, making exact morphological identification of pollen difficult, and is thus mostly only accurate to family or genus level (Mander, 2016; Rahl, 2008). The expectation of accurate species-level identifications based on NGS data being referenced against comprehensive sequence databases led to the exclusion of morphological comparison of pollen in this study. Optical and molecular identification methods were compared by Keller *et al*. (2015) and the level of diversity that could be detected in pollen was improved greatly upon by NGS. Richardson *et al*. (2015) also found corresponding identifications between microscopic and NGS analysis of pollen samples. Ideally, a reference database of plants from the regions of interest would been created using the selected barcode regions. However, the incredible diversity and number of plant species in South Africa, and in particular in the Succulent Karoo, make this an enormous task – one actively being addressed by the research community.

The interaction of bees with plants they may pollinate has not been studied on a molecular level in Africa before. This PhD study also focused on various indigenous bee species occuring in different ecosystems. The number of floral visits of bee species from the highly diverse Succulent Karoo in the western region of South Africa were compared to the Savanna region in the East of the country. Additionally, this study is also novel in that historic bee specimens were used to sample pollen loads, which provide temporal data on which the conclusions were drawn. This further broadens the scope of the work when compared to other pollen metabarcoding studies, and adding an additional dimension to pollination biology. This investigation showed that historic collections can be used for broader applications than taxonomic identification, and emphasises the important role that both taxonomy and molecular biology play in pollination biology.

1.3 Aims

The main aims of the research were:

- Development, testing and evaluation of pollen metabarcoding workflows using historic insect collection specimens.
- Investigation of the extent of interspecific differences of floral choice in long-tongued megachilid bees (Hymenoptera: Apoidea: Megachilidae) in the Succulent Karoo and the Savanna biomes of South Africa.
- To investigate whether widespread bee genera are more oligolectic in the Western Cape Province than elsewhere in South Africa, using *Megachile* as the research subject.

1.4 Objectives

The specific research objectives of the study were to:

- Refine molecular barcoding techniques to identify pollen sampled from bee specimens of different ages from historical insect collections
- Determine the lowest taxonomic level to which plants can accurately be identified from pollen DNA sampled from historic specimens using the publicly available barcode gene sequences as a reference database.
- Determine the plant origin of pollen found on sampled bees by metabarcoding.
- Determine the amount and identity of plant taxa visited by bees from the Succulent Karoo, the Savanna and those widespread in South Africa.

1.5 Outline of thesis structure

The thesis is structured as chapters which are written as scientific manuscripts, such that each chapter contains its own introduction, materials and methods, results, discussion and conclusion sections.

- Chapter 2 is a review of the literature and it is reiterated throughout the remaining chapters. This chapter is focused on the value of pollination as a whole, the role that bees play in this system, and highlights the importance of bee-pollination and the role of plant-pollinator interactions in South Africa. The chapter mentions recent advances in the field of molecular biology that have made studying these interactions using DNA barcoding possible.

- In Chapter 3 the process of developing and optimising pollen metabarcoding methodologies to enable pollen identification from samples obtained from historic insect collections is discussed. Historical samples are difficult to work with and this chapter includes details on the optimisation of the methodologies, such as DNA extractions, PCR, sequencing and bioinformatics analyses, required to perform such a study from this origin.

- In Chapter 4 the optimised pollen metabarcoding protocols developed in Chapter 3 is utilised to investigate the biological questions posed in the thesis. Different *Megachile* bee species from three distinct geographic regions across South Africa were sampled to compare pollen choices between the regions. Due to the high levels of plant and bee diversity in the Succulent Karoo biodiversity hotspot, it was expected that floral choices in bees from this region would be more specialised than those in bees from the Savanna region, or in those that occur widespread across South Africa. This hypothesis was tested in Chapter 4.

- In the final chapter, Chapter 5, the work from all preceding chapters is integrated and conclusions are made about the value of the research. This chapter also outlines future perspectives and discusses which potential research opportunities stem from this work.

- An appendix contains all information that could not form part of the chapters mentioned above.

1.6 References

Eardley, C.D., 1989. Diversity and endemism of southern African bees. Plant Protection News 18, 1–2.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6, e17497. doi:10.1371/journal.pone.0017497.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218.

Johnson, S.D., 1997. Pollination ecotypes of *Satyrium hallackii* (Orchidaceae) in South Africa. Bot. J. Linn. Soc. 123, 225–235. doi:10.1111/j.1095-8339.1997.tb01415.x.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. doi:10.1098/rspb.2006.3721.

Mander, L., 2016. A combinatorial approach to angiosperm pollen morphology. Proc. R. Soc. B Biol. Sci. 283, 20162033. doi:10.1098/rspb.2016.2033.

Michener, C.D., 2000. The bees of the world, 2nd ed. Johns Hopkins University Press, Baltimore.

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., Gascon, C., 2011. Global biodiversity conservation: The critical role of hotspots, in: Zachos, F.E., Habel, J.C. (Eds.), Biodiversity Hotspots. Springer, Heidelberg, pp. 3–22.

Packer, L., Zayed, A., Grixti, J.C., Ruz, L., Owen, R.E., Vivallo, F., Toro, H., 2005. Conservation genetics of potentially endangered mutualisms: Reduced levels of genetic variation in specialist versus generalist bees. Conserv. Biol. 19, 195–202. doi:10.1111/j.1523-1739.2005.00601.x.

Pauw, A., Stanway, R., 2015. Unrivalled specialization in a pollination network from South Africa reveals that specialization increases with latitude only in the Southern Hemisphere. J. Biogeogr. 42, 652–661. doi:10.1111/jbi.12453.

Pennisi, E., 2000. Taxonomic Revival. Science 289, 2306–2308.

Pitts-Singer, T.L., Cane, J.H., 2011. The alfalfa leafcutting bee, *Megachile rotundata*: The world's most intensively managed solitary bee. Ann. Rev. Entomol. 56, 221–237. doi:10.1146/annurev-ento-120709-144836.

Rahl, M., 2008. Microscopic identification and purity determination of pollen grains, in: Jones, M.G., Lympany, P. (Eds.), Allergy Methods and Protocols. Humana Press, Totowa, NJ.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015b. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl. Plant Sci. 3, 1400066. doi:10.3732/apps.1400066.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA barcoding for honey biodiversity. Diversity 2, 610–617. doi:10.3390/d2040610.

Van Wyk, A.E., Smith, G.F., 2001. Regions of floristic endemism in southern Africa: A review with emphasis on succulents. Umdaus Press, Hatfield.

vanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95. doi:10.1016/j.jip.2009.06.011.

Wilson, E.E., Sidhu, C.S., LeVan, K.E., Holway, D.A., 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol. Ecol. 19, 4823–4829. doi:10.1111/j.1365-294X.2010.04849.x.

Zayed, A., Packer, L., Grixti, J.C., Ruz, L., Owen, R.E., Toro, H., 2006. Increased genetic differentiation in a specialist versus a generalist bee: implications for conservation. Conserv. Genet. 6, 1017–1026. doi:10.1007/s10592-005-9094-5.

CHAPTER 2: POLLINATION: IMPACT, ROLE-PLAYERS, INTERACTIONS AND STUDY – A SOUTH AFRICAN PERSPECTIVE

2.1 Abstract

Plant-pollinator interactions are essential for maintaining both pollinator and plant communities in native and agricultural environments. Animal-instigated pollination can be complex. Plants are usually visited by a number of different animal species, which in turn may visit flowers of several plant species. Therefore, the identification of the pollen carried by flower visitors is an essential first step in pollination biology. It is of particular importance to understand the floral preferences of native pollinators, as it is sometimes suggested that foreign pollinators be imported to perform pollination in lieu of less efficient native counterparts. The skill and time required to identify pollen based on morphology has been a major stumbling block in this field. Advances in the genetic analysis of DNA, using DNA barcoding, extracted directly from pollen offers an innovative alternative to traditional methods of pollen identification. This technique, which is reviewed in detail, can be used on pollen loads sampled from bees in the field and from specimens in historic collections. Here the importance of pollination, the role-players involved, their management, and the evolution of their interactions, behaviour and morphology are reviewed – with special focus on South African bees.

Keywords: *floral choice, bees, plant-pollinator interactions, palynology, pollen, ITS*

2.2 Introduction

South Africa has one of the world's most diverse landscapes, with high plant and pollinator diversity and endemism (Mittermeier *et al*., 2011). Healthy plant-pollinator interactions are important to maintain both native plant and pollinator communities. The interactions between pollinators and their host plants are complex and very little is currently known about the floral choices of indigenous bees, specifically in South Africa. Most plant-pollinator interaction studies rely on lengthy field-based experiments (Gess and Gess, 2004). Due to pressure from urban development, overexploitation of natural resources, and climate change, many of the nine biomes in South Africa are under threat (Department of Environmental Affairs, 2012). The effects of an anthropogenic influence on the environment dictates improved methods of studying plant-pollinator interactions, to understand how they may be influenced by environmental and ecological changes.

Observing pollinators in the field is not the only way to study their interactions with plants. Pollinator pollen load identification gives insight into the species' floral visits. Pollen loads provide a snapshot of the interactions with the plant community at the time they were caught. Bees caught at a flower patch, for example, can have varying pollen loads since they can either be on their way to the flowers from the nest, or on their way from the flowers back to the nest, or busy foraging at the patch. Neither identifying the pollen found on a sampled bee, nor netting a bee on a flower, can give definite answers regarding the plants it pollinates, and similarly inferences on possible fruit and seed set cannot be made (Gess and Gess, 2014). However, some pollination inferences can be drawn nonetheless, especially if multiple bee samples are investigated. Pollen loads sampled from pollinators could be identified by classic microscopic palynology or genetic methods. These two approaches are discussed in detail below.

Here, the role of pollination in agriculture and natural plant populations, with a special focus on bee-pollination, is evaluated. The value of bee-pollination in a South African context is reviewed and discussed, as well as the potential impact that the introduction of a foreign bee species could have on highly diverse native bee populations. Advances in studying plant-pollinator interactions using genetic methods are also reviewed.

2.3 The value of pollination

Functional ecosystems require various essential ecosystem services to be performed. Ecosystem services are defined as services provided to human welfare by maintenance of the ecosystem by organisms that interact in the ecosystem and pollination is one such extremely important service (Losey and Vaughan, 2006). Plant-pollinator interaction is, in most instances, an intimate mutualistic relationship, where both parties are reliant on each other for survival; plants for reproduction and pollinators for food or other forms of reward. Although a plant might have multiple pollinators, it is possible that one or more of these pollinators are specialists and may therefore rely heavily on that specific plant taxon for survival (Klein *et al*., 2007). A decline in the host plant numbers would ultimately lead to a decline in its specialist pollinators, and *vice versa*, having an important impact in maintaining biodiversity and ecosystems.

Pollination is not only extremely important in natural ecosystems, but also in artificial production environments. Biotic pollination of crops is important from an agricultural production perspective since approximately one-third of all human food consumption results from animal-pollinated plants, of which up to 75% is used directly as food (Klein *et al*.,

2007). A decline in crop pollinator populations would thus negatively impact crop production. The importance of pollinators has been illustrated in a study conducted on 137 single crops and five commodities, where increases of 68.4% in production of the leading single crops and 71.6% in production of commodity crops were found with animal pollination (Klein *et al*., 2007). It was estimated that native insects in the United States of America (excluding the introduced honeybee) were solely responsible for $3.07 billion worth of fruit and vegetable production in 2003 (Losey and Vaughan, 2006). This clearly indicates the value of indigenous pollinators to human society for maximal crop production.

2.4 Plant-pollinator interactions

Pollination, although an important ecosystem service and often crucial for the survival of both parties involved, is in essence an inadvertent process. From the view of the pollinator, it is not its specific goal to provide this service to its mutualistic plant partner, but it is rather a coincidental result of its actions while visiting these hosts (Kearns *et al*., 1998). Bees, for example, can deliberately collect pollen, or pollen can passively adhere to bee bodies while they are visiting flowers (Michener, 2000). Mutualism between the plant and the pollinator is based on rewards (pollen, nectar and oil) that the pollinator receives from the plant, and the plant gains the service of successful pollination and securing its reproductive success (Johnson, 2010; Kearns *et al*., 1998; Michener, 2000). This mutualism can be facultative or obligate, depending on whether the plant is self-compatible or whether it is monoecious or dioecious (Kearns *et al*., 1998). Parasitic interactions are also possible when a potential pollinator takes pollen and nectar from the plant without playing any role in its pollination (Kearns *et al*., 1998; Michener, 2000). Interactions between plants and potential pollinators therefore range from parasitic to obligate mutualistic, with each plant-pollinator interaction a developing relationship, based on how the plant and pollinator adapt to suit each other.

2.4.1 Evolution of plant-pollinator interactions

Interactions between pollinators and their target plants are usually regarded as being either generalised or specialised (Minckley *et al*., 2000). Generalisation describes an interaction where a plant has flowers that are accessible and attractive to many different pollinator species. Specialisation refers to flowers that are sufficiently specialized as to be attractive and/or accessible to only a single type of pollinator (Minckley *et al*., 2000; Padyšáková *et al*., 2013). The same can be applied from the pollinator's perspective, where the range of plants a pollinator prefers relates to it being generalised or specialised (Bosch *et al*., 2009; Minckley *et al*., 2000), as discussed below. In some extreme cases, both parties could co-evolve

morphologically and behaviourally to only allow one-on-one plant-pollinator interactions, where the plant protects access to rewards for its specific pollinator - a feature of, among others, many genera in the Orchidaceae (Hetherington-Rauth and Ramírez, 2015).

It has been argued that the formation of specific floral structures in plants is largely driven by means of natural selection from their respective pollinators (Johnson, 2010). Better pollinator-flower compatibility would, therefore, result in higher selection, through increased fertilization events of these individuals. The pollinator-plant interaction is important since pollinator or floral adaptations can drive speciation as suggested by the diverse floras of the Cape Floristic Region, South Africa (Johnson, 1996). Flowers of angiosperms can gain suites of adaptive traits to make them more suitable to a certain type of pollinator or pollinator guild, also known as floral syndromes (De Merxem *et al*., 2009). These floral adaptations or syndromes can lead to reproductive isolation and drive speciation, but isolation and speciation are not necessarily connected (Johnson and Steiner, 2000). It is important to remember that floral syndromes are not an absolute definition of a plant's pollinators but rather a description of how unrelated plants evolved to have similar floral traits (Johnson and Steiner, 2000).

2.4.2 Bees as pollinators

The most common and invariably important biotic pollinators of angiosperms are bees (Hymenoptera: Apoidea, Klein *et al*., 2007) since they actively collect pollen as food for themselves and/or their larvae (Michener, 2000). There are an estimated 25,000 bee species in the world of which approximately two-thirds are taxonomically described (Johnson, 2010). Bees are important pollinators of tropical forest trees (Bawa, 1990) and play an essential role in the pollination of smaller trees, shrubs and herbaceous plants (Johnson, 2010). Many crop plants are bee-pollinated such as sunflower, tomato, canola, cowpea and coffee, to name but a few (Klein *et al*., 2007; Kwapong *et al*., 2013).

South Africa has high levels of both bee and plant diversity, especially in the southern and western predominantly winter rainfall parts of the country. Approximately 50% of the bee species known to occur in sub-Saharan Africa are also located in South Africa (Kuhlmann, 2009). Moreover, 95% of the bee species found in the winter rainfall region occur only in southern Africa (Eardley, 1989). The moist, eastern part of the country has also been shown to be diverse in its bee species composition, albeit less so than in the arid western part of South Africa. The eastern region contains a high endemism level of 75% (Eardley, 1989; Kuhlmann, 2009).

Although South Africa is particularly rich in pollinators and floristic diversity, relatively few comprehensive studies have investigated pollinator-plant interactions in the country as a whole. One such study that focused on bees in the arid western region of southern Africa, documented that 16,229 plants were visited by 924 species of non-*Apis* bees, wasps and pollen wasps (Gess and Gess, 2004). The bees in this study were represented by 420 different species that visited 34 out of the 36 available plant families in the study area. The four plant families most frequently visited by bees were the Fabaceae, Asteraceae, Aizoaceae and Zygophyllaceae. The foraging habits of all the different bee families studied, i.e. the Colletidae, Andrenidae, Halictidae, Mellitidae, Megachilidae and Apidae, were oligolectic (specialised) to narrowly or broadly polylectic (generalised) – though none of the families were nearly as polylectic as *A. mellifera*. However, in the Mellitidae, half of the observed species were found to be oligolectic for members of the *Wahlenbergia* plant genus (Campanulaceae). This work has shed light on the diversity and foraging habits of pollinators in the arid and semi-arid regions of the western part of southern Africa (Gess and Gess, 2004).

No similar regional-scale study of bees in the eastern part of South Africa is currently available. However, there are a few smaller studies of pollinators for specific plant species, for example in the KwaZulu-Natal midlands in the eastern part of South Africa, *Wahlenbergia* were visited by halictid species and *A. mellifera* (Welsford and Johnson, 2012). It is noteworthy that halictid bees have also been reported to be oligolectic for *Wahlenbergia* in Australia (Goulson, 2003). This is possibly an adaptation of bees to their locally available flora as was previously reported (Ginsberg, 1981). The few available studies, together with the high floristic and bee diversity and endemism in South Africa, highlights the need for further studies into the diversity of bee interactions with plant species – studies that are needed to elucidate floral choice patterns within South African bee populations.

2.4.3 Bee adaptations for foraging

Bees are active foragers, collecting various substances from flowers. During foraging activity, pollen grains become attached to their bodies. Specialised branched hairs on their bodies trap the pollen grains during the collection of pollen, nectar or oil (Thorp, 2000). Electrostatic charges on the hairs also aid in the transfer of pollen from anthers to the bee body. In addition, modifications of hairs on the mouthparts, undersides of the heads, or faces of bees all assist in extraction of pollen from flowers (Michener, 2000; Thorp, 2000). During foraging the pollen is groomed from the insect body into structures used to carry it to the nest. These transport structures, also known as scopae, are brushes of hairs located on the hind tarsi of most bee species, or on the bottom of the abdomen as in the Megachilidae (Michener, 2000). Structural and behavioural adaptations for the collection of pollen have previously been reviewed in

more detail (Thorp, 2000). When pollen is groomed into transport structures, these pollen grains are generally not available for pollination (Westerkamp, 1991) as they are tightly packed. Loosely adhered pollen grains on bee bodies are more important in pollination.

2.4.4 Generalist versus specialist interactions

Bees can exhibit generalist or specialist behaviour in their floral choices for specific requirements. Most bees within the eusocial groups, such as honeybees, bumblebees and most stingless bees, are polylectic in terms of pollen and nectar collection (Michener, 2000). They visit plants from a wide variety of taxa that are available for pollen and nectar collection. Polylectic bees still show floral constancy, that is, they make repetitive visits to plants from the same taxon that they have previously visited while the resource is available (Wilson and Stine, 1996). Floral constancy is likely a learned behaviour that increases foraging efficiency during a single trip (Michener, 2000). Some bees are more selective as far as pollen is concerned. Solitary bee groups show either polylectic or oligolectic foraging behaviour. When visiting only a single species of plant, bees can be said to be monolectic, but behaviour mostly tends to range from narrowly to broadly oligolectic, with the boundaries between them remaining unclear (Linsley, 1958). Oligolectic bees still visit flowers from plant taxa other than those from which pollen is collected for other resources, such as nectar, oils and others (Bosch *et al*., 2009).

Floral choices of bee pollinators play an important role in the sustainability of a plant community. According to food web theory, the more complex the plant-pollinator interactions are, the less susceptible the plant community is to disturbances or extinction (Melián and Bascompte, 2002). If one of the interactions fails for some reason, this interaction would likely be taken over by some other pollinator involved in the complex interaction matrix. On the contrary, a plant community with a high level of pollinator specialisation would be markedly more vulnerable to any disturbance in its interactions. Plant communities with high diversity would therefore be able to sustain an increased level of bee specialisation, whereas a low diversity plant community would evolve to increase the complexity level of its plant-pollinator interactions. This was experimentally verified by increasing plant diversity in a gradient and showing an increase in solitary bee specialisation as plant species richness increased (Ebeling *et al*., 2011). Oligolectic bees are also more susceptible to changes in their environment and thus extinction. Since specialist bees have a more restricted foraging range, their effective population size (N_e) and levels of genetic variation are lowered, making these bees and pollination systems vulnerable and in need of protection (Packer *et al*., 2005).

2.4.5 Studying bee-plant interactions

Plant-pollinator interactions have historically been studied through careful and patient observation. This usually involves lengthy field-based experiments (Gess and Gess, 2004) with plant species in a demarcated area studied for a specific time to see which, if any, animals visit the flowers (Gess and Gess, 2004; Kwapong *et al*., 2013). Visiting a plant still does not necessarily mean it is a pollinator of that plant. Even when pollen is transferred to a receptive stigma, genetic incompatibility between pollen and plant may still prevent fertilisation from taking place, such as no pollen tube germination, termination of pollen tube growth down the style, or pollen simply being unviable (Dafni, 1992). Unsuccessful fertilisation could occur due to pollen from a different plant species being deposited onto the stigma. Self-incompatibility also prevents fertilisation by pollen from the same plant (Dafni, 1992). Laboratory experiments with captive pollinators can also be conducted, especially for confirmation studies, but these do not reflect the pollinators' natural environment (Harder, 1988).

Field-based observations can be followed by determination of the pollen loads on potential pollinators and the assessment of pollination effectiveness, as measured by the degree of fruiting and seed set through examinations of the individual plants visited (Klein *et al*., 2003). Pollen load determination of any potential pollinator requires capturing the animal in question and the removal of the pollen it carries. The pollen morphology is then carefully analysed (palynology) to identify, or confirm, the plant species from which it originates. Palynology-based identification requires sufficient knowledge of the field and intimate familiarity and expertise with pollen morphological structures, especially of closely related plant species. Furthermore, some form of microscopy is required for visualising the pollen's morphological features used in the identification process, such as scanning electron microscopy (SEM) or compound light microscopy (LM) (Dafni, 1992; Rahl, 2008). These technologies require specialised sample preparation methods, skill to prepare and operate instruments, and experience to best obtain comparative morphological features between samples. Additionally, pollen morphological features of different plant species or genera can be extremely similar, especially if they are closely related, thereby requiring a wide palynology knowledge base to accurately distinguish between these samples (Rahl, 2008). Mixed pollen samples from closely related species would therefore require a highly skilled and knowledgeable palynologist, usually an expert familiar with the pollen from the area under investigation. The pollen-carrier must also be identified to make accurate inferences (Gess and Gess, 2004), a function normally performed by different taxonomic specialists in entomology. This makes plant-pollinator interaction studies time-consuming and highly multi-disciplinary, and requires expertise in the fields of taxonomy, botanical reproduction, palynology, entomology, and microscopy.

2.5 Pollinator declines

The most well known bee species is the honeybee, *Apis mellifera* and most bee-related research has focused on this species (Klein *et al*., 2007; vanEngelsdorp and Meixner, 2010). Honeybee populations have been reported to be declining in certain areas of the world, such as central Europe, the United States of America, and Mexico (Goulson *et al*., 2015; vanEngelsdorp and Meixner, 2010). Although there is as yet no consensus on what may be driving *A. mellifera* population decline, factors such as insecticide use on crops, pests, diseases and predators, a decrease in genetic variation of bee colonies, and the effects of climate change and limitations in the trade of bee colonies may all play a role (Goulson *et al*., 2015; vanEngelsdorp and Meixner, 2010).

In South Africa and some other countries, honeybee numbers are seemingly not declining. This is attributed to beekeeping (apiculture) and the introduction of numerous alien plant species, widening the honeybee foraging range (Jaffé *et al*., 2010). Honeybee colonies in South Africa were also seen previously to be resilient to most diseases. This view was supported when an outbreak of American foulbrood (AFB) in 2011 did not cause any major colony losses (Human *et al*., 2011). More recently in 2015, however, another AFB outbreak in South Africa reduced the number of colonies in the Western Cape by 40% (Kings, 2015).

Not only honeybee populations have been declining over time. Researchers in Britain and the Netherlands have found a correlation between declines in native bees, and declines in outcrossing plants dependent on these bees (Biesmeijer *et al*., 2006). The native bees in both countries had narrow habitat requirements and produce single broods per year. Honeybee data were specifically excluded, but data for all native species for both countries were included in the analyses. The ultimate cause and direction of the declines could not be determined from the data, but the aforementioned study supports the notion that species reliant on a wider range of interactions within a plant-pollinator system would be more resilient when threatened (Biesmeijer *et al*., 2006).

A changing climate, inappropriate land-management and a growing human population have contributed to the reduction of overall biodiversity, including native, wild bee populations across the world (Biesmeijer *et al*., 2006) An important determinant of the maintenance of plant-pollinator interactions is the way land is used and managed (Kearns and Inouye, 1997; Klein *et al*., 2007). When agriculture is intensified on a piece of land, bee diversity can decline due to less opportunities for them to nest, lower foraging diversity, and possible insecticide use on crops (Klein *et al*., 2007). In South Africa's Karoo, all of these factors have been documented to result in a decline of bee and wasp diversity (Gess and Gess, 2014). Game farming started to replace stock farming in this region, and land is often overexploited

due to the inability of game to move to different areas once the resources have been exhausted in the area where they are kept (Gess and Gess, 2014). Tourism opportunities availed from game farming resulted in the introduction of animal species not normally found in the area, and also no period of rest for the land to recover. In areas where large plots of single cultivated plants (monocultures) are found, such as in the wheat fields and wine lands of the Western Cape, very little of the natural vegetation remains. These areas are also likely sprayed with pesticides. In combination, these factors can cause the complete loss of native bee and wasp communities (Gess and Gess, 2014).

2.6 Managing pollinators

Native honeybees are currently the only pollinators that are being managed in South Africa. The management of honeybee colonies for pollination purposes has several advantages and disadvantages. Due to their generalist foraging habits they are suitable for use on many different crop species. Like many other bee species, they are nevertheless unable to pollinate all crops (Kearns and Inouye, 1997). Additionally, they pack their collected pollen into the corbiculae on their hind legs after moistening it with nectar or honey. This results in limited pollen available for pollination (Michener, 2000) and renders the honeybee a poorer pollinator when compared to other bee species (Westerkamp, 1991). African honeybees are also aggressive and care needs to be taken when working with them (Kearns and Inouye, 1997). Their susceptibility to pesticides, diseases and parasites also threaten their commerciality (vanEngelsdorp and Meixner, 2010) and it is consequently important that pollination management strategies using other native species be explored.

Crop production has increased dramatically over the past decades to meet the demands of growing populations. This means that pollinator population sizes are not adequate to deal with the demand. Managed pollinators provide a solution to this problem. It would be best to manage indigenous pollinator populations, such as the honeybee in Africa, to alleviate the problem. In light of some of the inadequacies of honeybees as pollinators, and it being the only group of managed pollinators in South Africa, investigations into the floral choices of native bees could identify candidates for management in the place of, or in addition to, honeybee populations. It is possible that a carpenter bee, *Xylocopa scioensis*, could be managed for tomato pollination in South Africa, much like the leafcutter bee, *Megachile rotundata*, is used for alfalfa (lucerne) pollination in North America (Pitts-Singer and Cane, 2011). Tomatoes require vibratile ("buzz") pollination, something not effectively achieved by honeybees. This could potentially avoid the need to import foreign pollinators as has been proposed for the Western region of South Africa, to aid in the pollination of particularly vibratile-pollinated crops (Rodger *et al*., 2013).

2.6.1 Effects of **Bombus** *introduction*

The introduction of generalist pollinators, like the honeybee, to provide pollination services to multiple crop species is an economic choice. The honeybee (*Apis mellifera*) naturally occurs throughout Africa, Europe and western Asia, but has been introduced to a significant proportion of the rest of the world as a successful pollinator (Goulson, 2003). The impact of the introduction of any alien species into an environment should be carefully considered, as it could be devastating to the native ecosystems. The European bumblebee, *Bombus terrestris*, is most commonly found throughout Europe but has also recently been introduced into other countries (Hingston *et al*., 2002; Torretta *et al*., 2006) and its effect on the environment is well documented. Suggestions have been made that South Africa would benefit from introducing *B. terrestris* and managing their populations for pollination (Rodger *et al*., 2013). South Africa does not have any native species of *Bombus*.

The impact of the introduction of any alien species into an environment should therefore be carefully considered, as it could be devastating to native ecosystems. Since its introduction into foreign habitats, *B. terrestris* has had major effects on native plant and bee populations, both positive and negative. It has been shown to increase pollination overall, but decrease efficiency of pollination in native plants, enhance pollination in weeds, and cause displacement of native pollinators (Hingston *et al*., 2002). It was discovered, by chance, to have invaded Neuquén Province in Argentina during a survey of floral visitors of shrubs. The bees were thought to have entered Argentina from Chile, as extensive studies of natural and museum populations in Argentina did not provide any historical evidence for the presence of *B. terrestris* (Torretta *et al*., 2006). Analysis of the pollen found on the *B. terrestris* individuals showed that they were competing with an indigenous *Bombus* species for food on seven out of the eight host plants. In Japan, *B. terrestris* was introduced to pollinate crops, but then escaped from greenhouses, became naturalised and has had negative consequences on the native bee populations (Goka, 2010). Resource competition between the introduced and native bumblebees was found in the Japanese study. Introduced species also interfered with the reproduction of both native plants and native bumblebees (by interspecies crosses). Additionally, new parasites were introduced to native populations. Native bumblebee populations have been displaced by *B. terrestris* before (Inoue *et al*., 2008), making its invasiveness of great concern. Previously, it has invaded Tasmanian national and urban gardens where it was found foraging on a wide variety of plant types (Hingston *et al*., 2002).

Evidently, the use of a foreign pollinator in South Africa should be carefully considered. Native, oligolectic bee species in South Africa would be particularly vulnerable to an introduction of *B. terrestris*, or any other polylectic species, that are to be managed for pollination services. So far, permits have not been granted to import *B. terrestris* into South

Africa, but in February 2014, Senegal received a shipment of *B. terrestris* colonies from Belgium (Siegmund, 2014) signifying the first introduction of this species in sub-Saharan Africa (Rodger *et al*., 2013). The preceding evidence clearly indicates the possibility that the bees introduced elsewhere in Africa could spread to South Africa.

2.7 Harnessing genetic methods to examine floral choice

The high species diversity of both plants and pollinators in South Africa makes the traditional methodology for studying plant-pollinator interactions cumbersome and impractical in projects encompassing many different species of plants and pollinators. Additionally, the few published works in this area suggest limited expertise within this field worldwide. Another approach is therefore needed to investigate these interactions more efficiently. Genetic methods can prove advantageous in revealing the floral choice patterns of native bees in South Africa. Insect taxonomists across the country have built, and are constantly adding, to large collections of native bees sampled from all over the country. Many of these bees have pollen attached to their bodies that can be used to genetically determine the taxa of plants that they visited in the flight before they were collected. Plant species within the country are also currently being collected, identified and barcoded (Lahaye *et al*., 2008).

2.7.1 DNA barcoding

DNA barcoding has been successfully used as a diagnostic tool to identify morphologically cryptic species (by comparison to reference libraries) and has highlighted previously unrecognised species, for example various fish and amphipod crustacean species (Hebert *et al*., 2003; Ward *et al*., 2005; Witt *et al*., 2006). The genetic barcoding of a specimen involves the amplification of a DNA region that has a higher level of interspecific variation and limited intraspecific divergence. Gene regions used in barcoding should also provide a DNA target that can be easily amplified across many taxa using universal primers (Hebert *et al*., 2003; Moritz and Cicero, 2004).

Pollen DNA has been used before to identify its plant origins, often in the reconstruction of ancient plant populations (Bennett and Parducci, 2006; Parducci et al., 2013; Schnell et al., 2010). In the last two years, research publications utilising barcoding in palynology have increased (Bell *et al*., 2016; Hawkins *et al*., 2015; Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015; Sickel *et al*., 2015). The use of next-generation sequencing (NGS) in identifying mixed pollen samples (pollen metabarcoding) has recently become possible, with a full laboratory protocol and bioinformatics analysis pipeline published (Sickel *et al*., 2015).

Even though the use of DNA barcoding is a powerful method in distinguishing pollen provenance from pollinators, it is important to consider its limitations as well. The barcode gene region sequenced needs to be considered, as not all plant barcode regions have the same level of discrimination (CBOL Plant Working Group, 2009). Another limiting factor of this approach is the completeness of the sequence reference databases with which the pollen DNA is compared. When reference databases are highly unrepresentative of the flora of the specific geographic region of interest, identifications made using the DNA obtained from pollen will likely not be accurate at genus or species levels. The massive amounts of data generated by NGS also require computing power capable of handling the complex bioinformatics algorithms necessary to turn these raw data into understandable information.

2.7.1.1 Pollen as a template for genetic studies

How pollen is collected can impact the success of downstream molecular applications. When sampling pollen from hives, nests, or even honey, sufficient sample quantities are usually available for processing (Bruni *et al*., 2015; Hawkins *et al*., 2015; Keller *et al*., 2015; Sickel *et al*., 2015). Conversely, when sampling pollen directly from insect specimens, pollinators may only have few pollen grains captured on their bodies. Small sample quantities may limit and complicate all further laboratory steps. The physical structure of pollen can also be problematic when used as a template. Pollen has an extremely hardy outer wall to protect it from various environmental factors, and this wall could influence DNA extraction and other processes (Edlund, 2004). Different methods are currently used to extract DNA from pollen, but most include a step to disrupt the tough pollen exine (Hawkins *et al*., 2015; Keller *et al*., 2015; Kraaijeveld *et al*., 2015). A standardised DNA extraction method for pollen barcoding purposes would greatly aid in the comparability between studies (Bell *et al*., 2016).

Using a NGS approach for plant-pollinator interaction studies allows the collection and barcoding of pollen, even if only a few pollen grains are available. Pollinator specimens can therefore be used as pollen sources. When bees from a natural history collections are to be used as a pollen source, some factors need to be kept in mind. Bees might have extremely limited quantities of pollen captured on their bodies, and DNA extraction and all subsequent steps should be optimised for use with low starting DNA concentrations in mind. The manner in which the collection has been maintained is also of primary concern. It is well known that bees collect fungi together with pollen (Eltz *et al*., 2002), but a collection kept in suboptimal conditions would see additional fungal and bacterial growth (Merritt, 2007). Depending on the research question, barcoding gene regions can be selected to amplify more than just plant DNA from pollen samples.

2.7.1.2 Amplifying DNA from pollen samples

The barcoding principle was first applied to animal groups using the mitochondrial cytochrome c oxidase I gene (*COI*, Hebert *et al*., 2003). In plants, the mitochondrial gene variation is not as great between species as in animals thus making the use of the *COI* barcode region ineffective as a barcode within this Kingdom. Many studies have been done to search for a suite of barcode markers for use in land plants, with varied outcomes and numerous suggestions of genes to target (CBOL Plant Working Group, 2009; Kress and Erickson, 2007). The focus has mainly been on the plastid genome, with ribulose-1,5-biphosphate carboxylase oxygenase (*rbcL*) and maturase K (*matK*) being the most studied genes and at first glance the most informative. The Consortium for the Barcode of Life (CBOL) Plant Working Group was established to develop all aspects with regards to plant barcoding. They have suggested the use of *rbcL* and *matK* as the standard barcode for plants (CBOL Plant Working Group, 2009) after evaluating the success of combinations of coding regions (*matK*, *rbcL*, *rpoB*, and *rpoC1*), and non-coding regions (*atpF–atpH*, *trnH–psbA*, and *psbK–psbI*). The internal transcribed spacer (ITS) region of the nuclear genome has been suggested as an additional region to barcode, with several of the plastid genes added to increase identification success (Yao *et al*., 2010).

The plastid genome is usually uniparentally inherited, and amplifying plastid DNA could potentially present a problem when evaluating pollen from plants with only maternal plastid inheritance (Corriveau *et al*., 1990). When plastids are exclusively maternally inherited, the ITS barcode could be invaluable in identifying the pollen parent plant as pollen contains two sperm cells which contains the nuclear genome of the plant (McCue *et al*., 2011). Sometimes organellar DNA is biparentally inherited and some plastid leakage from the non-contributing parent can also occur (Nagata *et al*., 1999).

2.7.1.3 PCR and sequencing

Sequencing pollen DNA has initially been done directly from the PCR template by traditional Sanger sequencing. A study on Hawaiian *Hylaeus* bees investigated the pollen composition in the bee's gut to determine their pollen foraging behaviour (Wilson *et al*., 2010). ITS barcodes were sequenced for samples that were preserved in 100% ethanol post-collection and the 28S ribosomal RNA (rRNA) region for samples preserved in 70% ethanol. Using Sanger sequencing, most pollen samples could only be identified to one plant species, but in two samples pollen belonging to two species could be identified. For mixed pollen samples, PCR products have been cloned and subsequently Sanger-sequenced (Bruni *et al*., 2015; Galimberti *et al*., 2014). In this approach, a number of clones are picked and sequenced prior to identification against a reference database. Sequencing clones of pollen found in multiflower

honey produced identifications of between 12 and 15 taxa per sample and 38 taxa overall (Bruni *et al*., 2015), and pollen from honeybee pollen pellets collected from hives produced between 21 and 31 taxa per sample and 52 taxa overall (Galimberti *et al*., 2014). Pollen identifications were made using the Basic Local Alignment Search Tool (BLAST) available for searching GenBank, which provides a best-hit similarity method of analysis (Altschul *et al*., 1990). However, using a cloning approach to obtain single identifiable barcodes from a mixed sample with a number of unknown species, is time consuming and expensive.

NGS has made it possible to process many samples simultaneously due to the parallel nature of the technology. It is hence much more cost-effective to sequence mixed-origin samples on an NGS platform (Liu *et al*., 2012). Each PCR strand is sequenced separately in NGS and this eliminates the need for prior microscopic sorting or cloning of mixed pollen samples. Studies published recently in the pollen barcoding field have combined barcoding with NGS as the preferred sequencing method (Hawkins *et al*., 2015; Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015; Sickel *et al*., 2015). For example, metabarcoding was used to investigate the floral composition of honey samples in commercial (Valentini *et al*., 2010) and domestic beekeeper-provided honeys (Hawkins *et al*., 2015). A larger region of the same barcode as was used in metabarcoding commercial honeys (Valentini *et al*., 2010), the chloroplast *trn*L (UAA) intron region, was used to test the efficiency of NGS in identifying the plant origins of airborne pollen (Kraaijeveld *et al*., 2015). A chloroplast gene was also used as barcode in the study on beekeeper-provided honeys (Hawkins *et al*., 2015), whereas several others (Keller *et al*., 2015; Richardson *et al*., 2015; Sickel *et al*., 2015) were successful using a nuclear region for pollen identification. Different sequencing platforms have been used for pollen metabarcoding, this is summarised in Table 2.1 together with the particular study's application in the field and choice of genetic barcode marker.

Table 2.1. A comparison of recently published pollen metabarcoding studies, focusing on the application of the study, the barcode region selected and the next generation sequencing (NGS) platform used

Pollen metabarcoding application	Barcode region	NGS platform	Reference
Aeroallergen monitoring	*trn*L	Ion Torrent PGM	Kraaijeveld *et al*., 2015
Provenance monitoring	ITS2	Illumina MiSeq	Richardson *et al*., 2015
Provenance monitoring	ITS2	Roche 454 GS junior	Keller *et al*., 2015
Provenance monitoring	ITS2	Illumina MiSeq	Sickel *et al*., 2015
Food quality and provenance monitoring	*rbcL*	Roche 454 GS FLX	Hawkins *et al*., 2015
Food quality and provenance monitoring	*trn*L	Roche 454 GS 20	Valentini *et al*., 2010

Multiplexed samples need a way to be separated post-sequencing. Adding unique sequence indexes to the sequencing adapters in NGS systems allows this to be done bioinformatically (Sickel *et al*., 2015). Various indexing methods have also been successfully used, with dual indexing of PCR products by far the most cost-effective as it allows for a higher degree of multiplexing. Illumina has published a workflow for 16S metagenomics that adds overhang adapters to gene-specific PCR primers, from which dual-indexing can be done directly with the Nextera® XT (Illumina Part #15044223 Rev. B) indexing PCR. This protocol can be adapted for use in any metagenomic application, making it ideal for metabarcoding of pollen.

2.7.1.4 Bioinformatics

The incorporation of NGS in the barcoding process produces considerable amounts of data. Bioinformatic pipelines catering to the specific metabarcoding needs of pollen analysis are essential to provide reliable identifications using sequence reference databases. Sequence similarity (or best hit) approaches (Altschul *et al*., 1990) have long been in use, but suffer from some drawbacks. Heuristic searches on local alignments are performed, and a value is given of the probability that another equally good hit will be found by chance. This is not comparable to a confidence score and relates only to the local alignment, not the taxonomic assignment of the sequence (Munch *et al*., 2008). Other software available for bacterial taxonomy assignments use classifiers, such as the Ribosomal Database Project (RDP) Classifier (Wang *et al*., 2007) and the UTAX command in USEARCH (currently not

published, http://www.drive5.com/usearch/manual/utax_algo.html) and these tend to perform better than best-hit approaches. Classifiers rely on the assignment of information in a hierarchical manner to provide taxonomic classifications together with a confidence score. Incorrect assignments can still be made when classifiers are trained on incomplete or incorrect sequence reference databases. Recently, a complete bioinformatics pipeline has been published for ITS2 (Sickel *et al*., 2015) providing much needed guidance to researchers in the field. Standardised bioinformatics methods still need to be developed so that data can be easily analysed across different studies.

The reliable use of barcoding in species identification requires high-quality sequence databases that connect specific species to their DNA barcodes and that hierarchically connect these species taxonomically. This is particularly important when mixed-species pollen, such as sampled from a bee's body, is being assigned to its taxonomic origin during analysis. Additionally, a database is required for each barcode region used, with the availability of barcode sequences for these databases dependent on the usage of the DNA region within the taxon under investigation. Most sequence databases are comprised of sequences obtained from publicly available databases, such as GenBank at the National Center for Biotechnology Information (NCBI). This is not ideal, since misidentified entries could be present and often the relevant barcode markers are not available in these public databases, thereby resulting in gaps for the gene region of interest in the barcode reference database. Additionally, incomplete barcoding of plant species within a region further compounds correct identifications. Pollen samples will subsequently be mismatched to available sequences in the reference database or left unidentified. Some sequences are available in databases that undergo quality checks, such as the ITS2-Database (Keller *et al*., 2009). Plant data contained in the Barcode of Life Database Systems (BOLD; www.boldsystems.org, Ratnasingham and Hebert, 2007) are all from *rbcL* and *matK*, the two proposed plant plastid barcodes. These sequences are submitted by researchers and must conform to certain standards to be accepted. The available ITS2-Database has also recently been expanded nearly 2.5 times for plants (Sickel *et al*., 2015). Curated databases provide higher confidence in the underlying sequence data, whereas sequences in NCBI are often taxonomically misclassified, but sometimes represent the only available entry for a particular species. This could lead to the underestimation of within-species diversity due to recent speciation (Sandionigi *et al*., 2012). Bioinformatics methods applied to barcoding sequence data are consequently a crucial part of producing reliable pollen identifications.

2.8 Conclusion

Given South Africa's rich flowering plant and bee diversity, the immense economic significance of pollination for agriculture, and the threats of climate change and poor land management on the country's biodiversity, investigations into plant-pollinator relationships are vital. Floral choice in bees gives a good indication of which plants they likely pollinate. Oligolectic bees are more vulnerable to disruptions in their relationships with plants. As it has been suggested that the Succulent Karoo Biome in the western part of South Africa contains many oligolectic species, this is a key region of interest for study. Should bumblebees be introduced to this area, as suggested previously, much of the bee biodiversity of South Africa could be at stake. The identification of pollen origins is important in understanding the floral choices of bees. Many advances have been made in recent years in molecular pollen identification. DNA metabarcoding can provide accurate taxonomic identifications of pollen origins when compared to comprehensive sequence databases of carefully selected barcode gene regions. However, the lack of barcoding information for the bulk of the South African flora is a major stumbling block still to be overcome. Pollen from both honeybees and their honey, and solitary bees, has successfully been identified using this technique. DNA metabarcoding could prove instrumental in the exploration of floral choice in South African bees.

2.9 References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Bawa, K.S., 1990. Plant-pollinator interactions in tropical rain forests. Annu. Rev. Ecol. Syst. 21, 399–422. doi:10.1146/annurev.es.21.110190.002151.

Bell, K., de Vere, N., Keller, A., Richardson, R., Gous, A., Burgess, K., *et al*., 2016. Pollen DNA barcoding: Current applications and future prospects. Genome 59, 629–640. doi:10.1139/gen-2015-0200.

Bennett, K.D., Parducci, L., 2006. DNA from pollen: principles and potential. The Holocene 16, 1031–1034. doi:10.1177/0959683606069383

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., *et al.*, 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313, 351–353. doi:10.1126/science.1129551.

Bosch, J., Martín González, A.M., Rodrigo, A., Navarro, D., 2009. Plant-pollinator networks: adding the pollinator's perspective. Ecol. Lett. 12, 409–419. doi:10.1111/j.1461-0248.2009.01296.x.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., *et al*., 2015. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 170, 308–315. doi:10.1016/j.foodchem.2014.08.060.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 106, 12794–12797. doi:10.1073/pnas.0905845106.

Corriveau, J.L., Goff, L.J., Coleman, A.W., 1990. Plastid DNA is not detectable in the male gametes and pollen tubes of an angiosperm (*Antirrhinum majus*) that is maternal for plastid inheritance. Curr. Genet. 17, 439–444. doi:10.1007/BF00334525.

Dafni, A., 1992. Pollination Ecology: A Practical Approach. Oxford University Press, Oxford.

De Merxem, D.G., Borremans, B., de Jäger, M.L., Johnson, T., Jooste, M., Ros, P., Zenni, R.D., Ellis, A.G., Anderson, B., 2009. The importance of flower visitors not predicted by floral syndromes. South Afr. J. Bot. 75, 660–667. doi:10.1016/j.sajb.2009.08.002.

Department of Environmental Affairs, 2012. Land, in: 2nd South Africa Environment Outlook. A report on the state of the environment. Department of Environmental Affairs, Pretoria, pp. 87–105.

Eardley, C.D., 1989. Diversity and endemism of southern African bees. Plant Prot. News 18, 1–2.

Ebeling, A., Klein, A.-M., Tscharntke, T., 2011. Plant–flower visitor interaction webs: Temporal stability and pollinator specialization increases along an experimental plant diversity gradient. Basic Appl. Ecol. 12, 300–309. doi:10.1016/j.baae.2011.04.005.

Edlund, A.F., 2004. Pollen and stigma structure and function: The role of diversity in pollination. Plant Cell 16, S84–S97. doi:10.1105/tpc.015800.

Eltz, T., Brühl, C.A., Görke, C., 2002. Collection of mold (*Rhizopus* sp.) spores in lieu of pollen by the stingless bee *Trigona collina*. Insectes Sociaux 49, 28–30. doi:10.1007/s00040-002-8274-2.

Galimberti, A., De Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., *et al*., 2014. A DNA barcoding approach to characterize pollen collected by honeybees. PLoS ONE 9, e109363. doi:10.1371/journal.pone.0109363.

Gess, S.K., Gess, F.W., 2014. Wasps and bees in southern Africa. South African National Biodiversity Institute, Pretoria.

Gess, S.K., Gess, F.W., 2004. A comparative overview of flower visiting by non-*Apis* bees in the semi-arid to arid areas of southern Africa. J. Kans. Entomol. Soc. 77, 602–618. doi:10.2317/E7.1.

Ginsberg, H.S., 1981. Historical development of bee foraging patterns in central New York State. Psyche (Stuttg.) 88, 337–346. doi:10.1155/1981/29040.

Goka, K., 2010. Introduction to the special feature for ecological risk assessment of introduced bumblebees: Status of the European bumblebee, *Bombus terrestris*, in

Japan as a beneficial pollinator and an invasive alien species. Appl. Entomol. Zool. 45, 1–6. doi:10.1303/aez.2010.1.

Goulson, D., 2003. Effects of introduced bees of native ecosystems. Annu. Rev. Ecol. Syst. 34, 1–26. doi:10.1146/annurev.ecolsys.34.011802.132355.

Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347, 1255957. doi:10.1126/science.1255957.

Harder, L.D., 1988. Choice of individual flowers by bumble bees: Interaction of morphology, time and energy. Behaviour 104, 60–77. doi:10.1163/156853988X00601.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., *et al*., 2015. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10, e0134735. doi:10.1371/journal.pone.0134735.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218.

Hetherington-Rauth, M.C., Ramírez, S.R., 2015. Evolutionary trends and specialization in the euglossine bee–pollinated orchid genus *Gongora*. Ann. Mo. Bot. Gard. 100, 271–299. doi:10.3417/2014035.

Hingston, A.B., Marsden-Smedley, J., Driscoll, D.A., Corbett, S., Fenton, J., Anderson, R., *et al*., 2002. Extent of invasion of Tasmanian native vegetation by the exotic bumblebee *Bombus terrestris* (Apoidea: Apidae). Austral Ecol. 27, 162–172. doi:10.1046/j.1442-9993.2002.01179.x.

Human, H., Pirk, C.W.W., Crewe, R.M., Dietemann, V., 2011. The honeybee disease American foulbrood - An African perspective. Afr. Entomol. 19, 551–557. doi:10.4001/003.019.0301.

Inoue, M.N., Yokoyama, J., Washitani, I., 2008. Displacement of Japanese native bumblebees by the recently introduced *Bombus terrestris* (L.) (Hymenoptera: Apidae). J. Insect Conserv. 12, 135–146. doi:10.1007/s10841-007-9071-z.

Jaffé, R., Dietemann, V., Allsopp, M.H., Costa, C., Crewe, R.M., Dall'Olio, R., *et al*., 2010. Estimating the density of honeybee colonies across their natural range to fill the gap in pollinator decline censuses. Conserv. Biol. 24, 583–593. doi:10.1111/j.1523-1739.2009.01331.x.

Johnson, S.D., 2010. The pollination niche and its role in the diversification and maintenance of the southern African flora. Philos. Trans. R. Soc. B Biol. Sci. 365, 499–516. doi:10.1098/rstb.2009.0243.

Johnson, S.D., 1996. Pollination, adaptation and speciation models in the Cape flora of South Africa. Taxon 45, 59. doi:10.2307/1222585.

Johnson, S.D., Steiner, K.E., 2000. Generalization versus specialization in plant pollination systems. Trends Ecol. Evol. 15, 140–143. doi:10.1016/S0169-5347(99)01811-X.

Kearns, C.A., Inouye, D.W., 1997. Pollinators, flowering plants, and conservation biology. BioScience 47, 297–307. doi:10.2307/1313191.

Kearns, C.A., Inouye, D.W., Waser, N.M., 1998. Endangered mutualisms: The conservation of plant-pollinator interactions. Annu. Rev. Ecol. Syst. 29, 83–112. doi:10.1146/annurev.ecolsys.29.1.83.

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., *et al*., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 17, 558–566. doi:10.1111/plb.12251.

Keller, A., Schleicher, T., Schultz, J., Müller, T., Dandekar, T., Wolf, M., 2009. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene 430, 50–57. doi:10.1016/j.gene.2008.10.012.

Kings, S., 2015. Honeybee crisis catches SA off-guard [Internet]. Mail & Guardian. URL http://mg.co.za/article/2015-04-16-honeybee-crisis-catches-sa-off-guard (accessed 6.1.16).

Klein, A.-M., Steffan-Dewenter, I., Tscharntke, T., 2003. Fruit set of highland coffee increases with the diversity of pollinating bees. Proc. R. Soc. B Biol. Sci. 270, 955–961. doi:10.1098/rspb.2002.2306.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. doi:10.1098/rspb.2006.3721.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., *et al*., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol. Ecol. Res. 15, 8–16. doi:10.1111/1755-0998.12288.

Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE 2, e508. doi:10.1371/journal.pone.0000508.

Kuhlmann, M., 2009. Patterns of diversity, endemism and distribution of bees (Insecta: Hymenoptera: Anthophila) in southern Africa. S. Afr. J. Bot. 75, 726–738. doi:10.1016/j.sajb.2009.06.016.

Kwapong, P.K., Danquah, P.O.A., Asare, A.T., 2013. Insect floral visitors of cowpea (*Vigna unguiculata* L.). Ann. Biol. Res. 4, 12–18.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., *et al*., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. U.S.A. 105, 2923–2928.

Linsley, E.G., 1958. The ecology of solitary bees. University of California, Berkeley, California.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., *et al*., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 1–11. doi:10.1155/2012/251364.

Losey, J.E., Vaughan, M., 2006. The economic value of ecological services provided by insects. BioScience 56, 311–323. doi:10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2.

McCue, A.D., Cresti, M., Feijo, J.A., Slotkin, R.K., 2011. Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. J. Exp. Bot. 62, 1621–1631. doi:10.1093/jxb/err032.

Melián, C.J., Bascompte, J., 2002. Food web structure and habitat loss. Ecol. Lett. 5, 37–46. doi:10.1046/j.1461-0248.2002.00280.x.

Merritt, J., 2007. Mold: Prevention Of Growth In Museum Collections. Conserve O Gram 3/4, 1–5.

Michener, C.D., 2000. The bees of the world, 2nd ed. Johns Hopkins University Press, Baltimore.

Minckley, R.L., Cane, J.H., Kervin, L., 2000. Origins and ecological consequences of pollen specialization among desert bees. Proc. R. Soc. Lond. B Biol. Sci. 267, 265–271. doi:10.1098/rspb.2000.0996.

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., Gascon, C., 2011. Global biodiversity conservation: The critical role of hotspots, in: Zachos, F.E., Habel, J.C. (Eds.), Biodiversity Hotspots. Springer, Heidelberg, pp. 3–22.

Moritz, C., Cicero, C., 2004. DNA barcoding: Promise and pitfalls. PLoS Biol. 2, e354. doi:10.1371/journal.pbio.0020354.

Munch, K., Boomsma, W., Huelsenbeck, J., Willerslev, E., Nielsen, R., 2008. Statistical assignment of DNA sequences using Bayesian phylogenetics. Syst. Biol. 57, 750–757. doi:10.1080/10635150802422316.

Nagata, N., Saito, C., Sakai, A., Kuroiwa, H., Kuroiwa, T., 1999. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. Planta 209, 53–65. doi:10.1007/s004250050606.

Packer, L., Zayed, A., Grixti, J.C., Ruz, L., Owen, R.E., Vivallo, F., *et al*., 2005. Conservation genetics of potentially endangered mutualisms: Reduced levels of genetic variation in specialist versus generalist bees. Conserv. Biol. 19, 195–202. doi:10.1111/j.1523-1739.2005.00601.x.

Padyšáková, E., Bartoš, M., Tropek, R., Janeček, Š., 2013. Generalization versus specialization in pollination systems: Visitors, thieves, and pollinators of *Hypoestes aristata* (Acanthaceae). PLoS ONE 8, e59299. doi:http:dx.doi.org/10.1371/journal.pone.0059299.

Parducci, L., Matetovici, I., Fontana, S.L., Bennett, K.D., Suyama, Y., Haile, J., Kjaer, K.H., Larsen, N.K., Drouzas, A.D., Willerslev, E., 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Mol. Ecol. 22, 3511–3524. doi:10.1111/mec.12298

Pitts-Singer, T.L., Cane, J.H., 2011. The alfalfa leafcutting bee, *Megachile rotundata*: The world's most intensively managed solitary bee. Annu. Rev. Entomol. 56, 221–237. doi:10.1146/annurev-ento-120709-144836.

Rahl, M., 2008. Microscopic identification and purity determination of pollen grains, in: Jones, M.G., Lympany, P. (Eds.), Allergy Methods and Protocols. Humana Press, Totowa, NJ.

Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol. Ecol. Notes 7, 355–364. doi:10.1111/j.1471-8286.2007.01678.x.

Richardson, R.T., Lin, C.-H., Quijia, J.O., Riusech, N.S., Goodell, K., Johnson, R.M., 2015. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl. Plant Sci. 3, 1500043. doi:10.3732/apps.1500043.

Rodger, J., Donaldson, J., Eardley, C., Pauw, A., Johnson, S., 2013. SA should put bumblebees to flight [Internet]. Farmers Weekly. URL http://www.farmersweekly.co.za/article.aspx?id=40333&h=SA-should-put-bumblebees-to-flight (accessed 3.12.16).

Sandionigi, A., Galimberti, A., Labra, M., Ferri, E., Panunzi, E., De Mattia, F., *et al*., 2012. Analytical approaches for DNA barcoding data – how to find a way for plants? Plant Biosyst. 146, 805–813. doi:10.1080/11263504.2012.740084.

Schnell, I.B., Fraser, M., Willerslev, E., Gilbert, M.T.P., 2010. Characterisation of insect and plant origins using DNA extracted from small volumes of bee honey. Arthropod Plant Interact. 4, 107–116. doi:10.1007/s11829-010-9089-0

Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., *et al*., 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol. 15. doi:10.1186/s12898-015-0051-y.

Siegmund, H., 2014. Brussels Airlines cargo welcomes 16 queen bees aboard [Internet]. URL http://www.cargoforwarder.eu/2014/03/03/brussels-airlines-cargo-welcomes-16-queen-bees-aboard/ (accessed 3.12.16).

Thorp, R.W., 2000. The collection of pollen by bees. Plant Syst. Evol. 222, 211–223. doi:10.1007/BF00984103.

Torretta, J.P., Medan, D., Abrahamovich, A.H., 2006. First record of the invasive bumblebee *Bombus terrestris* (L.)(Hymenoptera, Apidae) in Argentina. Trans. Am. Entomol. Soc. 132, 285–289. doi:10.3157/0002-8320(2006)132[285:FROTIB]2.0.CO;2.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA barcoding for honey biodiversity. Diversity 2, 610–617. doi:10.3390/d2040610.

vanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95. doi:10.1016/j.jip.2009.06.011.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. doi:10.1128/AEM.00062-07.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. B Biol. Sci. 360, 1847–1857. doi:10.1098/rstb.2005.1716.

Welsford, M.R., Johnson, S.D., 2012. Solitary and social bees as pollinators of *Wahlenbergia* (Campanulaceae): single-visit effectiveness, overnight sheltering and responses to flower colour. Arthropod-Plant Interact. 6, 1–14. doi:10.1007/s11829-011-9149-0

Westerkamp, C., 1991. Honeybees are poor pollinators—why? Plant Syst. Evol. 177, 71–75. doi:10.1007/BF00937827.

Wilson, E.E., Sidhu, C.S., LeVan, K.E., Holway, D.A., 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol. Ecol. 19, 4823–4829. doi:10.1111/j.1365-294X.2010.04849.x.

Wilson, P., Stine, M., 1996. Floral constancy in bumble bees: handling efficiency or perceptual conditioning? Oecologia 106, 493–499. doi:10.1007/BF00329707.

Witt, J.D.S., Threloff, D.L., Hebert, P.D.N., 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Mol. Ecol. 15, 3073–3082. doi:10.1111/j.1365-294X.2006.02999.x.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., et al., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5, e13102. doi:10.1371/journal.pone.0013102.

CHAPTER 3: POLLEN METABARCODING FROM BEES IN A HISTORIC BEE COLLECTION IDENTIFIES PROVENANCE AND POSSIBLE MICROBIAL ASSOCIATIONS

3.1 Abstract

Pollination is a key component in both agricultural food production and ecosystem maintenance. Natural history collections containing pollinators can provide us with unique access to pollen samples collected at different spatial and temporal scales. Identification of the plant origins of pollen trapped on the bodies of pollinators in these collections can provide insight into historic plant communities and pollinators' preferred floral taxa. In this study, pollen has been sampled from *Megachile venusta* Smith bees taken from the National Collection of Insects, South Africa, spanning 93 years. Two barcode regions were sequenced on an Illumina MiSeq, namely the internal transcribed spacer 1 (ITS1) and the internal transcribed spacer 2 (ITS2). Reference sequence databases were generated by mining data for Viridiplantae from GenBank for both barcodes, and comparing the ITS2 database with a previously published ITS2 database for plants. Amplification universality of the barcode primers enabled the investigation of both pollen and some associated fungi from the pollen. Plant identification was more diverse with ITS2 than with ITS1 barcode data. Limited local plant sequence representation in reference databases resulted in higher-level taxon classifications being more confidently interpreted. Sequences that were not of plant origin were mostly assigned to fungi, especially *Malassezia*. This study successfully used pollen from bee specimens collected from as early as 1914 to obtain both the pollen and fungal metabarcodes, thereby allowing the identification of the pollen's plant origin and possible fungal contaminants linked to historic insect collections.

Keywords: *plant-pollinator interaction, ITS, insect collection, historic bee specimens, pollen identification, palynology,* Megachile

3.2 Introduction

Our daily diet contains many plant products produced as a result of pollination, such as fruits, vegetables, nuts and seed-derived commodities. This crucial ecosystem service not only ensures food on our tables, but also the diversification and maintenance of natural plant populations (Daily *et al*., 1997; Klein *et al*., 2007; Kremen *et al*., 2007). Studying the interaction between plants and their pollinators has traditionally been done by field-based observation (Johnson, 1997; Wester *et al*., 2009) and palynology (Dafni, 1992; Wilcock and Neiland, 2002) using light and electron microscopes. These methods are tedious and time-

consuming, and require experts in the fields of palynology and taxonomy to identify both the pollen and the pollinator. Similar pollen morphologies, especially from closely related taxa, further complicate plant identification by microscopic palynology (Hargreaves *et al*., 2004; Mullins and Emberlin, 1997; Williams and Kremen, 2007). These requirements have limited studies on plant-pollinator interactions for many pollinator genera, especially in species rich regions where plants and pollinators are abundant.

Taxonomic activities in the areas of entomology and botany drive pollinator and palynology related work, usually in studies conducted independently of each other. Samples are therefore often collected for taxonomically related purposes, such as species identification, distribution pattern determination or identifying new introductions. Individual specimens are labelled with descriptive collection information, including collection date, location, collector, and other relevant information (Pennisi, 2000). Flower visiting animals housed within natural history collections may have pollen on their bodies. Although flower visitors were likely not collected with the aim of utilising the pollen that was inadvertently collected along with the specimen, this pollen holds important information on the food plant of the insect visitor, the identity of a possible pollinator, and the plant community structure where the organism was collected. Additionally, a number of specimens from the same area, but from different temporal points can be selected from a collection to provide a chronological map of the area's plant and pollinator history. Historic collections may therefore provide a meaningful resource to investigate not only pollinator-plant interactions over time, but also plant communities, their diversity and distribution.

DNA barcoding allows for identification and classification of organisms based on a short nucleotide sequence (Hebert *et al*., 2003). Ideal DNA barcodes have significant interspecific genetic variation, but are flanked by conserved regions for universal primer binding to allow easy amplification of a reasonably short fragment for a wide range of taxa (Kress and Erickson, 2008). Projects are still taking place to find the optimal DNA barcode for plants (Dong *et al*., 2015; Ferri *et al*., 2015; Kress *et al*., 2015), but the ribulose-1,5-biphosphate carboxylase (*rbcL*) and maturase K (*matK*) chloroplast genes have been suggested as good candidate genes to target (CBOL Plant Working Group, 2009). Other chloroplast genes and regions have also been used successfully to barcode plants and pollen, such as *trn*L (Kraaijeveld *et al*., 2015; Valentini *et al*., 2010), *rpoC1* and *trnH-psbA* (CBOL Plant Working Group, 2009). Another accepted choice is the internal transcribed spacer 2 (ITS2) region that is found between the 5.8S and 26S rRNA genes in plants (Chen *et al*., 2010; Yao *et al*., 2010). ITS2 was also used as the DNA barcode in recent pollen barcoding studies (Keller *et al*., 2015; Richardson *et al*., 2015a, 2015b; Sickel *et al*., 2015). The internal transcribed region 1 (ITS1) was also assessed as a potential barcode together with ITS2 (Chen *et al*., 2010), but it

proved difficult to amplify and was rejected. This is in contrast to a recent study that found that ITS1 was more efficient overall at identifying plants to species level with little amplification issues (Wang *et al*., 2015). Generally, a multi-locus approach to identification yields better results due to increased discriminatory power (Burgess *et al*., 2011; CBOL Plant Working Group, 2009, and as reviewed in Bell *et al*., 2016). DNA barcoding has been successfully used to identify animals (Hajibabaei *et al*., 2006; Koch, 2010; Sheffield *et al*., 2009) and plants, including pollen (CBOL Plant Working Group, 2009; de Vere *et al*., 2012; Zhou *et al*., 2007). It has also been used to identify cryptic species (Hebert *et al*., 2004; Witt *et al*., 2006) and has recently been increasingly used to identify organisms in environmental studies (Hajibabaei *et al*., 2011; Sheffield *et al*., 2009).

Mixed origin, environmental samples, such as pollen, are characterised by the presence of DNA from different organisms that may or may not be degraded. In traditional DNA barcoding using Sanger sequencing it is necessary to separate the organisms in the sample by taxon prior to sequencing, thus obtaining a single-specimen sequence that could be assigned to a single species. Multiple organisms in a sample result in overlapping electropherograms when sequenced, thereby preventing successful species assignment. Separation of individual species is usually obtained through a cloning step, after which a number of clones are sequenced. This is extremely arduous and expensive, and in many cases not practical (Galimberti *et al*., 2014; Hajibabaei *et al*., 2011). Wilson *et al*. (2010) used Sanger sequencing to identify pollen collected from the crops of bees. Their strategy to overcome the problem of potentially having multiple plant origins per sample was to collect pollen from plants in the area, sequence these in known mixtures and then use the results to determine the level of pollen purity required to successfully identify a plant species. This provided an external measure from which they established the Basic Local Alignment Search Tool (BLAST) requirements to identify a dominant pollen species. Their strategy was limited to identifying dominant pollen represented in a sample, even though they indicated that three bees' crops contained a number of plant species. Next generation sequencing (NGS) technologies removed the need for separating species in samples by allowing high-throughput sequencing of complex DNA libraries (Liu *et al*., 2012) without prior cloning. This technology has been used in a limited number of studies that evaluated its use in the identification of plant species from mixed pollen samples (Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015a; Sickel *et al*., 2015). These studies have demonstrated that NGS DNA barcoding (metabarcoding) could improve on the identification assessments of pollen when compared to traditional microscopic identification methods employed by palynology.

In this study, we investigated the possibility of using a bee collection as a pollen source for ITS1, ITS2 and *rbcL* metabarcoding and examined the usefulness of this approach to identify

plant species from limited pollen carried by bee specimens collected over 100 years ago. Besides the pollen component, the taxonomic composition of the remainder of the environmental DNA sampled was also briefly explored.

3.3 Materials and methods

3.3.1 Pollen sample collection from bee specimens

Selecting an appropriate bee species for this study was not only dependant on the availability of the species within the collection, but also when specimens were collected. Pollen loads of the specimens also varied depending on whether the individual bee was captured on their way to, or on their way from a floral visit. We focused on an indigenous species that is widely distributed within South Africa for our initial search. Based on these criteria, *Megachile venusta* Smith (Megachilidae) specimens were selected from the South African National Collection of Insects housed at Biosystematics, Plant Protection Research: Plant Health, of the Agricultural Research Council (ARC), Pretoria, South Africa. The bees in the collection were already identified and classified prior to their incorporation into the collection by taxonomic experts using standard morphological taxonomic features for the group. The collection is mainly utilised for taxonomic classification of indigenous bee species, housing type specimens and determining local distribution patterns. The collection has been kept in a secure storage facility at a controlled temperature (22°C) since November 2010. Before then, the samples were kept at room temperature.

Megachile venusta bee specimens used for pollen sample collection were from across South Africa and covered a period of 93 years (1914 – 2007). Three bee samples containing pollen from each decade, starting from the 1910s up to the 2000s, were selected. No samples were available in the collection for the 1930s or 1950s and these decades are thus not represented here. Only 1 sample was available and included for the 1940 decade. Accession information of the bee specimens used in this study is provided in Table S3.1 (supporting information).

Pollen from the selected *M. venusta* specimens was scraped off bee abdomens with sterile micropipette tips dipped in sterilised glycerol while viewing specimens with a stereo dissection microscope (SteREO Discovery.V8 microscope, Carl Zeiss Microscopy GmbH, Jena, Germany). Care had to be taken when working with the aged, fragile bee specimens since only the exoskeletons generally remained. Scraping pollen too forcefully resulted in detachment of the metasoma from the mesosoma. Each pollen sample was transferred to a sterile 1.5 ml Eppendorf tube and crushed with the micropipette tip while still under

magnification. The micropipette tip for each sample was placed inside its respective tube after scraping off the pollen because some pollen inadvertently entered the micropipette tip during scraping.

3.3.2 DNA extraction, barcode amplification and sequencing

DNA extraction was initially optimised prior to the extraction of pollen samples from bee specimens. Both fresh pollen, arbitrarily collected from plants growing at the ARC Biotechnology Platform's grounds at Onderstepoort, Pretoria, and historical pollen on bee specimens were used as test samples for pollen extraction optimisation. The following commercial kits were tested: QIAamp DNA Micro Kit (Qiagen, Hilden, Germany), DNeasy® Plant Mini Kit (Qiagen) and Nucleospin® DNA Trace Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). All kits were used according to manufacturer's protocols to test their pollen DNA extraction capabilities. Standard fresh pollen mixtures from the same supply were used for all test extractions. Three test samples were divided into two and extracted in parallel for each kit, with one reaction subjected to 3 mm steel bead disruption using a TissueLyser II (Qiagen). The other reaction was not subjected to disruption. Pollen samples collected from six different bees were selected for DNA extraction optimisation based on sample age. Bee specimen pollen loads did not vary significantly in size when judged by eye and were therefore not taken into account in selection. Two bee specimens were selected from three different decades, respectively (1980s, 1960s, and 2000s). Pollen samples from each decade were used for DNA extraction with the DNeasy® Plant Mini Kit (Qiagen), one sample with, and one sample without bead disruption. Bead disruption was performed for 2 min at 25 Hz, with addition of lysis buffers both before, and after disruption in different samples. Direct amplification from the pollen template was also tested on fresh pollen, as previously performed by Petersen *et al*. (1996).

After optimisation, all DNA from pollen collected from *M. venusta* was extracted using the DNeasy® Plant Mini Kit (Qiagen) without bead disruption, but with micropipette tip crushing. Lysis buffer AP1 and Proteinase K (0.2 mg/ml) were added directly to the Eppendorf tubes containing the micropipette tips used for scraping pollen off the bees. Before transferral of the lysate to the QIAshredder Mini Spin Columns, the micropipette tips were carefully removed using a pair of sterile forceps and excess liquid expelled with a micropipette, in all cases cleaning equipment with 10% bleach and 70% ethanol solutions between sampling. The risk of sample contamination was mitigated by performing extractions in a clean laboratory as suggested by Willerslev *et al.* (2004). Reagents certified free of DNA or RNA were also used. The remainder of the DNA extraction was performed following the manufacturer's protocol, with the elution step using the protocol recommendations for increasing DNA yield with a

minor modification; the first eluate of 20 µl was reapplied to the DNeasy Mini Spin Column and eluted for a second time.

3.3.3 Taxonomic classification of pollen

3.3.3.1 Barcode amplification and sequencing

Three regions were targeted for DNA barcoding to identify pollen origins, namely ITS1, ITS2 and *rbcL*. The primers selected for PCR amplification were previously published for ITS1 and ITS2 (White *et al*., 1990) and *rbcL* (de Vere *et al*., 2012; Fazekas *et al*., 2008; Kress *et al*., 2005). ITS1 and ITS2 barcode sizes differ between plant taxa but have expected sizes of 100 - 700 bp (Yao *et al*., 2010), and *rbcL* barcodes are expected to be between 500 and 700 bp long (Burgess *et al*., 2011; Fazekas *et al*., 2008). Primers were modified to add overhang adapters to be compatible with the standard Illumina indexing and adapter PCR as described in the Illumina 16S Metagenomic Sequencing Library Preparation Guide (Illumina, 2013). The final modified primer sequences can be seen in Table 3.1. Two reverse primers were tested for *rbcL* due to poor amplification and sequencing results. Amplification products using primer rbcLajf634R_Tag_IL produced some sequence reads, whereas sequencing products after amplification with rbcLr506_Tag_IL did not produce usable results. All oligonucleotide primers were synthesised by Integrated DNA Technologies (IDT, Coralville, IA, USA).

Table 3.1. Primer sequences for ITS1 and ITS2 barcodes with the added Illumina adapter overhangs. Primer sequences were obtained from de Vere *et al*. (2012), Fazekas *et al*. (2008), Kress *et al*. (2005) and White *et al*. (1990). Illumina nucleotide sequences (indicated in bold and underlined) were used in accordance to the workflow from the Illumina 16S Metagenomics protocol (Illumina, 2013). These tags allow Nextera indexing and Illumina adapter addition through PCR.

Barcode region	Primer name	Primer sequence (5' to 3')
ITS1	ITS5F_Tag_IL	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** GGA AGT AAA AGT CGT AAC AAG
	ITS2R_Tag_IL	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**GC TGC GTT CTT CAT CGA TGC
ITS2	ITS3F_Tag_IL	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** GCA TCG ATG AAG AAC GCA GC
	ITS4R_Tag_IL	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**TC CTC CGC TTA TTG ATA TGC
rbcL	rbcLF_Tag_IL	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** ATG TCA CCA CAA ACA GAG ACT
	rbcLajf634R_Tag_IL	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**GA AAC GGT CTC TCC AAC GCA T
	rbcLr506_Tag_IL	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**AG GGG ACG ACC ATA CTT GTT CA

Barcode amplification reactions consisted of a final concentration of 0.5 µM of each primer (initial concentration of 10 µM each), 200 µM of dNTPs, 1× Phusion® High-Fidelity Buffer, 0.02 U/µl Phusion® High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, MA, USA) and 5 µl of DNA template, irrespective of DNA concentration. Reaction volumes were adjusted to a final reaction volume of 50 µl with Milli-Q® H2O (Merck Millipore, KGaA, Darmstadt, Germany). All amplification reactions commenced with a denaturation step at 98°C for 3 minutes, followed by 30 cycles of denaturation (98°C for 7 seconds), primer annealing (65°C for 30 seconds) and extension (72°C for 30 seconds). Reactions were concluded with a final elongation step at 72°C for 10 minutes.

Amplified barcodes were visualised using 2% agarose gel electrophoresis, and were purified with the QIAamp® MinElute™ PCR Purification Kit (Qiagen), according to the manufacturer's protocol. Reapplication of the eluate to the MinElute Spin Column was performed to increase DNA concentration. DNA quantification was done using a Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA) and the Qubit® dsDNA

High Sensitivity Assay Kit (Invitrogen, Life Technologies), both according to the protocols provided by the manufacturer.

Sequencing libraries were prepared according to the preparation protocol (Illumina, 2013). Nextera XT (Illumina, Inc. San Diego, CA, USA) indexes were used to multiplex the individual samples and barcodes. Sequencing of a single multiplexed sample was performed using the MiSeq Reagent Kit v3 (2 × 300 bp paired end, Illumina, Inc.) on a MiSeq desktop sequencer (Illumina, Inc.) at the ARC's Biotechnology Platform, Pretoria, South Africa.

3.3.3.2 Bioinformatics and sample analyses

The MiSeq sequencer performed all primary analysis, such as image analysis and base calling. Sample demultiplexing was done using MiSeq Reporter v2.5.1 by separating the samples on perfect index matches (Illumina, 2014). Quality and adapter trimming of reads was done using Trimmomatic 0.33 (Bolger *et al*., 2014). Reads that passed quality trimming were merged in MacQiime 1.9.1-20150604 (Caporaso *et al*., 2010).

A curated, plant sequence database for ITS1 was not available at the time of data analyses. A reference database was therefore constructed by downloading all Viridiplantae sequences for ITS from GenBank (Benson *et al*., 2015, accessed 12 May 2015) using custom Python scripts. A Hidden Markov modeller, ITSx 1.0.11 (Bengtsson-Palme *et al*., 2013), was used to detect ITS1 and ITS2 in the complete ITS sequences downloaded from GenBank, and to exclude any sequences detected as anything other than bryophytes, chlorophytes, marchantiophytes and tracheophytes. Pollen samples were taxonomically classified by plant origin against this downloaded reference sequence database using the "rdp" option in *assign_taxonomy.py* in MacQiime 1.9.1-20150604 (Caporaso *et al*., 2010), which used RDP Classifier 2.2 (Wang *et al*., 2007). A minimum confidence level of 80% (c = 0.80) was chosen in order to assign a pollen DNA sequence to species level using the supplied reference database of plant sequences. ITS2 sequence data were analysed with the sequence database generated in this paper (hereafter referred to as gITS2), as well as compared to the annotated and curated ITS2 database created by Sickel *et al.* (2015) using their published bioinformatics pipeline. The latter workflow and database are hereafter referred to as sITS2. Sequence reads not classifying to plant taxa, were uploaded to the MG-RAST server v3.6 (Meyer *et al*., 2008) for standard analysis. Taxa represented in a proportion less than 0.1% of the total number of reads per sample were discarded (Sickel *et al*., 2015).

Only reads identified as plant taxa were used to determine species richness. To determine whether samples were sequenced to a sufficient depth to identify all possible plant taxa,

rarefaction curves were drawn using the vegan v. 2.3.4 package in R v. 3.2.4 (R Core Team, 2016). Taxonomic assignments for each sample were checked for their presence in the area in which bees were originally sampled using the local Plants of southern Africa (POSA) v. 3.0 database (www.posa.sanbi.org, accessed 10 Feb 2016). Spearman rank-order correlations were performed in R v 3.2.4 (R Core Team, 2016).

3.4 Results

3.4.1 Pollen DNA extraction from historic bee specimens

DNA from 22 pollen samples was obtained from a historic bee collection and was successfully extracted, although suitable specimens with sufficient pollen loads limited sampling. Available pollen varied between specimens and collection dates. The manner of specimen collection, netting and malaise traps for example, and the subsequent handling may have inadvertently contributed to pollen losses and further limit suitable specimens. In the case of *M. venusta*, approximately 25% of the total specimens within the collection had sufficient pollen for sampling, of which a subset was used here. However, the large number of specimens in the collection ensured sufficient material.

Several protocols were tested to find the optimal method for DNA extraction from pollen samples taken fresh and from a historic bee collection. All pollen DNA concentrations post-extraction were lower than the accurate quantifiable range of the Qubit® assays. This was expected of the small volumes of pollen obtained from bee specimens. PCR results were consequently used as the measure of success of extractions. The only extraction protocol that resulted in consistent PCR amplification was the DNeasy® Plant Mini Kit (Qiagen), without bead disruption. Other DNA extraction methods produced PCR results that were too inconsistent for developing a NGS protocol for barcoding insect-derived pollen samples.

3.4.2 Pollen DNA high-throughput sequencing

A total number of 660,837 high-quality merged reads were obtained for ITS1 and 1,130,803 for ITS2 after quality, adapter and length trimming across 22 pollen samples. This was on average 30,038 reads for ITS1 and 51,400 reads for ITS2 per sample, respectively. Due to consistently poor amplification results, only 40,646 reads with poor quality were obtained for *rbcL* in total after 2 sequencing runs, with a mean of less than 2,000 reads per sample. Mean read lengths for both forward and reverse reads for *rbcL* were very close to the range of our length quality cut-off (forward mean length = 142 bp; reverse mean length = 103 bp). One sample failed to produce any reads, and less than 200 reads were obtained for four samples.

The number of reads obtained per sample for rbcL was significantly lower than for ITS1 ($t = 4.48$, $p < 0.001$) and ITS2 ($t = 4.03$, $p < 0.001$). Table 3.2 provides summary statistics for ITS1, ITS2 and *rbcL* processed reads

Table 3.2. Summary of processed reads for ITS1, ITS2 and *rbcL* after next generation sequencing. Numbers indicated are the remaing number of reads after processing for quality with Q20 filtering, Nextera adapter trimming, fragments discarded that were less than 100 bp in length and forward and reverse reads merged.

	ITS1	**ITS2**	***rbcL***
Sum of total combined reads	660,837	1,130,803	40,646
Mean of total combined reads	30,038	51,400	1,936
Median of combined reads	20,135	24,668	1,570
Standard deviation	27,246	56,826	1,427

The percentages of reads of both ITS1 and ITS2 assigned only to the kingdom Viridiplantae varied between samples. Sequence reference databases were based on this kingdom, and all samples consisting of less than 1,000 reads that were identified to plant species level were regarded as unsuccessful and were discarded prior to further analyses. The presence of unidentified reads did not influence the identification of plant origins of samples, even though a higher amount of total reads were necessary to reach sequence saturation for plant identification. Identification of rbcL reads to plant origins produced very variable results. In 45% of the samples less than 1,000 reads were produced. Due to the extremely variable nature of amplification and sequencing results, *rbcL* data were not analysed further.

After removal of taxa representing less than 0.1% of reads per sample, only one or two plant genera per sample for ITS1 could be identified against the sequence reference database generated in this study. Between one and eight plant species were identified per sample using sITS2. Rarefaction curves show that the sequencing depth for all samples was sufficient to obtain maximum taxon richness (Figure 3.1). When all raw read data are included in rarefaction analyses, a maximum of ten species per sample for ITS2 was reached, with curves still reaching a plateau (supplementary Figure S3.1), thus further supporting sufficient sequencing depth was reached.

A total of 81.8% of ITS1 samples remained that had more than 1,000 reads identified to Viridiplantae. One of the ITS2 samples only had 1,154 high quality, merged reads that could be identified to Viridiplantae, but were still sequenced to saturation, as indicated by the rarefaction curve (Figure 3.1b). A single ITS2 sample had less than 1,000 reads identified to Viridiplantae and was removed prior to further analyses, with 95.5% of samples remaining for further analyses. Only five (22.7%) of the ITS2 samples identified using the gITS2 database contained more than 1,000 reads that could be confidently identified to plant genus level. A rarefaction curve was consequently not created for gITS2 identifications, and the remainder of the analyses were performed only on sITS2 classifications. Only one or two plant genera could be distinguished per sample using the gITS2 database.

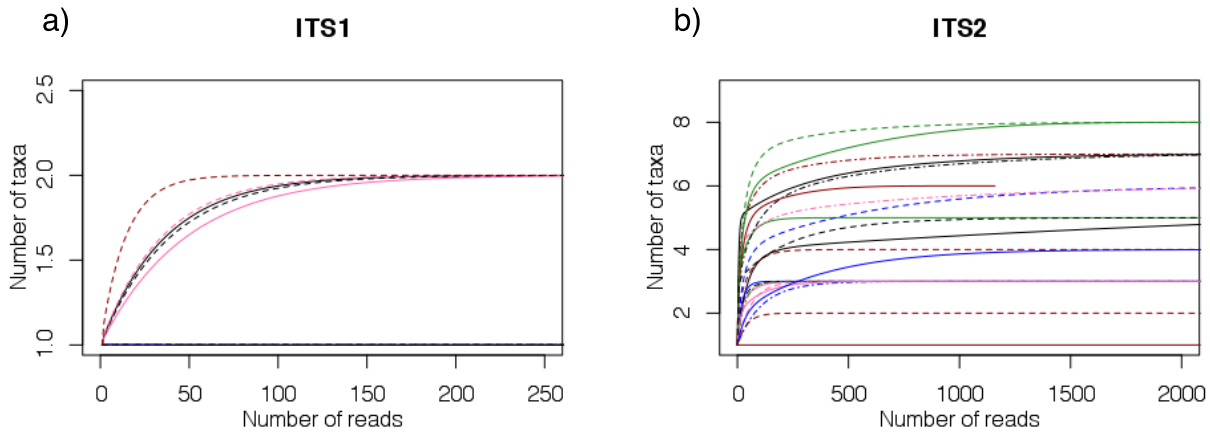

Figure 3.1. Rarefaction curves for a) ITS1 and b) ITS2 samples. ITS1 samples reached sequence saturation at approximately 250 reads, whereas ITS2 samples needed approximately 1,000 to 2,000 high quality sequence reads to obtain maximum plant taxon richness per sample. Rarefaction curves were created after taxa representing less than 0.1% of reads per sample were removed.

3.4.3 Plant origins of pollen collected from Megachile venusta *specimens*

When identifying sequence reads to the ITS1 database, two plant genera (*Helianthus* and *Oryza*) were identified. Only *Helianthus* was identified in 72.2% of the samples, and both genera were identified in the remaining 27.8% of samples. On average 3.3% (SD = 0.25) of reads per sample could only be assigned to the phylum level (Streptophyta) and 50.3% (SD = 0.09) of reads remained unidentified at the assignment level of kingdom. Classification to species level was not possible with the ITS1 database.

ITS2 read classifications varied distinctly between the two different sequence databases used. Using the gITS2 database, most reads of the five samples included in the analyses (mean =

74.8%, SD = 0.06) could only be classified to the Streptophyta phylum level. Between 2.2% and 17.7% (mean = 9.0%, SD = 0.19) of reads per sample remained unidentified (kingdom Viridiplantae). Up to two genera were identified (*Helianthus* and *Amaranthus*) per sample with the gITS2 database. These genera originated from two different families and orders, but were from the same class. Species level classifications could not confidently be made with the gITS2 database.

Identification with the sITS2 database produced identification only up to kingdom level in 0.6% of the reads per sample, on average (SD = 0.02) and only up to phylum level for an average of 68.4% (SD = 0.22) of reads per sample. Significantly more lower ranking taxon identifications could be made using the sITS2 database. With the confidence set at the recommended level of 80%, an average of four species, four genera and four families were identified per sample when classifying reads with the sITS2 database. In total, 25 species from 21 different genera could be confidently identified with the sITS2 database. These species belonged to 19 different families, 16 orders and six classes. The five most dominant plant species identified overall were *Pteris vittata* (34.6%), *Helianthus annuus* (32.4%), *Astragalus membranaceus* (17.2%), *Magnolia kwangtungensis* (3.3%), and *Macrothamnium leptohymenioides* (3.2%). Two algae species, *Caulerpa webbiana* and *Pirula salina*, were identified in one, and three samples, respectively. Due to the inherent biases that PCR amplification presents, abundance data in metabarcoding should be interpreted with care. A summary of all species identified and their respective abundances can be seen in Figure 3.2. However, several sequence reads could not be classified confidently to species level with the sITS2 database. Eight taxa could only be classified to genus level, five to family level, two to class level, and another one to order level. Of the eight genera identified, three correspond to prior species level classifications, with five genera newly identified (Table 3.3). Both *Amaranthus* and *Helianthus* were also recognised by the gITS2 database.

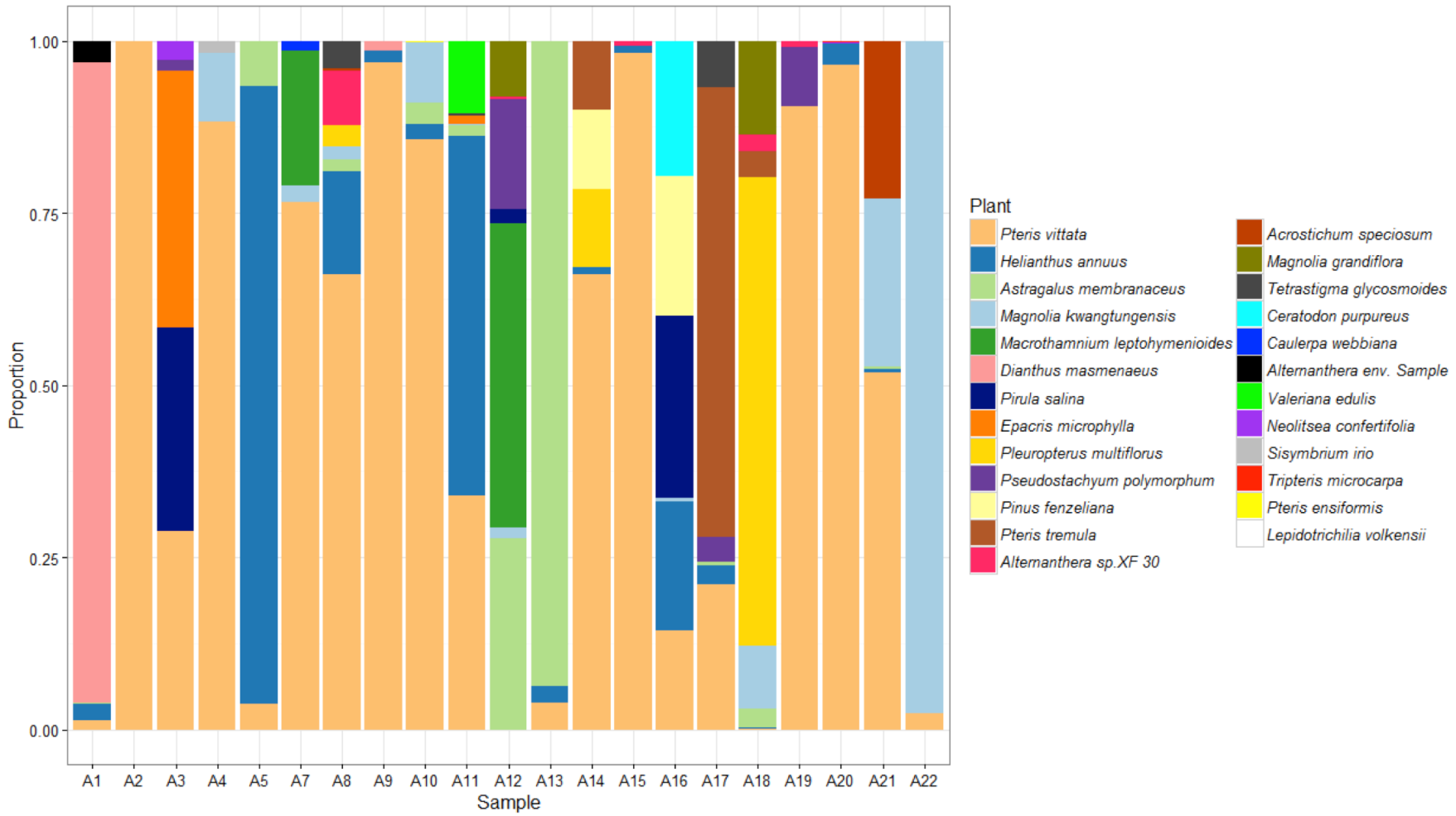

Figure 3.2. Bar graph representing 22 samples classified with the sITS2 database. Samples are arranged from the oldest on the left (A1 sampled in 1914) to the newest on the right (A22 sampled in 2007). The different colours indicate different plant species identified with the sequence database, and in which percentage it was detected in each sample.

Table 3.3. Viridiplantae taxa that were not classified to species-level, but to genus and family-level. Two classes were also identified, Liliopsida and Magnoliopsida, as well as the order Cucurbitales.

Family	**Genus**
Asteraceae[#]*	*Helianthus**
	Lactuca
Amaranthaceae[#]*	*Amaranthus*
	*Alternanthera**
Magnoliaceae	*Magnolia**
Proteaceae	*Macadamia*
Moraceae	*Morus*
Trebouxiaphyceae	*Trebouxia*
Cucurbitaceae[#]	-
Fabaceae[#]*	-
Poaceae[#]*	-

[#] Five taxa identified for which sequence reads could be classified up to family level. Four of these families ([#]*) have been identified during species-level classification.
* Taxa also identified during species-level classification.

From the different taxa distinguished in the pollen from *M. venusta* samples, data for 15 genera, and for only six species were available in the POSA v. 3.0 database. The available plant distribution data overlap well with the geographic origins of the bee samples. For 86.4% of samples, all identified genera occurred within the area where *M. venusta* was sampled and from which pollen was sequenced. Four genera (*Magnolia, Helianthus*, *Astragalus*, and *Acrostichum*) in three samples did not have representatives in the POSA v 3.0 database originating from the Northern Cape.

Combined sITS2 classification results of all samples from *M. venusta* specimens provide insight into the floral choice of the bee species. When adding how many times a plant species was identified across all samples, the most commonly collected plant species was *P. vittata*, followed by *Helianthus annuus*, *Magnolia kwangtungensis* and *Astragalus membranaceus* (Figure 3.3).

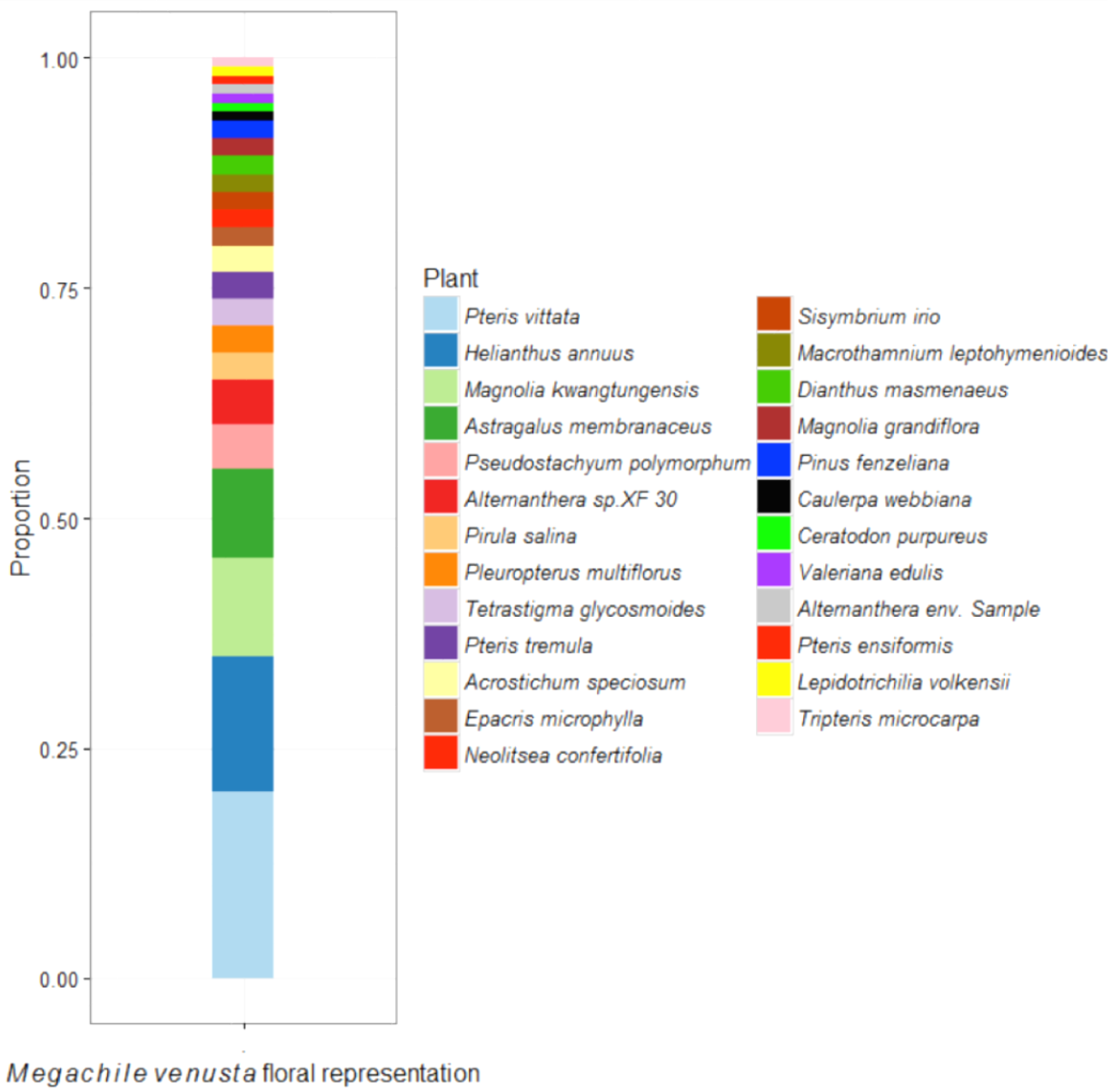

Figure 3.3. Floral representation of pollen sampled from all *M. venusta* bee specimens. *P. vittata, H. annuus, M. kwangtungensis* and *A. membranaceus* are the most highly represented plant species from pollen of *M. venusta* specimen.

3.4.4 Fungal and other contamination associated with specimen-collected pollen samples

The focus of this study was specifically on the utilisation of historic collections to determine the plant origins of pollen sampled from collection specimens. As such, only a brief overview will be given on the fungal contamination in pollen samples. A large number of reads in each sample remained unclassified during the classification with plant reference databases, regardless of which sequence database was used in the analysis. As with plant taxon classification, rare taxa were excluded from the analysis of previously unclassified reads.

Unclassified reads consisted mainly of fungi. The most abundant fungal genus was *Malassezia*, and it was the only one present in all samples of ITS1 (49.9% of total reads), although only detected in six samples of sITS2 (4.5% of total reads). *Cladosporium* was present in all but two of the sITS2 samples (23.7% of total reads), but only in negligible proportions in three samples of ITS1. Several plant pathogenic fungal genera were detected in varying proportions, such as *Alternaria*, *Cladosporium*, *Fusarium*, *Kabatiella*, *Mucor*, *Myrothecium*, *Penicillium*, *Peniophora*, *Peyronellaea*, *Phanerochaete*, *Phoma*, *Ulocladium*, and *Ustilago*. As discussed previously, the large extent of fungal and other contamination present in the pollen samples did not have an adverse effect on the classification of plant taxa, provided that a sufficient sequencing depth was maintained. Cross-contamination of pollen samples within a tray was not assessed.

3.4.5 Effect of sample age on sequence success and taxon identification

Pollen from as far back as 1914 was metabarcoded in this study. A Spearman's correlation was run to determine the relationship between the proportion of unidentifiable reads in 22 pollen samples and the year from which they originated for both ITS1 and sITS2. No correlation was found for either ITS1 (r_s = -0.17, n = 22, p = 0.45) or sITS2 (r_s = -0.05, n = 22, p = 0.81). Therefore, no relationship was found between the age of pollen samples and the proportion of reads that could be assigned to plant species. Since the majority of unidentified reads were found to be fungi, the age of the sample also had no significant correlation with the percentage fungus present in the pollen load.

3.5 Discussion

Historic specimen collections can essentially be seen as large, untapped resources of genetic data, especially for plant-pollinator interaction investigations. Where pollination is concerned, honeybee-related studies far outweigh studies of other bees. It is nonetheless important to focus attention to the lesser-studied plant-pollinator interactions as this can hold key information about both plant and bee communities. As pollinator specimens in historic collections usually have accompanying metadata, these specimens are invaluable to researchers interested in pollination and change in ecosystems over time. Sampling pollen directly from pollinators in collections allows for retracing pollinator interactions over time, and thus recreating plant communities in both time and space, as the sample’s place of collection was documented.

DNA barcoding has been used for years to successfully identify unknown plants (Burgess *et al*., 2011), and the recent uptake of pollen sequencing in the metabarcoding community

(Bruni *et al*., 2015; Galimberti *et al*., 2014; Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015b) has sparked new interest in this field. In this study, DNA metabarcoding was used to determine the plant origins of limited pollen sampled directly from *M. venusta* bees taken from a historic collection. The pollen exine is exceptionally resilient ensuring DNA contained within the pollen grain maintains its integrity for very long time periods, making ancient pollen studies possible (Parducci *et al*., 2005). Different DNA extraction methods were tested here to determine which one would provide the most reliable barcoding results from the limited starting material collected from bee specimens. The DNeasy® Plant kit (Qiagen) was found to provide DNA templates that allowed repetitive and consistent barcode amplification. The same kit was also used for DNA extraction from larger pollen samples in other metabarcoding studies (Bruni *et al*., 2015; Galimberti *et al*., 2014; Hawkins *et al*., 2015; Richardson *et al*., 2015a, 2015b), thereby providing an abundance of material for subsequent laboratory workflows. Differing methods of macerating the pollen exine were performed in extractions previously performed, except in Bruni *et al*. (2015) where the pollen was extracted directly after isolation from honey. Adding a maceration step to any of the extraction protocols tested here appeared to either destroy the entire sample, or dilute the limited starting material to such low levels that PCR was unsuccessful or highly variable. Amplification of DNA directly from the pollen template failed, likely due to unsuccessful disruption of the pollen exine and unavailability of the DNA during amplification. Proteinase K was added to the extraction kit's lysis buffer to optimise the breakdown of the pollen exine in the absence of physical maceration. This step is also incorporated in the Macherey-Nagel NucleoSpin® Food Kit used by Keller *et al*. (2015) and Sickel *et al*. (2015), and was also included in the extraction of pollen from honey by Hawkins *et al*. (2015). It is important to keep in mind that since pollen is only macerated with a pipette tip in the proposed extraction method, it is possible that some bias could be introduced during this step. This bias could be introduced since radical differences in pollen exine structures between taxa exists (Edlund, 2004), and it is possible that DNA from some pollen taxa was more easily accessible during extraction, leading to a bias in subsequent steps. Although it was not the preferred method in this study due to the small sample sizes, it is advisable to include a pollen maceration step such as bead beating when initially selecting a DNA extraction method for historic samples to decrease chances of bias. Sufficient PCR amplification was obtained with this extraction method using ITS1 and ITS2 primers on these collection-based pollen samples to continue with sequencing.

The choice of DNA barcode to identify plant origins from historic collections is important, not only due to its ease of amplification from older samples, but also due to the availability of reference barcode databases to identify species. Since the starting material is very limited, there is not much room for optimisation of procedures across multiple barcodes or even

within a single barcode. Sequencing results indicate that historic pollen identification using DNA barcoding on an NGS platform was successful for ITS1 and ITS2, regardless of having very limited starting material. In this study ITS1 and ITS2 amplified equally well across samples but *rbcL* produced variable amplification results. Two different reverse primers were tested for this gene and only amplicons produced with rbcLajf634F_Tag_IL produced sequence results. However, low clustering on the MiSeq flow cells occurred during both sequence runs that did produce results, with variability between samples. *RbcL* has been successfully used before in pollen barcoding with both Sanger sequencing (Bruni *et al*., 2015; Galimberti *et al*., 2014) and with NGS (Richardson *et al*., 2015a), however, the amounts of pollen used for DNA extraction in these cases were notably more (50 mg in Richardson *et al*., 2015a and 100 mg in Galimberti *et al*., 2014) than sampled in this study from a single bee specimen. Although there are studies that have shown success by using plastid DNA in pollen barcoding (Hawkins *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015a), it remains important to remember that plastids are maternally inherited in many floral taxa and that they will consequently not be present in all pollen grains (Bennett and Parducci, 2006; Corriveau *et al*., 1990). This poses a false negative versus true negative PCR amplification problem for a barcoding approach, with uncertainty about what causes the amplification failure. The low plant species diversity identified with ITS1 and ITS2 barcodes could also explain the low amplification and sequencing success of *rbcL* from pollen samples. Many of the samples had only identified a few plants. This could mean that plant families remained genetically undetected due to pollen not containing plastid DNA, even though they were represented in the pollen load of the insect. *RbcL* amplicon lengths were also expected to be longer than ITS1 or ITS2 amplicons, and these longer amplicons could potentially not have been obtained due to DNA degradation (Willerslev *et al.*, 2004), subsequently resulting in poor sequencing results. Due to the variable amplification and sequencing results obtained with the *rbcL* amplicons, this gene does not appear to be a good choice for an affordable, reliable metabarcode workflow for pollen sampled from bee specimens in a natural history collection.

A dual-indexing strategy was used to lower costs, reduce error-introducing steps and save time. Tagmentation kits (Richardson *et al*., 2015a, 2015b) and single-index systems (Hawkins *et al*., 2015; Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Valentini *et al*., 2010) have so far been the norm in pollen metabarcoding. Cornman *et al*. (2015) recently also used Illumina's 16S Metagenomic Library Preparation method to perform the dual-indexing for their samples, whereas Sickel *et al*. (2015) used a similar method developed by Kozich *et al*. (2013). Dual-indexing for pollen metabarcoding is particularly useful as the sequencing depth required may vary between samples depending on the number of pollen grains and number of taxa present per sample (Bell *et al*., 2016). As indexing is the most expensive part of the NGS process,

driving sequencing cost down at the indexing stage allows for more freedom where sequence depth is concerned.

For plant origin tracing of pollen, rarefaction curves indicate that enough reads were sequenced for ITS1 and sITS2 samples to reach sequence saturation. Only approximately 250 reads per sample were necessary to reach a plateau during rarefaction for ITS1 since only two species were identified using this barcode. For sITS2, between 750 and 2000 reads were necessary to reach sequence saturation, with the upper limit consistent with previous findings in pollen metabarcoding (Sickel *et al*., 2015). Less taxa overall were identified during this study than in the aforementioned, and the suggested number of reads to be sequenced (2000 – 3000) in that study would have been more than enough. When interest in the pollen sample is wider than simply its plant origin, such as investigating the pollen microbiome, many more reads are needed to obtain sequence saturation for plant identification, which increased research cost but did not limit the application of the method. However, in combination with dual indexing and low per-sample read requirements, a NGS approach provided a highly cost efficient pollen metabarcoding strategy to screen large pollen sample numbers from historical specimen collections.

Sequencing results from pollen were not compared with microscopic methods, as suggested by Keller *et al*. (2015) and Kraaijeveld *et al*. (2015). This was mainly because of the extremely small volume of pollen present on most of the bees sampled and the absence of an available palynologist. Primer biases can occur, and ideally triplicate PCR reactions should be performed to overcome this issue (as performed in Keller *et al*., 2015 and Sickel *et al*., 2015). Triplicate PCR reactions posed a problem here, as a very small volume of DNA was extracted and the DNA concentration remained extremely low. Two recent studies had found that microscopic and metabarcoding results correspond significantly (Keller *et al*., 2015; Kraaijeveld *et al*., 2015), but another showed no correlation between them (Richardson *et al*., 2015b). Any proportional plant identification data should therefore be interpreted carefully. In this study, PCR bias was likely introduced due to the amplification cycles used to obtain results, and proportional results were interpreted with this in mind. Increased sample numbers, obtained from historic specimen collections could, however, provide the statistical support required to identify a pollinator's floral interactions. Another means of interpreting sequence data without relying on read percentages per sample is to combine pollen data from all samples for a particular bee species. By counting how many times a certain taxon was observed across all samples, normalisation within the bee species occured and more confident conclusions could be drawn about the floral choice of this bee. Including mock community analysis of known pollen mixtures is an important future step that was not feasible in this study due to a lack of age- and site-appropriate pollen samples for comparison. The removal

of rare taxa in each pollen sample was aimed to overcome this limitation to some degree (Brown *et al*., 2015).

The lower number of plant taxa identified per sample in this study is concordant with the observed floral constancy behaviour in foraging bees (Michener, 2000), where bees tend to visit flowers from plants of the same taxa during one foraging trip as long as this resource remains available. Pollen in this study was sampled directly from bee specimens that were actively foraging during their capture, and therefore the lower number of plant taxa obtained, was expected. Sampling pollen from pollen traps (Keller *et al*., 2015; Richardson *et al*., 2015a) or honey (Bruni *et al*., 2015; Hawkins *et al*., 2015) would yield considerably more plant taxa after metabarcoding as these pollen samples originated from multiple bees and cover many foraging trips. Indeed, the plant families identified here with metabarcoding corresponded with the foraging information known for *M. venusta*, with the Asteraceae, Fabaceae, and Poaceae associations observed for specimens both in South African and collections abroad (Eardley, 2013).

All plant families discovered were present in the localities where bee samples were collected when compared with the POSA v3.0 database. The absence of four genera in the database for the Northern Cape Province is potentially explained by sampling efforts in different regions. As an example, botanists would typically sample intensively in the area where they are based or have long-term studies, and also more so on their plants of interest, leaving gaps in the record in other plant taxa. *Helianthus annuus* has been observed to grow in fields in the Northern Cape by the authors, for example, but was absent from the database, probably because this crop is not native to the region. The POSA database was last updated in 2012, limiting its usefulness. Only six species identified with the sITS2 database were represented in POSA. Once again, this could be due to poor species representation in the plant database, or sequence misclassification due to limited representation in the sequence reference database, allowing closely related sequences to be assigned with high enough confidence even though it does not represent the true plant origin. Some of the species identified are not native to South Africa, such as *Magnolia* spp., and *Pseudostachyum polymorphum*. A simple internet search for *Magnolia* in South Africa, however, reveals that particularly *Magnolia grandiflora* is readily traded.

Species of interest were *Pteris vittata* and *Pteris ensiformis*, a genus of fern that was present in all sITS2 samples. Ferns do not produce pollen and are unknown to have any animal involvement in reproduction. Ferns usually produce large numbers of spores that are easily dispersed into the environment. Initial thoughts were that these identifications could be due to environmental spore contamination. However, alternative explanations for the detection of

ferns are that sequences could also have been misclassified or the sequences representing this genus in the underlying database could potentially have been incorrectly assigned. When constructing and testing the gITS2 database, similarly high proportions of *Pteris* classifications were present. However, upon investigation of the underlying entries in GenBank, it was found that five entries from the same batch are identical to fungal sequences, which could indicate that these samples were misrepresented in NCBI. When those sequences were removed from the database, no more identifications of *Pteris* occurred. This demonstrates the importance of the quality of the reference sequence database. The same entries are also present in the sITS2 database but were not removed prior to classification, as a retraining of the classifier would have needed to be done. This would alter the Sickel *et al*. (2015) database and have defeated the purpose of a comparison between ITS2 databases.

The identification from a reference database will also only occur if the specific species was barcoded before, correctly classified and phylogenetically assigned. The International Barcode of Life (iBOL, www.ibol.org) project aims to achieve this. Currently there are only approximately 7,100 entries for the internal transcribed spacer regions on GenBank for South African flora, when compared to the almost 275,000 total entries. This means that any sequence database based on data drawn from NCBI will only include approximately 2.6% of local sequence data. This is unquestionably a limiting factor for pollen barcoding interpretation in South Africa. Misclassifications of sequences will occur more frequently as the more likely candidates are possibly not represented in the reference database. Species-level interpretation should subsequently be approached with extreme caution. More confidence can be placed in higher-level classifications, with family-level interpretation most likely being accurate. ITS1 and gITS2 databases were based solely on sequence data sourced from GenBank, and consequently does not include many local entries. Also, data submitted to GenBank are not always reliable (Harris, 2003). The sITS2 database created by Sickel *et al*. (2015) resulted in more identifications when compared to the gITS2 database, and this is likely due to their database being subjected to quality control measures, such as structure validation (Keller *et al*., 2015; Sickel *et al*., 2015). Sequence databases were used with a Bayesian classifier that implements hierarchical decision-making (Wang *et al*., 2007), and this provides further confidence in taxon assignments. Simple BLAST searches, for instance, may only rely on local alignments that could potentially provide information on only a part of the underlying sequence. The underlying taxonomic information may also be incorrect or incomplete. BLAST also does not provide any measure of confidence in identifications, as multiple sequences could have the same similarity score when locally aligned.These identifications should be interpreted with extreme caution and should rather be avoided.

DNA degradation in pollen samples could lead to shorter amplification products, the introduction of base modifications, and non-amplification (Willerslev *et al.*, 2004). Sequencing errors due to base modifications can lead to incorrect taxon identification. Also, some taxa with longer ITS1 or ITS2 regions may selectively not amplify if pollen show high levels of DNA degradation. This would lead to unamplified taxa not being sequenced and not being detected in metabarcoding results. In such cases, it is important to combine analysis with observational studies and microscopic identification of pollen.

Specimens housed in historic collections can be retained for hundreds of years, and it is therefore important that specimen storage conditions are optimal to prevent fungal and microbial growth that could adversely affect the collection (Merritt, 2007). Many fungal taxa are naturally associated with plants and bees inadvertently collect fungal spores together with pollen or, sometimes, instead of pollen (Shaw, 1999). The fungal microbiome of pollen sampled from historic collection specimens can also be studied using the primers chosen in this study. The ITS5-ITS2 primers for the ITS1 region, and ITS3-ITS4 primers for ITS2 are not fungi-specific, and amplify both plant and fungal DNA well with a three-nucleotide mismatch (Bellemain *et al*., 2010). Sequencing results show highly variable amounts of microbiological (both fungal and bacterial) material between ITS1 and ITS2, as well as between samples within a barcoding region. The primers used to amplify ITS1 appear to be more universal at a three-nucleotide mismatch, in terms of amplifying plant species. Using this parameter, ITS5-ITS2 primers amplified 12,100 plant sequences vs 9,293 fungal sequences, whereas ITS3-ITS4 primers amplified only 8,852 plant sequences versus 22,078 fungal sequences *in silico* (Bellemain *et al*., 2010). It makes sense then that, on average, less unidentified reads were present in ITS1 (53.6%) than in sITS2 (69%). No relationship was found between the proportion of fungi per sample and the age of the sample. This could suggest that fungal growth did not increase in the collection significantly over the last 102 years, despite the fact that storage conditions had not been standardised for the samples analysed in this study. Temperature of the storage unit was only controlled at 22°C from 2010, and the most recent sample analysed here was dated 2007. It should be kept in mind that the sample size is small, and a positive correlation with increasing age could be possible if more data were available. Not only the age of the samples, but also the preservation of DNA and extraction efficiency could play a role in the success of amplification and metabarcoding of pollen.

3.6 Conclusion

Pollen metabarcoding of historic collections opens up the possibility to reconstruct the plant communities that pollinators visited in the past. By doing this, changes in their floral choices

can be tracked both temporally and spatially, giving insight in how different environmental factors affect them. Understanding the influences of factors such as climate change and land use change on plant-pollinator interactions could prove vital in the conservation of vulnerable species, both plant and animal. Pollen sampled from historic *M. venusta* bees, dating back 102 years, was successfully used for DNA extraction, amplification and sequencing on an NGS platform. This showed that a museum collection housing *Megachile venusta* specimens could indeed be a valuable resource in pollinator-plant studies. DNA metabarcoding was used to identify the plant origins in pollen, as well as its microbiome for the ITS1 and ITS2 regions. Using ITS2 as a barcode provided much better resolution for plant identification than ITS1. Multi-locus approaches to DNA barcoding for plants are recommended and ITS1 data should therefore be considered with ITS2 data. Species-level plant classification was possible with ITS2, but without a comprehensive local plant sequence reference database, family-based interpretations are more reliable. The fungal microbiome of pollen sampled reflected storage, environmental and handling contaminants and this need to be considered when planning collection-based metabarcoding studies, as well as how future collection specimens are handled.

3.7 Acknowledgements

Charles Hefer and Jurgens de Bruin from the ARC's Biotechnology Platform are thanked for bioinformatics assistance and advice on some aspects of the work. An anonymous reviewer is also thanked for their insightful comments and contributions towards improving the manuscript.

3.8 References

Bell, K., de Vere, N., Keller, A., Richardson, R., Gous, A., Burgess, K., *et al*., 2016. Pollen DNA barcoding: Current applications and future prospects. Genome 59, 629–640. doi:10.1139/gen-2015-0200.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., Kauserud, H., 2010. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. BMC Microbiol. 10, 1.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., *et al*., 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. 4, 914–919. doi:10.1111/2041-210X.12073.

Bennett, K.D., Parducci, L., 2006. DNA from pollen: principles and potential. The Holocene 16, 1031–1034. doi:10.1177/0959683606069383.

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2015. GenBank. Nucleic Acids Res. 43, D30–D35. doi:10.1093/nar/gku1216.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, btu170. doi:10.1093/bioinformatics/btu170.

Brown, S.P., Veach, A.M., Rigdon-Huss, A.R., Grond, K., Lickteig, S.K., Lothamer, K., *et al*., 2015. Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecol. 13, 221–225. doi:10.1016/j.funeco.2014.08.006.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., *et al*., 2015. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 170, 308–315. doi:10.1016/j.foodchem.2014.08.060.

Burgess, K.S., Fazekas, A.J., Kesanakurti, P.R., Graham, S.W., Husband, B.C., Newmaster, S.G., *et al*., 2011. Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode: Barcoding plants in a local flora. Methods Ecol. Evol. 2, 333–340. doi:10.1111/j.2041-210X.2011.00092.x.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. doi:10.1038/nmeth.f.303.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 106, 12794–12797. doi:10.1073/pnas.0905845106.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., *et al*., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5, e8613. doi:10.1371/journal.pone.0008613.

Cornman, R.S., Otto, C.R.V., Iwanowicz, D., Pettis, J.S., 2015. Taxonomic characterization of honey bee (*Apis mellifera*) pollen foraging based on non-overlapping paired-end sequencing of nuclear ribosomal loci. PLoS ONE 10, e0145365. doi:10.1371/journal.pone.0145365.

Corriveau, J.L., Goff, L.J., Coleman, A.W., 1990. Plastid DNA is not detectable in the male gametes and pollen tubes of an angiosperm (*Antirrhinum majus*) that is maternal for plastid inheritance. Curr. Genet. 17, 439–444. doi:10.1007/BF00334525.

Dafni, A., 1992. Pollination Ecology: A Practical Approach. Oxford University Press, Oxford.

Daily, G.C., Alexander, S., Ehrlich, P.R., Goulder, L., Lubchenco, J., Matson, P.A., *et al*., 1997. Ecosystem services: Benefits supplied to human societies by natural ecosystems. Issues Ecol. 2, 1–18.

de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., *et al*., 2012. DNA barcoding the native flowering plants and conifers of Wales. PLoS ONE 7, e37945. doi:10.1371/journal.pone.0037945.

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., *et al*., 2015. *ycf1*, the most promising plastid DNA barcode of land plants. Sci. Rep. 5, 8348. doi:10.1038/srep08348.

Eardley, C., 2013. A taxonomic revision of the southern African leaf-cutter bees, *Megachile* Latreille *sensu stricto* and *Heriadopsis* Cockerell (Hymenoptera: Apoidea: Megachilidae). Zootaxa 3601. doi:10.11646/zootaxa.3601.1.1.

Edlund, A.F., 2004. Pollen and stigma structure and function: The role of diversity in pollination. Plant Cell 16, S84–S97. doi:10.1105/tpc.015800.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., *et al*., 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3, e2802. doi:10.1371/journal.pone.0002802.

Ferri, G., Corradini, B., Ferrari, F., Santunione, A.L., Palazzoli, F., Alu', M., 2015. Forensic botany II, DNA barcode for land plants: Which markers after the international agreement? Forensic Sci. Int. Genet. 15, 131–136. doi:10.1016/j.fsigen.2014.10.005.

Galimberti, A., De Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., *et al*., 2014. A DNA barcoding approach to characterize pollen collected by honeybees. PLoS ONE 9, e109363. doi:10.1371/journal.pone.0109363.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. U.S.A. 103, 968–971.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6, e17497. doi:10.1371/journal.pone.0017497.

Hargreaves, A.L., Johnson, S.D., Nol, E., 2004. Do floral syndromes predict specialization in plant pollination systems? An experimental test in an "ornithophilous" African Protea. Oecologia 140, 295–301. doi:10.1007/s00442-004-1495-5.

Harris, D.J., 2003. Can you bank on GenBank? Trends Ecol. Evol. 18, 317–319. doi:10.1016/S0169-5347(03)00128-9.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., *et al*., 2015. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10, e0134735. doi:10.1371/journal.pone.0134735.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. doi:10.1098/rspb.2002.2218.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. U.S.A. 101, 14812–14817.

Illumina, 2014. MiSeq Reporter User Guide.

Illumina, 2013. 16S Metagenomic Sequencing Library Preparation.

Johnson, S.D., 1997. Pollination ecotypes of *Satyrium hallackii* (Orchidaceae) in South Africa. Bot. J. Linn. Soc. 123, 225–235. doi:10.1111/j.1095-8339.1997.tb01415.x.

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., *et al*., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 17, 558–566. doi:10.1111/plb.12251.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., *et al*., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. doi:10.1098/rspb.2006.3721.

Koch, H., 2010. Combining morphology and DNA barcoding resolves the taxonomy of Western Malagasy *Liotrigona* Moure, 1961 (Hymenoptera: Apidae: Meliponini). Afr. Invertebr. 51, 413–421.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120. doi:10.1128/AEM.01043-13.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., *et al*., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol. Ecol. Res. 15, 8–16. doi:10.1111/1755-0998.12288.

Kremen, C., Williams, N.M., Aizen, M.A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., *et al*., 2007. Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. Ecol. Lett. 10, 299–314. doi:10.1111/j.1461-0248.2007.01018.x.

Kress, W.J., Erickson, D.L., 2008. DNA barcodes: Genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. U.S.A. 105, 2761–2762.

Kress, W.J., García-Robledo, C., Uriarte, M., Erickson, D.L., 2015. DNA barcodes for ecology, evolution, and conservation. Trends Ecol. Evol. 30, 25–35. doi:10.1016/j.tree.2014.10.008.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U.S.A. 102, 8369–8374. doi:10.1073/pnas.0503123102

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., *et al*., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 1–11. doi:10.1155/2012/251364.

Merritt, J., 2007. Mold: Prevention of growth in museum collections. Conserve O Gram 3/4.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., *et al*., 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386. doi:10.1186/1471-2105-9-386.

Michener, C.D., 2000. The bees of the world, 2nd ed. Johns Hopkins University Press, Baltimore.

Mullins, J., Emberlin, J., 1997. Sampling pollens. J. Aerosol Sci. 28, 365–370. doi:0021-8502/97.

Parducci, L., Suyama, Y., Lascoux, M., Bennett, K.D., 2005. Ancient DNA from pollen: a genetic record of population history in Scots pine. Mol. Ecol. 14, 2873–2882. doi:10.1111/j.1365-294X.2005.02644.x.

Pennisi, E., 2000. Taxonomic Revival. Science 289, 2306–2308.

Petersen, G., Johansen, B., Seberg, O., 1996. PCR and sequencing from a single pollen grain. Plant Mol. Biol. 31, 189–191.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, R.T., Lin, C.-H., Quijia, J.O., Riusech, N.S., Goodell, K., Johnson, R.M., 2015a. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl. Plant Sci. 3, 1500043. doi:10.3732/apps.1500043.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015b. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl. Plant Sci. 3, 1400066. doi:10.3732/apps.1400066.

Shaw, D.E., 1999. Bees and fungi, with special reference to certain plant pathogens [Internet]. Daniel McAlpine Memorial Lecture. URL http://www.appsnet.org/Publications/McAlpine/DM12.htm (accessed 3.14.16).

Sheffield, C.S., Hebert, P.D.N., Kevan, P.G., Packer, L., 2009. DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol. Ecol. Res. 9, 196–207. doi:10.1111/j.1755-0998.2009.02645.x.

Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., *et al*., 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol. 15. doi:10.1186/s12898-015-0051-y.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA barcoding for honey biodiversity. Diversity 2, 610–617. doi:10.3390/d2040610.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. doi:10.1128/AEM.00062-07.

Wang, X.-C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J.-H., *et al*., 2015. ITS1: a DNA barcode better than ITS2 in eukaryotes? Mol. Ecol. Res. 15, 573–586. doi:10.1111/1755-0998.12325.

Wester, P., Stanway, R., Pauw, A., 2009. Mice pollinate the Pagoda Lily, *Whiteheadia bifolia* (Hyacinthaceae) — First field observations with photographic documentation of rodent pollination in South Africa. S. Afr. J. Bot. 75, 713–719. doi:10.1016/j.sajb.2009.07.005.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: PCR Protocols: A Guide to Methods and Applications. Academic Press Inc., New York.

Wilcock, C., Neiland, R., 2002. Pollination failure in plants: why it happens and when it matters. Trends Plant Sci. 7, 270–277. doi:S1360-1385(02)02258-6.

Willerslev, E., Hansen, A.J., Poinar, H.N., 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. Trends Ecol. Evol. 19, 141–147. doi:10.1016/j.tree.2003.11.010

Williams, N.M., Kremen, C., 2007. Resource distributions among habitats determine solitary bee offspring production in a mosaic landscape. Ecol. Appl. 17, 910–921. doi:10.1890/06-0269.

Wilson, E.E., Sidhu, C.S., LeVan, K.E., Holway, D.A., 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol. Ecol. 19, 4823–4829. doi:10.1111/j.1365-294X.2010.04849.x.

Witt, J.D.S., Threloff, D.L., Hebert, P.D.N., 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Mol. Ecol. 15, 3073–3082. doi:10.1111/j.1365-294X.2006.02999.x.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., *et al*., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5, e13102. doi:10.1371/journal.pone.0013102.

Zhou, L.-J., Pei, K.-Q., Zhou, B., Ma, K.-P., 2007. A molecular approach to species identification of Chenopodiaceae pollen grains in surface soil. Am. J. Bot. 94, 477–481.

3.9 Supplementary data

The files that are referred to in this chapter as supplementary data are presented in this section in the order they are mentioned in the text.

Table S3.1. Collection information of *Megachile venusta* bee specimens used for pollen sample collection in this study. Bees were obtained from the National Insect Collection housed at the ARC's Biosystematics, Pretoria, South Africa. Collection information, such as the date, province, GPS coordinates and nearest town are given for each sample, where available.

Bee collection identifier[1]	**Sample identifier in this study**	**Bee collection date**	**Province[2]**	**GPS**	**Bee collection description**
HYMA05682	A1	10.12.1916	Transvaal (Gauteng)	25.43S 28.11E	Pretoria
HYMA05679	A2	30.10.1914	Orange Free State (Free State)	29.07S27.28E	Modderpoort
HYMA05680	A3	05.11.1914	Orange Free State (Free State)	29.27S26.13E	Bloemfontein
HYMA05677/1	A4	18.10.1921	North West	27.32S 24.48E	Taung
HYMA05678	A5	11.1921	Transvaal (Mpumalanga)	25.28S 30.59E	Nelspruit (Mbombela)
HYMA05677/2	A6	18.10.1921	North West	27.32S 24.48E	Taung
HYMA27297/1	A7	12.1948	Eastern Cape	33.50S 25.34E	Redhouse near Port Elizabeth
HYMA05723	A8	21.05.1969	Transvaal(Mpumalanga)	25.21S 31.53E	Kruger National Park at Crocodile bridge
HYMA05505	A9	23.09.1962	Transvaal (North West)	25.39S 26.41E	Swartruggens
HYMA05503	A10	5.03.1963	Natal (KwaZulu-Natal)	29.00S 29.53E	Estcourt
HYMA05769	A11	10.02.1977	Northern Cape	27.27S 23.26E	Kuruman
HYMA05520	A12	03.01.1970	Transvaal (Gauteng)	25.56S 28.13E	Olifantsfontein

HYMA05608	A13	11.12.1977	Transvaal (North West)	26.42S 27.05E	Potchefstroom (Tlokwe)
HYMA05804/1	A14	29.10.1985	Transvaal (Gauteng)	25.45S 28.12E	Pretoria at Gardens of Union Building
HYMA05804/2	A15	29.10.1986	Transvaal (Gauteng)	25.45S 28.12E	Pretoria at Gardens of Union Building
HYMA05890	A16	10.11.1981	Natal (KwaZulu-Natal)	28.55S 29.14E	Cathedral Peak Forestry Area
HYMA27298	A17	11.01.1993	Gauteng	25.41S 28.18E	Roodeplaat Research Station
HYMA05911	A18	24-25.02.1993	Free State	27.40S 25.45E	Sandveld Nature Reserve
HYMA27299	A19	24-25.02.1993	Free State	27.40S 25.45E	Sandveld Nature Reserve
HYMA27300	A20	09.09.2007	Northern Cape	28.07S 17.00E	Richtersveld National Park at Hand of God
HYMA27301	A21	19.01.2004	Northern Cape	27.13S 22.55E	4 km West of Hotazel
HYMA27302	A22	19.09.2005	Western Cape	32.12S 18.53E	South of Clanwilliam

[1]National Insect Collection's (ARC, Pretoria) unique identifiers.
[2]South Africa has re-divided and renamed some of their provinces and towns. The original collection province and towns are given with the new names provided in brackets.

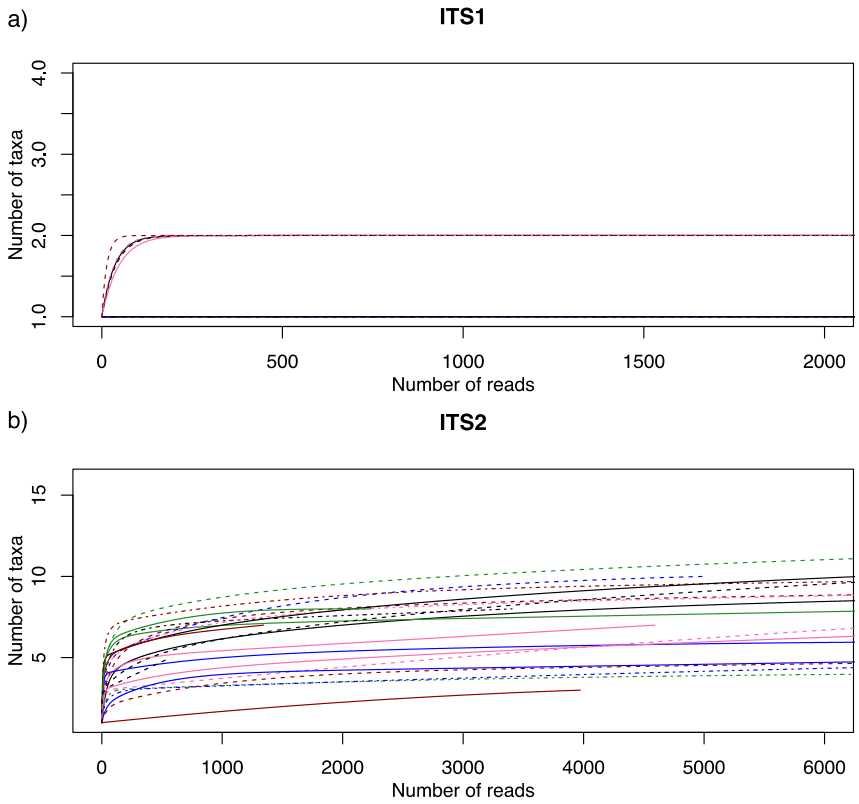

Figure S3.1. Rarefaction curves for a) ITS1 and b) ITS2 samples. Rarefaction curves were created on all reads, prior to removal of 0.1% rare taxa per sample. No difference was observed for ITS1, but an increase of two species was observed in the maximum number of species per sample for sITS2.

CHAPTER 4: FLORAL CHOICE IN BEES (*MEGACHILE*: MEGACHILIDAE) FROM SOUTH AFRICA AS REVEALED BY POLLEN METABARCODING OF HISTORIC SPECIMENS

4.1 Abstract

South Africa has high levels of plant and animal endemism and diversity. Bees in particular show high diversity and endemism in the western part of the country. Not much is currently known about the pollen preferences of indigenous bees in South Africa, with data only available from observational studies. Pollen metabarcoding provides provenance information by utilising DNA analyses instead of traditional microscopic identifications. In this study, pollen was sampled from bee specimens from a historic insect collection (National Collection of Insects, South Africa) from two florally important areas, as well as a group distributed countrywide covering various biomes. Bees were collected in the Succulent Karoo, the Savanna, and from across the country. The nuclear internal transcribed spacer 2 (ITS2) region was amplified, sequenced and compared to a sequence reference database to assign taxonomic classifications to family level. Sequence reads were also clustered to operational taxonomic units (OTUs) based on 97% sequence similarity to obtain plant species visit ranges. General linear models (GLM) showed no significant difference in the mean number of plant taxa visited by bees in the genus *Megachile* between the Succulent Karoo and Savanna, but the widespread group visited significantly more taxa than the other two groups on average. The number of floral visits made by *Megachile niveofasciata* accounted for the difference in the three groups. Bees from the widespread group were characterised by a significantly different composition in pollen assemblage than the other two groups. Time since sampling did not have an effect on the mean number of taxa visited by any of the bee species studied.

Keywords: *pollen preference, South Africa, insect collection, ITS2, Succulent Karoo, Savanna*

4.2 Introduction

The use of next-generation sequencing (NGS) technology and DNA barcoding in high-throughput identification of plant origins from pollen samples has been on the rise in recent years (Bruni *et al*., 2015; Keller *et al*., 2015; McFrederick and Rehan, 2016; Sickel *et al*., 2015). There are many advantages to this method over traditional microscopic palynology, including increased time-efficiency, being able to multiplex samples that in turn reduce costs, and the relative ease of the process that does not require a trained palynologist (Bell *et al*., 2016). Genetic analyses of pollen also allows for more accuracy, since pollen morphology is a

limiting factor for identifications due to the pollen of some taxa being indistinguishable (Rahl, 2008). Pollen metabarcoding results have been shown to be consistent in comparison to microscopy-based identifications (Keller *et al*., 2015; Kraaijeveld *et al*., 2015; Richardson *et al*., 2015a), but there is no consensus yet regarding the consistency of pollen quantification between the two methods (Keller *et al*., 2015; Richardson *et al*., 2015b).

The applications of pollen metabarcoding are extensive, and include monitoring food- and air-quality, forensic studies, and reconstructing ancient plant communities, among others (Bell *et al*., 2016). Metabarcoding pollen loads sampled from pollinators, as opposed to conducting lengthy field experiments, can provide insight into plant-pollinator interactions in a much more efficient manner. As there is still a great deal to learn about interactions between most pollinators and the plants they visit, this optimised way of studying them is invaluable. In particular, very little is known about the pollen plants of many endemic bee species in South Africa, as their interactions with plants are only studied using observation (such as in Johnson, 1997; Pauw, 2006; Pauw and Stanway, 2015).

South Africa is exceptionally rich in plant and animal diversity, with insects accounting for the majority of animal species present within the country (Da Silva and Willows-Munro, 2016; Hamer, 2013; Van Wyk and Smith, 2001). Plant and bee diversity and endemism is particularly high in the western part of the country (Eardley, 1989; Eardley *et al*., 2009; Kuhlmann, 2009; Van Wyk and Smith, 2001) with high levels of specialisation reported for this biologically important area (Pauw and Stanway, 2015). Specialist bee and plant species are more susceptible to changes in the environment and consequently more vulnerable to declines and extinction (Packer *et al*., 2005; Zayed, 2009). This is particularly troubling in the face of global climate change. Given this susceptibility and the high levels of endemism in South Africa, it is of great importance to study the interactions between bees and plants of diversity hotspots.

Taxonomists collecting bees for identification purposes have filled insect collections across South Africa with specimens from a wide range of taxa. These collections tend to be excellently maintained with information included with each specimen on when, where and by whom it was collected (Pennisi, 2000). Specimen labels sometimes also include information about plant associations. Bees in these collections are not stripped of their pollen loads during their taxonomic identification and thus could potentially contain a wealth of information about the plants bees visited before capture. This untapped resource can be taken advantage of by sampling pollen from insect collection specimens for metabarcoding purposes. Recently, research showed that plant origin identifications could be made from pollen sampled from historic insect collection specimens dating back to 1914 using the internal transcribed spacer

regions (ITS1 and ITS2; Chapter 3). Due to the lack of South African sequences represented in the reference sequence database, species- and genus-level plant classification were problematic, and family-level interpretations were recommended.

The use of historically collected specimens in answering biological questions can potentially greatly increase the scope of questions that can be answered, as well as increase sample sizes by supplementing data obtained from contemporary field work. In this study, bee specimens from a National Insect Collection in South Africa were sampled for their pollen loads to determine whether the floral choice of bees in the genus *Megachile*, family Megachilidae, differ between the high-diversity Succulent Karoo in the west and less diverse Savanna regions in the east of the country. As comparison, specimens of selected Megachilidae species that occur throughout South Africa were included in the analyses. The ITS2 region of the pollen nuclear genome was analysed to reveal plant provenance and plant family affinities for each group discussed.

4.3 Materials and methods

4.3.1 Specimen selection

Bee specimens are housed in the National Insect Collection of the Agricultural Research Council (ARC), at Biosystematics in Pretoria, South Africa. Taxonomic experts have previously identified all specimens housed in the collection by standard methods. The conditions of bee storage were as described in Chapter 3. Six bee species from the same genus, *Megachile*, were selected from different biomes of interest in South Africa; two species restricted to the Succulent Karoo, two species restricted to the Savanna biome, and two widespread species for comparison. The widespread species were selected on their larger geographic occurrence and not necessarily based on their foraging behaviour (specialists/generalists). It was postulated that they should be more generalist in foraging behaviour. The species selected from the different biomes can be seen in Table 4.1, and specimen sampling localities are mapped in Fig. 4.1. Not all the specimens in an insect collection would carry pollen on their bodies (Chapter 3), therefore the numbers of specimens available to sample for pollen varied for each bee species. All specimens with pollen visible on their bodies were selected, regardless of age, for five of the six species. Specific temporal points were consequently not selected when sampling. One of the two widespread species selected, *Megachile venusta*, was used in a previous pollen metabarcoding study (Chapter 3), where it was shown that that DNA could be successfully sequenced from a wide temporal range of samples. The original study consisted of 22 *M. venusta* specimens, with one sample

(sample 6) being excluded because too few reads could be confidently classified to species level. This sample was also excluded from this study, as it did not match the criteria for analysis when classified to family level as discussed below.

Table 4.1. Species from the genus *Megachile* (Megachilidae) selected for study from different biomes of interest in South Africa. The number of specimens in the National Insect Collection at the ARC that contained pollen for sampling is also indicated, as well as the age of specimens included.

Species	Biome	Number of specimens	Years sampled
Megachile karooensis	Succulent Karoo	20	1982 - 1990
Megachile murina	Succulent Karoo	27	1982 - 1990
Megachile felina	Savanna	17	1966 - 1990
Megachile maxillosa	Savanna	32	1914 - 2003
Megachile niveofasciata	Widespread	10	1984 - 2000
Megachile venusta	Widespread	21	1914 - 2007

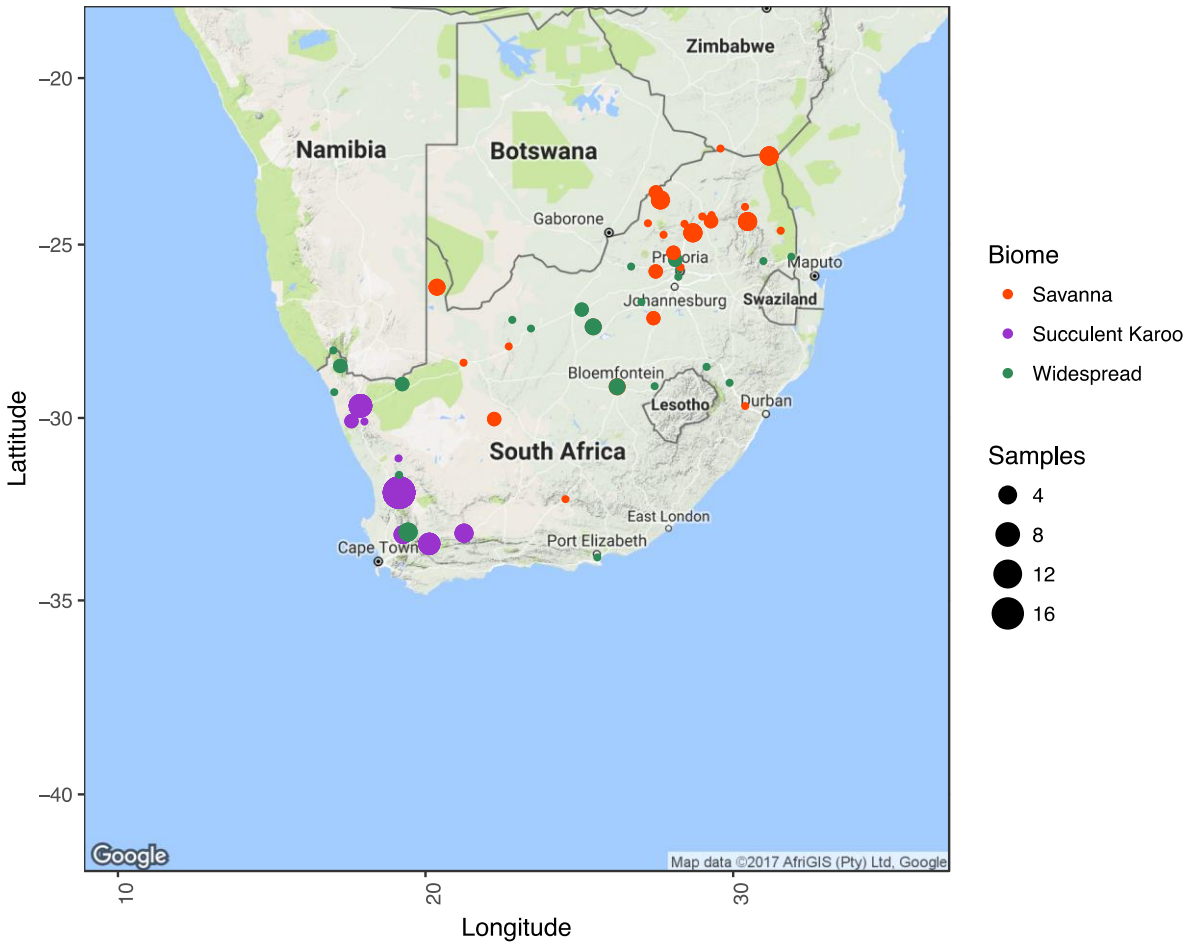

Figure 4.1. A map of South Africa with the sampling localities of bee specimens indicated.

Information on all bee specimens, including accession numbers, original sampling dates of bees, sampling areas and GPS coordinates (if available) is provided as supplementary information (supplementary Tables S4.1 – S4.5, and Table S3.1 from Chapter 3).

A stereo dissection microscope (SteREO Discovery.V8 microscope, Carl Zeiss Microscopy GmbH, Jena, Germany) was used to view the specimens to confirm the presence of pollen, as well as to scrape pollen from the selected bee specimen's scopae. Sterile micropipette tips were dipped in sterilised glycerol as described previously (Chapter 3) and used to remove pollen from bee specimens. Micropipette tips were saved in the respective sample's 1.5 ml Eppendorf tube after crushing of the pollen to save pollen fragments inside the tip until DNA extraction.

The optimal approach to extract DNA from pollen sampled from specimens selected from an insect collection was previously investigated and documented (Chapter 3). In this study, the extraction method previously found to be the most suitable was used for DNA extraction of all pollen samples. The DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used for DNA extraction, without any bead disruption during lysis. The only maceration done was by micropipette tip crushing after pollen was scraped off the specimens. Extraction was performed according to the manufacturer's protocol, with incorporation of the modifications as described in Chapter 3. DNA was eluted in 50 µl of buffer EB, with reapplication of the eluate to the DNeasy Mini Spin Column for a second elution step to increase DNA yield.

The nuclear ITS2 region was selected as the barcode to be targeted for the identification of pollen's plant origins. The primers for ITS2 were used successfully before for pollen metabarcoding (Chapter 3) and were modified to include overhang adapters as described in the Illumina *16S Metagenomic Sequencing Library Preparation Guide* (Illumina, 2013). These overhang adapters allow the primers to be used directly in the standard Illumina indexing and adapter PCR. The primers used are ITS3F_Tag_IL 5' **TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** GCA TCG ATG AAG AAC GCA GC 3' and ITS4R_Tag_IL 5' **GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**TC CTC CGC TTA TTG ATA TGC 3' (overhang adapters indicated in bold and underlined). Oligonucleotides were synthesised by Inqaba Biotechnical Industries (Pty) Ltd. (Pretoria, South Africa).

Barcode amplification was achieved in reactions with a final concentration of 1× Phusion® High Fidelity PCR Master Mix with HF Buffer (Thermo Scientific, Waltham, MA, USA), 0.5 µM of each primer, and 5 µl of DNA template. Milli- Q® H_2O (Merck Millipore, KGaA,

Darmstadt, Germany) was added to a final reaction volume of 50 µl. Amplification cycling was done as described in Chapter 3.

Amplification products were purified using the Agencourt AMPure XP (Beckman Coulter, Brea, California, USA) bead purification system according to the manufacturer's protocol. The DNA concentration of approximately half of the samples was evaluated using a Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA) and a Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen, Life Technologies). Samples were randomly selected for evaluation and both kits were used according to manufacturer instructions.

Nextera XT (Illumina, Inc. San Diego, CA, USA) indexes were added according to the sequencing preparation protocol (Illumina, 2013) to multiplex samples. Indexed samples were pooled equimolarly and sequenced on a HiSeq 2500 sequencer (Illumina, Inc.) at the ARC's Biotechnology Platform, Pretoria, South Africa, using the HiSeq Rapid SBS Kit v2, with 2 × 250 bp paired end reads (Illumina, Inc.).

4.3.2 Bioinformatics and statistical analyses

Demultiplexing of samples was done with CASAVA v1.8.2 (Illumina Part #15011196 Rev D) based on the Nextera index sequences used. Low quality bases and adapter sequences were trimmed from reads with Trimmomatic 0.35 (Bolger *et al*., 2014) using a sliding window of eight bases with an average quality of 20 required per window. Trimmed reads were merged in MacQiime 1.9.1-20150604 (Caporaso *et al*., 2010) using the *multiple_join_paired_ends.py* script.

The ITS2 database for Viridiplantae created by Sickel *et al*. (2015) was used for identification of the plant family origins of pollen samples. Classification of sequences was performed using the Ribosomal Database Project (RDP) classifier v. 2.10.1 (Wang *et al*., 2007). Analysis was performed following the bioinformatics workflow described in the aforementioned publication. Reads not meeting the required 0.8 confidence level at family level after classification, and rare taxa (less than 0.1% of the total amount of reads identified as plant) were removed prior to further analyses to remove sequencing artifacts (Brown *et al.*, 2015). Each plant family identified by RDP classifier was treated as an operational taxonomic unit (OTU). Data were handled in two different ways. Biases during DNA extraction of pollen and DNA amplification may lead to a skew in the abundances of certain taxa being identified (Keller *et al*., 2015; Shokralla *et al*., 2012). To circumvent this, plant family data were converted to absence/presence counts for each plant family for each sample in a biome, referred to as detection counts. These counts were added across samples for each family to

obtain plant family prevalence within a biome group across all samples. Data were also analysed as read counts per sample. OTU tables were created for read and detection counts for family abundances.

The Sickel *et al*. (2015) ITS2 database did not represent South African plant diversity adequately to make species-level classifications (Chapter 3). In order to get an estimation of angiosperm species abundance ranges for the possible different bee species, sequence reads from each sample were clustered into OTUs using the *cluster_otus* command in usearch v8.0.1517 (Edgar, 2010). All reads per sample were then assigned to OTUs using *usearch_global* alignment, and taxonomy added to the OTU sequences with the *utax* command using the utax-compatible ITS2 database from Sickel *et al*. (2015). During the taxonomy assignment, confidence scores were not assigned, and a raw score cut-off of 10 was chosen for a species assignment to be included. Species names were changed to Sp1 to Sp49 as we were not interested in the actual species assignment, but only their abundance ranges. OTU tables were again created for both read counts and detection counts for species abundances, while disregarding species not from angiosperm families.

Rarefaction curves were drawn for plant family and species assignments for all bee species using vegan v. 2.3.4 (Oksanen *et al*., 2016) in R v. 3.2.4 (R Core Team, 2016). Taxonomic assignments were checked for local family occurrence against the Plants of southern Africa (POSA) database (Germishuizen *et al*., 2003, accessed 31 July 2016). For the Succulent Karoo biome, the search was confined to the Western Cape, Northern Cape and Cape Region (as defined in the database), for the Savanna biome the remaining regions were selected as search criteria and for the widespread species all South African provinces were selected in addition to the plant family name. The Plants of southern Africa database (POSA, Germishuizen *et al*., 2003), the vegetation map of South Africa (Low and Rebelo, 1996), and a quarter degree grid square (QDGS) shape file were used to calculate the frequency of occurrences per family within each biome in South Africa. Spatial analyses were carried out using rgdal v.1.2-4 in R v. 3.2.4 (R Core Team, 2016). A list of recorded flower visits was also created for each bee species using Eardley (2012), Eardley (2013), Gess and Gess (2014), and the Catalogue of Afrotropical Bees, accessed through the online Global Biodiversity Information Facility (GBIF, www.gbif.org, Eardley and Urban, 2010) for comparison.

To compare bee taxa and to assess the effect of time since collection on the mean number of taxa detectable in pollen loads, generalised linear models that incorporated a Poisson distribution and log link function were used in SPSS 23.0 (IBM Corp., Armonk, New York). Separate models were run with angiosperm species, angiosperm families, or all families (including lower plants) as response variables. To account for statistical non-independence

among bees sampled from the same localities, locality was treated as a subject in generalised estimating equations that used an exchangeable correlation matrix. Bee species were treated as a fixed factor and time was treated as a covariate. Significance was assessed using Wald statistics, and post-hoc comparisons among means were carried out using the Dunn-Sidak procedure. Marginal (model-adjusted) means were obtained by back-transformation from the log scale, which also resulted in asymmetrical standard errors.

To assess whether bee species were characterised by different assemblages of pollen, a similarity matrix for square-root transformed data on detection counts were calculated using the Bray-Curtis method and then plotted in two-dimensions with non-metric multi-dimensional scaling (NMDS) using Past 3.14 (Hammer *et al*., 2001). The significance of differences in pollen assemblages among bee species was assessed using ANOSIM, a non-parametric permutation procedure based on the similarity matrix underlying the ordination. Observed R-values were compared with the distribution of R-values generated by up to 10,000 random permutations of the sample labels in order to assess statistical significance.

4.4 Results

4.4.1 Next-generation sequencing (NGS) of bee-specimen derived pollen DNA

Pollen from five of the six bee species was sequenced using an Illumina HiSeq 2500. Pollen from *M. venusta* samples was sequenced using an Illumina MiSeq and classified in a previous study (Chapter 3). Summary statistics of the merged reads produced in this study are provided first. Quality and adapter trimming of samples sequenced on the HiSeq 2500 resulted in a total number of 2,954,892 high-quality, merged reads obtained across all bee species. This resulted in a mean of 28,412 merged reads per sample (median = 13,816 and SD = 43,161). Twenty-four of the 104 samples (22.6%) produced less than 1,000 reads per sample, and were discarded prior to further analyses. Pollen from 21 *M. venusta* specimens from Chapter 3 produced a total of 1,124,324 reads across all samples, with a mean of 53,539 reads per pollen sample (SD = 57,314). When combined with samples sequenced on the HiSeq 2500, a total of 4,071,658 reads were produced across all six *Megachile* species, after the 24 failed samples were removed. This results in a mean of 40,313 (median = 24,193; SD = 48,742) reads per pollen sample. The mean number of merged reads per *Megachile* species is provided in Table 4.2.

Table 4.2. Summary statistics of merged reads (after quality and adapter trimming was performed, and subsequent merging of forward and reverse reads) of the six *Megachile* species investigated in this study.

Bee species*	Sum of reads	Mean of reads	Median of reads	Standard deviation
Megachile felina (**18**, *12*)	664,462	51,113	39,250	50,581
Megachile maxillosa (**32**, *18)*	831,287	39,585	26,956	56,087
Megachile karooensis (**20**, *15*)	362,278	22,642	9,954	26,406
Megachile murina (**27**, *18)*	787,116	39,356	23,517	52,441
Megachile niveofasciata (**10**, *10*)	302,191	30,219	20,800	21,917
Megachile venusta# (**22**, *21*)	1,124,324	53,539	26,117	57,314

* Original pollen sample numbers and the number of pollen samples after data filtering are given in bold and italics after each bee species, respectively.
# The summary statistics represented for *Megachile venusta* are based on sequencing performed on an Illumina MiSeq, and disregards one sample which was excluded in the previous analysis (Chapter 3).

ITS2 sequence classification resulted in 71.9% of total reads being confidently identified to family level across all samples. Samples with less than 1,000 reads confidently classified to family level were discarded. Samples that passed the family level read cut-off were all subjected to species level taxonomic classification in utax. An additional read cut-off was not introduced at species level classification, as only reads classified as angiosperm species were recorded. Thirty-four pollen samples (26.6%) across all bee species were discarded because less than 1,000 reads were confidently classified to family level in each sample.

Only read counts confidently identified to family level and reads classified as angiosperm species were used to draw rarefaction curves to determine whether sequence saturation was achieved. Rarefaction curves indicated that for all bee species examined, sufficient reads were sequenced to resolve all possible plant taxa present in the pollen samples (supplementary Figure S4.1).

4.4.2 Pollen preferences of **Megachile** *bee species from different biomes*

Based on detection counts, the most prevalent families in all three groups were the Amaranthaceae and Pteridaceae (Figure 4.2). From the two bee species representing the Succulent Karoo biome, Pteridaceae were detected in 31 samples (93.9%), Amaranthaceae in 29 samples (87.9%), and Poaceae in 24 samples (72.7%) as the top three represented plant families. Similarly, in the Savanna biome, Pteridaceae were detected in 30 samples (100%), Amaranthaceae in 26 samples (86.7%), and Poaceae in 24 samples (80%). The samples from

the two bee species in the regional group also detected Pteridaceae as the most prevalent family (29 samples, 93.5%), with Asteraceae and Amaranthaceae detected in equal abundance across samples (26 samples, 83.9%). When plants not belonging to Angiosperm families were removed, the third most prevalent family across all samples in both the Succulent Karoo and the Savanna biomes was Polygonaceae, being detected in 9 (27.3%) and 8 (26.7%) samples, respectively. Figure 4.2 represents the different proportions in which the less abundant plant families were detected between biomes. Notably, Asteraceae was identified in 26 samples (83.9%) in the widespread bees but only in very low proportions of samples for the other two groups (2 samples (6.4%) in the Succulent Karoo and 2 samples (6.7%) in the Savanna bee species). Fabaceae was also identified in a much higher proportion of widespread samples (20 samples, 64.5%) than those in the Savanna (6 samples, 20%) or Succulent Karoo (7 samples, 22.6%) bee species. In general, higher proportions of samples with positive identifications were seen for more plant families for the widespread bee group, with many of the plant families only visited by bees from this group.

Of the three biome groups, the widespread bee species visited the widest range of plant families with 30 families visited compared to 15 visited by the Savanna group and 18 visited by the Succulent Karoo group. After classification of reads into species OTUs, the widespread bee species again visited the most plant species, with a total of 39 different plant species recognised. The Succulent Karoo bees visited a total of 25 plant species and the Savanna bees visited a total of 16 plant species. The numbers of plant families and species visited by each bee species in each area studied can be seen in Table 4.3 and Figure 4.3.

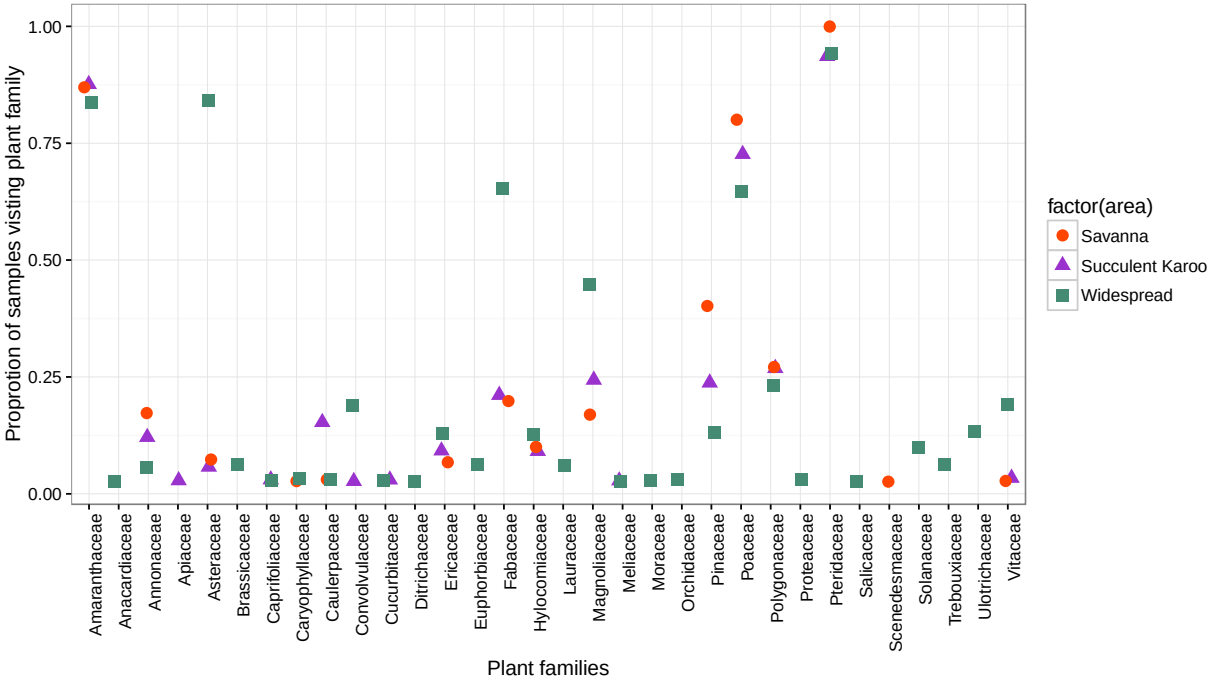

Figure 4.2. Visual representation of the proportion of samples in which each plant family was identified in each biome group. Families from the phylum Chlorophyta were removed prior to plotting of the graph. The family Pteridaceae was identified in nearly all samples in all three biome groups, with clear differences in the prevalence of Asteraceae, Fabaceae and Magnoliaceae between the widespread bee group and the other two biome groups. The absence of a group marker in a column indicates that the particular plant family was not detected in pollen loads from any bee specimens in that biome group.

Table 4.3. Summaries of the numbers of plant families and species that were visited by the bee species from the three regions studied. Data is based on detection counts for all plant families detected (including lower plants), counts of families not belonging to the angiosperms have been removed, and counts for angiosperm families identified with usearch.

OTU Dataset	Bee species (Region[1])	Number visited	Mean number visited	Overall mean number visited per area
All plant families	*M. niveofasciata* (W)	6 – 10	7.9	6.26 (SD = 2.14)
	M. venusta (W)	3 – 10	5.48	
	M. karooensis (SK)	2 – 9	4.53	4.27 (SD = 1.75)
	M. murina (SK)	2 – 8	4.06	
	M. felina (SV)	2 – 7	4.25	4.23 (SD = 1.30)
	M. maxillosa (SV)	2 – 8	4.22	
Angiosperm plant families	*M. niveofasciata* (W)	5 – 9	6.50	4.81 (SD = 2.01)
	M. venusta (W)	2 – 9	4.00	
	M. karooensis (SK)	1 – 6	3.00	2.85 (SD = 1.52)
	M. murina (SK)	1 – 7	2.72	
	M. felina (SV)	1 – 5	2.58	2.67 (SD = 1.09)
	M. maxillosa (SV)	1 – 5	2.72	
Angiosperm plant species	*M. niveofasciata* (W)	1 – 14	8.8	7.39 (SD 3.02)
	M. venusta (W)	2 – 14	6.71	
	M. karooensis (SK)	2 – 7	4.27	4.30 (SD = 1.74)
	M. murina (SK)	1 – 8	4.33	
	M. felina (SV)	2 – 9	5.25	4.13 (SD = 2.16)
	M. maxillosa (SV)	2 - 8	3.39	

[1] The regions in which the bee species belong are abbreviated for convenience. The widespread area is abbreviated as “W”, the Succulent Karoo biome is abbreviated as “SK”, and the Savanna biome is abbreviated as “SV”.

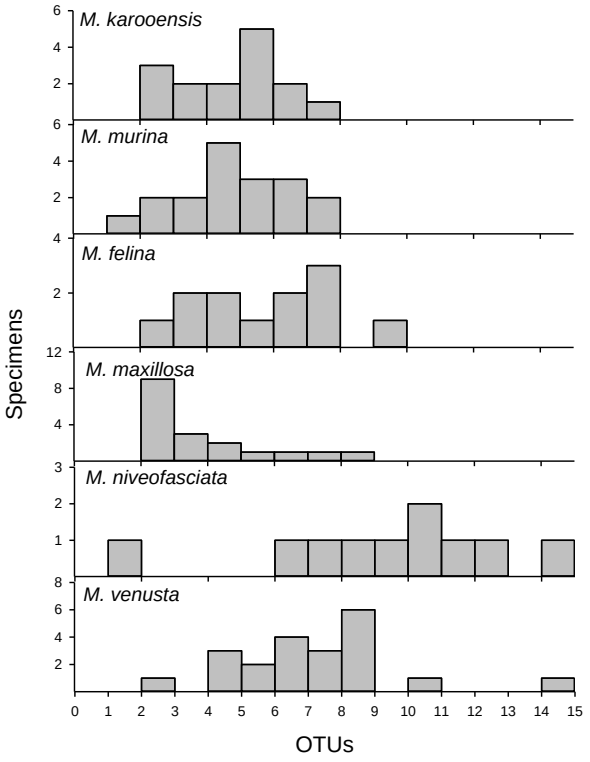

Figure 4.3. Frequency distribution of plant species detected in each bee species studied. Bees from the Succulent Karoo (*M. karooensis* and *M. murina*) and the Savanna (*M. felina* and *M. maxillosa*) show similar species abundances, with the two widespread bee species (*M. niveofasciata* and *M. venusta*) both detecting up to 14 different plant species each.

The mean number of angiosperm species detected in pollen loads varied significantly among bee species ($\chi 2 = 115.49$, $P < 0.0001$, Fig. 4.4). Similarly, the mean number of plant families also varied significantly between bee species ($\chi 2 = 130.51$, $P < 0.0001$, Figure 4.5a for angiosperm families, and $\chi 2 = 124.87$, $P < 0.0001$, Figure 4.5b for all families, including lower plants). In Figure 4.4 and Figure 4.5, the widespread group has higher means than the other two groups based on the Dunn-Sidak post-hoc tests. However, *M. venusta* did not differ significantly from *M. felina* in the mean number of OTU species visited. When comparing bee species on the mean number of Angiosperm families or all plant families visited, *M. venusta* also did not differ significantly from the other bee species. Sample-based rarefaction of species OTU data indicated that *M. venusta* has higher expected species richness than the other bee species (Figure 4.5), however. No significant effect of time since bee collection was found, however, for either plant species ($\chi 2 = 2.28$, $P = 0.131$), or plant family ($\chi 2 = 0.56$, $P = 0.454$ for angiosperm families, and $\chi 2 = 0.90$, $P = 0.344$ for all families, including lower plants).

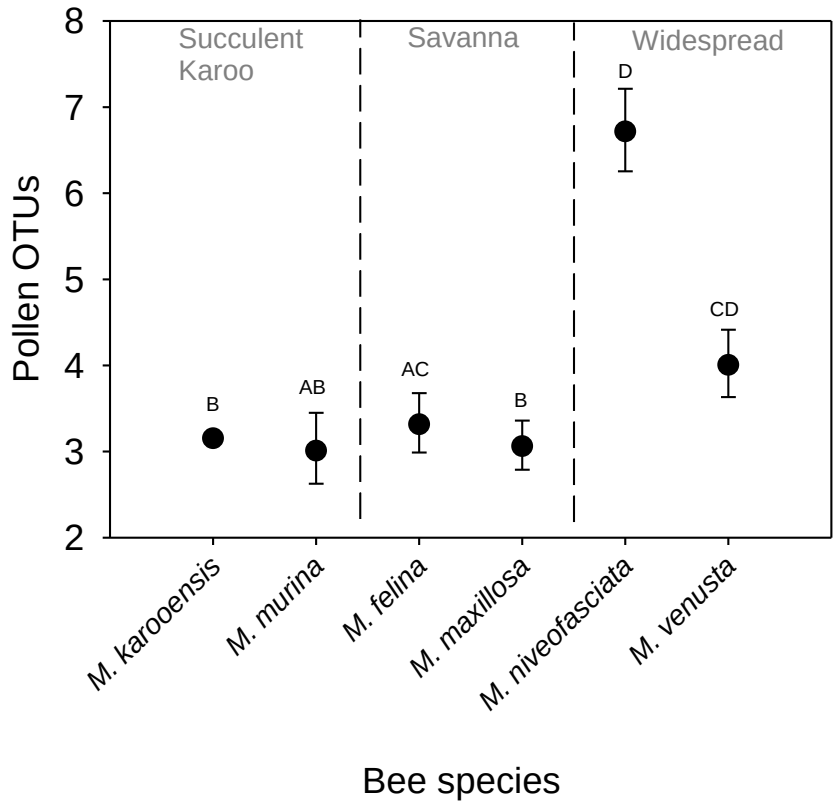

Figure 4.4. Model-adjusted means for plant species OTUs detected in pollen samples in the three regions studied. The mean number of species detected in bees from the Succulent Karoo (*M. karooensis* and *M. murina*) and the Savanna (*M. felina* and *M. maxillosa*) are lower than those in widespread bee species (*M. niveofasciata* and *M. venusta*). Means sharing a letter do not differ significantly at a 95% significance level after Sequential Dunn-Sidak post-hoc testing.

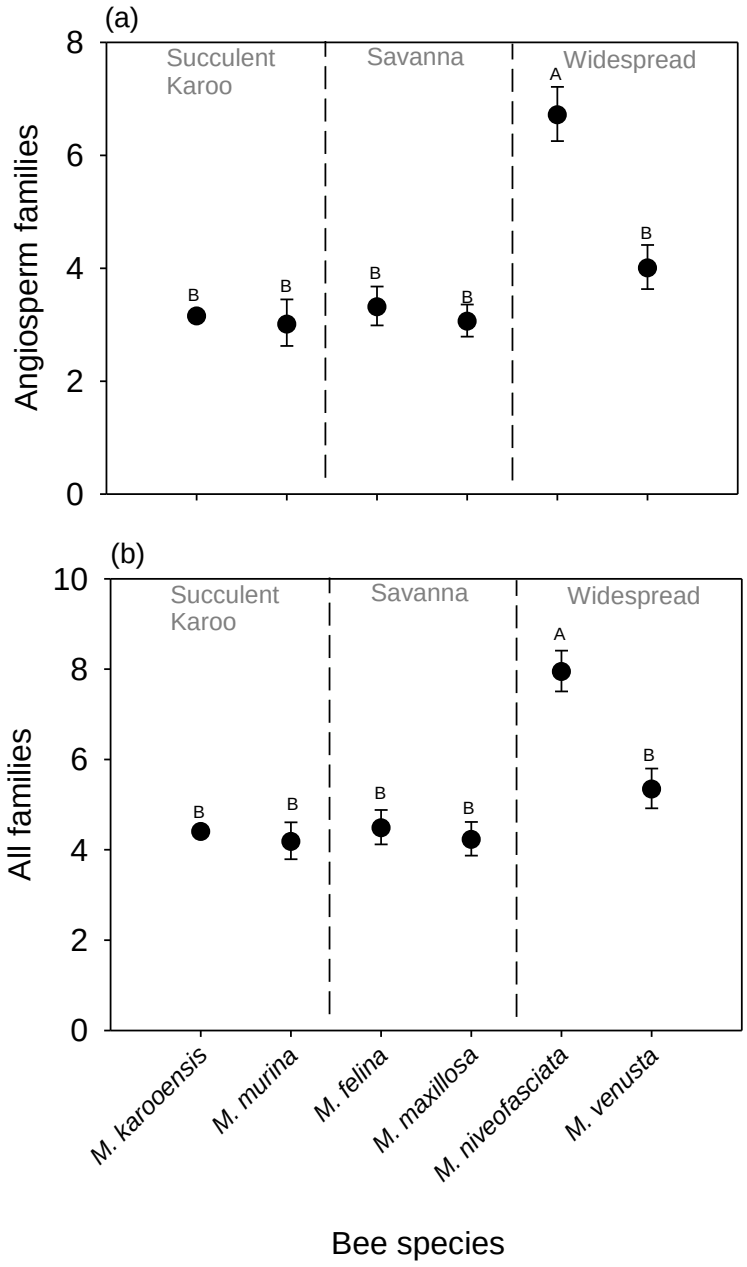

Figure 4.5. Model-adjusted means of plant family OTUs detected in pollen samples in the three regions studied. Angiosperm family means and plant family means overall were lower in bees from the Succulent Karoo (*M. karooensis* and *M. murina*) and the Savanna (*M. felina* and *M. maxillosa*) than those in widespread bee species (*M. niveofasciata* and *M. venusta*). Means sharing a letter do not differ significantly at a 95% significance level after Sequential Dunn-Sidak post-hoc testing. Only the means in *M. niveofasciata* differed significantly from all other bee species. In Figure 4 (a) the model-adjusted means of angiosperm plant families detected in bee species from different regions are provided and in (b) the model-adjusted means of plant families overall (including lower plants) that were detected in bee species from different regions are given.

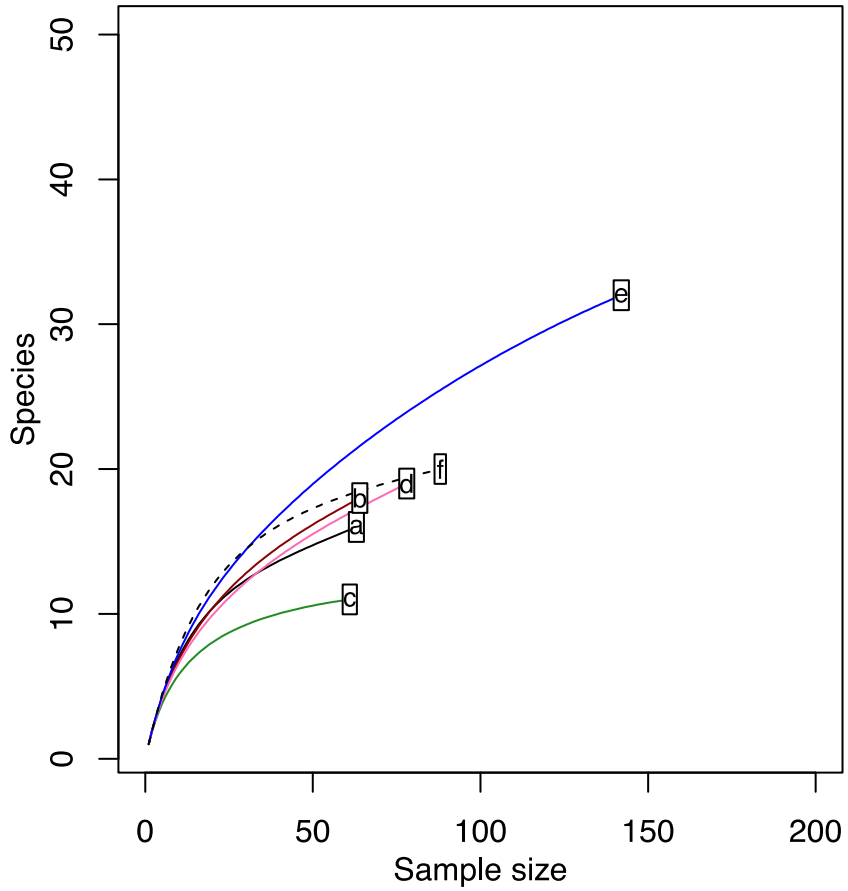

Figure 4.6. Sample-based rarefaction curve to estimate plant species richness for each bee species studied. *Megachile venusta* (e) was shown to have the highest estimated species richness, with *M. maxillosa* (c) having the lowest. The other species, *M. felina* (a), *M. karooensis* (b), *M. murina* (d), and *M. niveofasciata* (f) had similar species richness values.

There is a small, but significant difference in pollen composition in terms of plant species between regional groups of bees as shown by NMDS (Figure 4.7) and ANOSIM analysis ($R = 0.24$, $P < 0.0001$). Uncorrected significance values between the different groups compared in ANOSIM (supplementary Table S4.6) indicated that *M. venusta* and *M. niveofasciata* were the only two species differing significantly from the others, accounting for the overall difference in pollen composition shown.

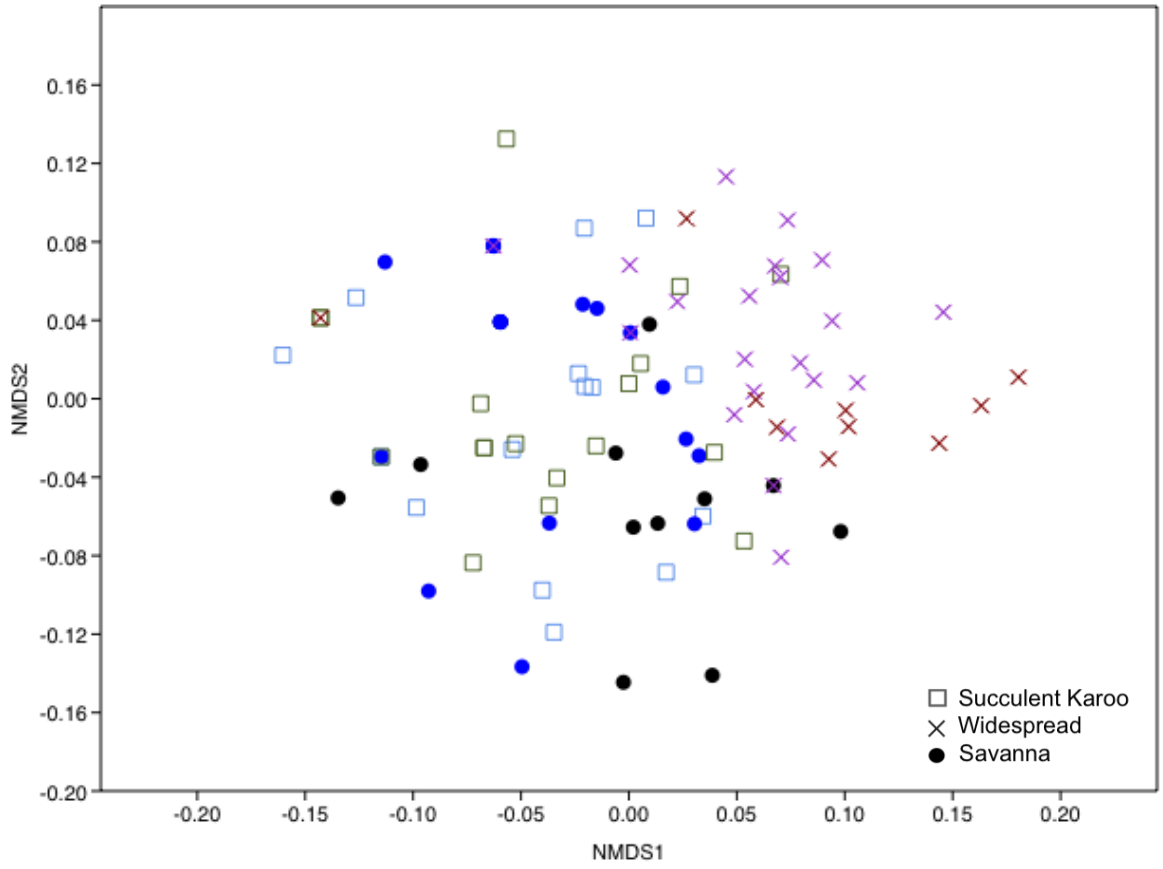

Figure 4.7: Non-metric multidimensional scaling (NMDS) analysis using Bray-Curtis distances representing plant species OTUs detected in bees from different regional groups. Bee species from the same region are represented with the same shape but with the different bee species having different colour codes. Bees from the Succulent Karoo are represented by squares, circles represent bees from the Savanna, and crosses represent widespread bees.

Different plant families identified in each biome group were compared to the plant family list created from the POSA database to confirm their occurrence in South Africa. All but three angiosperm families (87.5%) in all the regions studied were found to occur in the specific region detected. Magnoliaceae was not found in any of the regions searched. Additionally, Annonaceae and Caprifoliaceae were not found to occur in the Succulent Karoo. When plant families detected in pollen loads from different bee species were compared with the list of plants on which the particular bee species were observed, it was clear that Fabaceae is an important plant family for all six of the bee species (see supplementary Table S4.7). Fabaceae was the only plant family with both observed flower visits and detection in pollen loads in all six bee species. Amaranthaceae was detected by NGS in all six bee species, but only recorded in flower visits for *M. karooensis* and *M. venusta*. Additional recorded flower visits corresponding to detected families in pollen loads in this study includes Asteraceae in *M.*

murina, *M. felina*, *M. maxillosa*, *M. niveofasciata*, and *M. venusta*, and Solanaceae and Poaceae in *M. venusta*. The 20 plant families with the most species in each region were selected from the plant database and cross-referenced with the ITS2 sequence reference database to determine whether all these families are represented in the sequence database. Three of the top 20 most species-rich families in the Succulent Karoo, one family from the Savanna, and two widespread families were not represented at all in the ITS2 database.

4.5 Discussion

Pollen DNA barcoding is an emerging field and very limited research has been done on this topic in South Africa. Previously, taxonomic identifications of South African pollen data from the Sickel *et al*. (2015) ITS2 database was advised to be restricted to family level (Chapter 3). Here, pollen metabarcoding was used successfully on historic bee specimens from an insect collection to investigate their floral choice differences on plant family and species level. Pollen sequence data were classified into species OTUs and very useful species abundance data could be obtained, even though taxonomic identification remained unclarified. Clustering sequence reads obtained from pollen loads into OTUs based on 97% sequence identity allowed the plant species abundances in pollen loads from bees to be ascertained.

The higher floral diversity of the Succulent Karoo presents bees with more floral opportunities than the Savanna biome. One may therefore expect to see more plant families and species OTUs in pollen loads sampled from the bees from this region, rather than in those from the less florally diverse Savanna region. In contrast, the data presented here showed no significant difference in the number of either family, or species OTUs visited between the Succulent Karoo and Savanna biomes, meaning that the bees in the genus *Megachile* studied in the highly plant-diverse Western region of South Africa seem to be no more specialised than those to the east. More diverse plant taxa are available for bees to forage on in the Succulent Karoo biome than in the Savanna biome and this could indicate some level of specialisation in Succulent Karoo *Megachile* bees. These bees may visit a maximum number of plant taxa, regardless of the number of taxa available to them, whereas bee species from the Savanna biome visits approximately the same number of taxa, with fewer available to them from the start.

However, a distinct difference was shown by GLM in mean number of both plant family and species OTUs visited between widespread bees and those of the Succulent Karoo and Savanna. NMDS analysis shows that the widespread bee group is also characterised by a different pollen assemblage than the other two biome groups. Rarefaction of species OTU data revealed that *M. venusta* might visit a wider range of plant taxa than the other bee

species, but the species ranges visited by the other five bee species were similar. The wide ranges of families and species OTUs detected in *M. niveofasciata* pollen loads could indicate that these bees visit a wide range of plant taxa per bee, but are more preferential on a species level. The higher number of taxa visited by the widespread bees can possibly be explained by their access to a wider variety of plants spread across the country, as opposed to the bees restricted to the Succulent Karoo or Savanna regions only being able to collect pollen from plant families available in those areas. It is important to note that *M. venusta* did not differ significantly from the Savanna species, *M. felina*. The mean number of Angiosperm families, as well as the mean number of all plant families visited by the widespread bee species did differ significantly from the other two groups of bees. The difference seems to be explained by the higher mean number of visits made by *M. niveofasciata*.

Most plant families identified in pollen samples from the widespread bees, Succulent Karoo, and Savanna occur in all three these areas. Three plant families identified in the Succulent Karoo could not be found in the POSA database, but species from the Magnoliaceae, for instance, have been shown to be readily for sale in South African nurseries (Chapter 3). The absence of these families in the POSA database does not necessarily mean that they do not occur in the Succulent Karoo, they are simply not recorded in the database or might be artificially introduced into residential areas, which is not covered by indigenous plant databases. The database was undergoing its first update since 2009 at the time of writing, which should hopefully provide a more accurate representation of actual plant distribution in South Africa.

An interesting finding was the identification of high proportions of Amaranthaceae in the overall representation of both read and detection counts in all three groups investigated. In the arid Western region of South Africa, the three plant families most visited by bees are Fabaceae, Asteraceae and Aizoaceae (Gess and Gess, 2014). Amaranthaceae was not in the top ten most bee-visited families, or in the list of the 15 largest plant families in the area. Only seven percent of the total visits bees made in the study were to Amaranthaceae, of which the bee family with the most visits was Megachilidae (Gess and Gess, 2014). *Megachile venusta* were indeed recorded to visit flowers of Amaranthaceae by both Gess and Gess (2014), as well as Eardley and Urban (2010, Table S4.7). The presence of this family in the Succulent Karoo is not that peculiar. Amaranthaceae is one of the top ten families in the Tankwa National Park (Steyn *et al*., 2013) and in the top 20 families within the Extra Cape Subregion (Snijman, 2013). Its inclusion in the top 20 is not deemed too unusual in these dry areas, as is also seen in Australia and Eurasia. It was also found to be one of the dominant families in the Gannaveld (Wheeler, 2010). Unfortunately, no similar study of the same magnitude has been done for the Savanna region of South Africa.

As Asteraceae is the largest plant family in South Africa and the Succulent Karoo (Cowling *et al*., 1999), it was expected that it would be one of the most highly represented plant families identified in pollen from all biome groups, and particularly in pollen from the Succulent Karoo specimens. Interestingly, according to Gess and Gess (2014) varying rates of flower visitation has been observed between the three tribes in the bee subfamily Megachilinae. The Megachilini tribe visits Fabaceae flowers markedly more than the other two tribes, and the reverse is true for their visits to Asteraceae. Also, a low flower visitation rate of only 20% was observed in the Megachilini for Asteraceae (Gess and Gess, 2014). Our results showed that Fabaceae was present in the top five most visited families for all three groups but Asteraceae was only a highly represented family in pollen from widespread bees. However, these bees visiting Asteraceae flowers only for nectar could explain these differences. Many more visits to Fabaceae species than to Asteraceae species were also recorded previously for the six bee species studied (Table S4.7).

Grasses (Poaceae) were another highly represented family in all three studied groups. There have been reports over many years that bees also pollinate grasses (Bogdan, 1962; Koshy *et al*., 2001; Koshy and Harikumar, 2001; Pojar, 1973; Soderstrom and Calderon, 1971). The Poaceae family is very highly represented in the Savanna and is one of the 15 largest families in the Succulent Karoo (Gess and Gess, 2014) and it is thus plausible that these *Megachile* species could collect pollen from this family.

Pteridaceae was identified in almost all pollen samples and mostly in high proportions as well. The legitimacy of this identification is under contention as it is possible that the entries in the underlying reference sequence database originating from NCBI were problematic (Chapter 3). As Pteridaceae is a family that has 7,643 species across South Africa and 331 species occur in the Succulent Karoo and 2,177 in the Savanna (Germishuizen *et al*., 2003), it is quite plausible that airborne fern spores are included in pollen loads. Also, the presence of fern spores in pollen metabarcoding studies of honey (Hawkins *et al*., 2015; Valentini *et al*., 2010) confirm the possibility of bees foraging on spores, even though its presence was detected in less samples and in lower frequencies.

When comparing the plant families identified using the pollen loads from bees in this study to the plant visits recorded in Table S4.7, several of the families overlapped. In the widespread group, Amaranthaceae, Asteraceae, Fabaceae, Poaceae, Solanaceae, Brassicaceae and Anacardiaceae were identified by metabarcoding pollen, as well as through observational studies. Visits to Fabaceae and Asteraceae, and Amaranthaceae, Fabaceae and Asteraceae

were identified by both methods for bees from the Savanna, and bees from the Succulent Karoo, respectively. The higher number of overlapping families identified by both methods in the widespread group could be due the higher mean number of plant families identified by metabarcoding in this group overall. Of the plant species visits made by the six *Megachile* bee species (Table S4.7), less than half (42.7%) have at least one ITS2 sequence entry in GenBank. The families to which these species belong, however, are mostly well represented. The only visited families not represented in the ITS2 reference database used for identification are Mesembryanthemaceae and Asclepiadaceae. These two families would therefore not be detected in any mixed origin pollen load when identified against this reference database. ITS2 amplification is variable across plant groups (Chen *et al*., 2010), which means that by using only one set of primers in barcoding pollen, some plant family and species OTU identifications could be missed, which could also explain not detecting some families by NGS of pollen loads that were identified through observation.

This study has illustrated that it is possible to use historical insect specimens for pollen sampling in order to answer questions about floral choice, especially since the time since bee collection did not have any significant effect on the mean ranges of plant taxa observed. This indicates that pollen DNA was sufficiently preserved within the insect collection used and this increases the value of these collections. Scheper *et al*. (2014) analysed pollen loads from similarly stored museum specimens to investigate the effects of different factors on bee species decline in the Netherlands. They found a significant effect for time, where the size of floral ranges observed in pollen loads from bees prior to 1950 played a key role in bee population trends. There are limitations in the present study that need to be taken into account when considering the results. First, the results obtained were contingent upon the bee species chosen for study. All six of the species chosen were within the same genus (Eardley, 2013, 2012). Choosing bee specimens from different subfamilies or tribes within the same family might produce different results, as could the choice of completely different bee families. Second, not all specimens for the species of interest in the collection necessarily contained pollen for sampling, which means all specimens with pollen loads were selected for the study. In this study, it meant that some original sampling dates and localities were overrepresented in some of the bee species, in particular in *M. karooensis* and *M. murina*. In the widespread bee *M. niveofasciata*, specimens with pollen loads were coincidentally all caught in the Western region of South Africa, whereas *M. venusta* had a broader spread across the country. A limited spread of data hampers the ability to infer on floral choice differences on a species-wide scale. The number of plant families detected per specimen for, in particular the widespread group, could be influenced dramatically when specimens were uniformly spread

across the region. The plant taxa occurring naturally in this region would therefore be represented more robustly in the sampling. A uniform spread of sampling sites across the regions studied would have been ideal, but could not be obtained here due to the limiting factor of pollen availability on historic specimens.

4.6 Conclusion

In summary, this study showed that historic insect specimens could be used to answer questions about floral choice in bees from three different geographic areas. It was possible to show that bee species from the Succulent Karoo and Savanna do not differ significantly in the number of plant families or species that they visit, and that widespread bees tend to visit more plant species OTUs than ones from the aforementioned areas. Clustering sequence reads into species OTUs allowed floral choice ranges to be investigated in more detail than if only family level classifications were to be performed. However, bias in the specimen localities and dates of collection in this study likely influenced the number of plant taxa detected per geographic region. A larger sample size and a more evenly distributed spread of bee specimens sampled for pollen should ideally be used to overcome this bias.

4.7 Acknowledgements

Charles Hefer and Genevieve Thompson for their assistance with bioinformatics.

4.8 References

Bell, K., de Vere, N., Keller, A., Richardson, R., Gous, A., Burgess, K., *et al*., 2016. Pollen DNA barcoding: Current applications and future prospects. Genome 59, 629–640. doi:10.1139/gen-2015-0200.

Bogdan, A.V., 1962. Grass pollination by bees in Kenya. Proc. Linn. Soc. Lond. 173, 57–60.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, btu170. doi:10.1093/bioinformatics/btu170.

Brown, S.P., Veach, A.M., Rigdon-Huss, A.R., Grond, K., Lickteig, S.K., Lothamer, K., *et al*., 2015. Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecol. 13, 221–225. doi:10.1016/j.funeco.2014.08.006.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., *et al*., 2015. A DNA barcoding approach to identify plant species in multiflower honey. Food. Chem. 170, 308–315. doi:10.1016/j.foodchem.2014.08.060.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. doi:10.1038/nmeth.f.303.

Cowling, R.M., Esler, K.J., Rundel, P.W., 1999. Namaqualand, South Africa–an overview of a unique winter-rainfall desert ecosystem. Plant Ecol. 142, 3–21.

Da Silva, J.M., Willows-Munro, S., 2016. A review of over a decade of DNA barcoding in South Africa: a faunal perspective. Afr. Zool. 51. doi:10.1080/15627020.2016.1151377.

Eardley, C., 2013. A taxonomic revision of the southern African leaf-cutter bees, *Megachile* Latreille *sensu stricto* and *Heriadopsis* Cockerell (Hymenoptera: Apoidea: Megachilidae). Zootaxa 3601. doi:10.11646/zootaxa.3601.1.1.

Eardley, C.D., 2012. A taxonomic revision of the southern African species of dauber bees in the genus *Megachile* Latreille (Apoidea:Megachilidae). Zootaxa 3460, 1–139.

Eardley, C.D., 1989. Diversity and endemism of southern African bees. Plant Protection News 18, 1–2.

Eardley C.D., Urban R.P. 2010 Catalogue of Afrotropical bees (Hymenoptera: Apoidea:Apiformes). Magnolia Press, Auckland.

Eardley, C.D., Gikungu, M., Schwarz, M.P., 2009. Bee conservation in Sub-Saharan Africa and Madagascar: diversity, status and threats. Apidologie 40, 355–366. doi:10.1051/apido/2009016.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi:10.1093/bioinformatics/btq461.

Germishuizen, G., Meyer, N.L., National Botanical Institute (South Africa) (Eds.), 2003. Plants of southern Africa: an annotated checklist, Strelitzia. National Botanical Institute, Pretoria.

Gess, S.K., Gess, F.W., 2014. Wasps and bees in southern Africa. South African National Biodiversity Institute, Pretoria.

Hamer, M., 2013. A national strategy for zoological taxonomy (2013-2020). South African National Biodiversity Institute, Pretoria.

Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological statistics software package for education.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., *et al*., 2015. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10, e0134735. doi:10.1371/journal.pone.0134735.

Johnson, S.D., 1997. Pollination ecotypes of *Satyrium hallackii* (Orchidaceae) in South Africa. Bot. J. Linn. Soc. 123, 225–235. doi:10.1111/j.1095-8339.1997.tb01415.x.

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., Rost, S., *et al*., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 17, 558–566. doi:10.1111/plb.12251.

Koshy, K.C., Harikumar, D., 2001. Reproductive biology of *Ochlandra scriptoria*, an endemic reed bamboo of the Western Ghats, India. Bamboo Sci. Culture 15, 1–7.

Koshy, K.C., Harikumar, D., Narendran, T.C., 2001. Insect visits to some bamboos of the Western Ghats, India. Curr. Sci. 81, 833–838.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., *et al*., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol. Ecol. Res. 15, 8–16. doi:10.1111/1755-0998.12288.

Kuhlmann, M., 2009. Patterns of diversity, endemism and distribution of bees (Insecta: Hymenoptera: Anthophila) in southern Africa. S. Afr. J. Bot. 75, 726–738. doi:10.1016/j.sajb.2009.06.016.

Low, A.B., Rebelo, A.G. (Eds.), 1996. Vegetation of South Africa, Lesotho and Swaziland. Department of Environmental Affairs and Tourism, Pretoria.

McFrederick, Q.S., Rehan, S.M., 2016. Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. Mol. Ecol. 25, 2302–2311. doi:10.1111/mec.13608.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*., 2016. vegan: Community Ecology Package. R package version 2.3-4.

Packer, L., Zayed, A., Grixti, J.C., Ruz, L., Owen, R.E., Vivallo, F., *et al*., 2005. Conservation genetics of potentially endangered mutualisms: Reduced levels of genetic variation in specialist versus generalist bees. Conserv. Biol. 19, 195–202. doi:10.1111/j.1523-1739.2005.00601.x.

Pauw, A., 2006. Floral syndromes accurately predict pollination by a specialized oil-collecting bee (*Rediviva peringueyi*, Melittidae) in a guild of South African orchids (Coryciinae). Am. J. Bot. 93, 917–926. doi:10.3732/ajb.93.6.917.

Pauw, A., Stanway, R., 2015. Unrivalled specialization in a pollination network from South Africa reveals that specialization increases with latitude only in the southern Hemisphere. J. Biogeogr. 42, 652–661. doi:10.1111/jbi.12453.

Pennisi, E., 2000. Taxonomic Revival. Science 289, 2306–2308.

Pojar, J., 1973. Pollination of typically anemophilous salt marsh plants by bumble bees, *Bombus terricola occidentalis* Grne. Am. Midl. Nat. 89, 448–451. doi:10.2307/2424049.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahl, M., 2008. Microscopic identification and purity determination of pollen grains, in: Jones, M.G., Lympany, P. (Eds.), Allergy Methods and Protocols. Humana Press, Totowa, NJ.

Richardson, R.T., Lin, C.-H., Quijia, J.O., Riusech, N.S., Goodell, K., Johnson, R.M., 2015a. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl. Plant Sci. 3, 1500043. doi:10.3732/apps.1500043.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015b. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl. Plant Sci. 3, 1400066. doi:10.3732/apps.1400066.

Scheper, J., Reemer, M., van Kats, R., Ozinga, W.A., van der Linden, G.T.J., Schaminée, J.H.J., *et al*., 2014. Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in The Netherlands. Proc. Nat. Acad. Sci. 111, 17552–17557. doi:10.1073/pnas.1412973111.

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794–1805. doi:10.1111/j.1365-294X.2012.05538.x.

Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., *et al*., 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol. 15. doi:10.1186/s12898-015-0051-y.

Snijman, D.A. (Ed.), 2013. Plants of the Greater Cape floristic region, Vol. 2: the Extra Cape flora, Strelitzia. South African National Biodiversity Institute, Pretoria.

Soderstrom, T.R., Calderon, C.E., 1971. Insect pollination in tropical rain forest grasses. Biotropica 3, 1–16. doi:10.2307/2989701.

Steyn, H.M., Bester, S.P., Bezuidenhout, H., 2013. An updated plant checklist for Tankwa Karoo National Park, South Africa. S. Afr. J. Bot. 88, 247–251. doi:10.1016/j.sajb.2013.07.018.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA barcoding for honey biodiversity. Diversity 2, 610–617. doi:10.3390/d2040610.

Van Wyk, A.E., Smith, G.F., 2001. Regions of floristic endemism in southern Africa: A review with emphasis on succulents. Umdaus Press, Hatfield.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microb. 73, 5261–5267. doi:10.1128/AEM.00062-07.

Wheeler, A.D., 2010. Impacts of degradation on critically endangered Oudtshoorn Gannaveld. University of the Western Cape.

Zayed, A., 2009. Bee genetics and conservation. Apidologie 40, 237–262. doi:10.1051/apido/2009026

4.9 Supplementary data

Supplementary data referred to in Chapter 4 are presented in the order mentioned in the text.

Table S4.1. Collection information of *Megachile karooensis* bee specimens used for pollen collection in this study. Available information regarding the specimen collection, such as date, province, GPS coordinates and collection locality are given for each sample. All bee specimens are housed at the National Insect Collection, Agricultural Research Council, South Africa.

Bee collection identifier	Pollen sample identifier	Bee collection date	Province	GPS	Bee collection description
HYMA06233	b1	19.11.1982	Western Cape	33.23S 19.27E	Mitchell's Pass near Ceres
HYMA06345	b2	09.09.1987	Western Cape	31.13S 19.12E	Brandkop
HYMA05853/1	b3	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA05853/2	b4	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA05852	b5	27.09.1990	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA05853/3	b6	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA06359	b7	01.10.1990	Western Cape	30.09S 17.59E	Dassiefontein Farm near Kamieskroon
HYMA22079	b8	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29217	b9	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley

HYMA29218	b10	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29219	b11	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29220	b12	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29221	b13	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29222	b14	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29223	b15	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29224	b16	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29225	b17	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29226	b18	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29227	b19	05-07.09.1987	Western Cape	32.08S 19.14E	Clanwilliam District, Biedouw Valley
HYMA29228	b20	03.09.1987	Western Cape	NA	Namaqualand (Grid 2917DB)

Table S4.2. Collection information of *Megachile murina* bee specimens used for pollen collection in this study. Available information regarding the specimen collection, such as date, province, GPS coordinates and collection locality are given for each sample. All bee specimens are housed at the National Insect Collection, Agricultural Research Council, South Africa.

Bee collection identifier	Pollen sample identifier	Bee collection date	Province	GPS	Bee collection description
HYMA06346	d1	16.11.1984	Western Cape	32.08S 19.01E	Pakhuis Pass
HYMA06342	d2	11.09.1987	Western Cape	30.10S 18.01E	Kamiesberg
HYMA06233/1	d3	19.11.1982	Western Cape	33.23S 19.27E	Mitchell's Pass near Ceres
HYMA06233/2	d4	19.11.1982	Western Cape	33.23S 19.27E	Mitchell's Pass near Ceres
HYMA06233/3	d5	19.11.1982	Western Cape	33.23S 19.27E	Mitchell's Pass near Ceres
HYMA06316/1	d6	16.12.1988	Western Cape	33.19S 21.25E	North of Seweweekspoort
HYMA06316/2	d7	16.12.1988	Western Cape	33.19S 21.25E	North of Seweweekspoort
HYMA06316/3	d8	16.12.1988	Western Cape	33.19S 21.25E	North of Seweweekspoort
HYMA06343/1	d9	07.09.1987	Western Cape	32.08S 19.02E	Pakhuis Pass
HYMA06316/4	d10	16.12.1988	Western Cape	33.19S 21.25E	North of Seweweekspoort
HYMA06343/2	d11	07.09.1987	Western Cape	32.08S 19.02E	Pakhuis Pass
HYMA05855	d12	01.10.1990	Western Cape	30.09S 17.59E	Dassiefontein Farm near Kamieskroon

HYMA29232	d13	07.09.1987	Western Cape	32.08S 19.02E	Pakhuis Pass
HYMA29233	d14	07.09.1987	Western Cape	32.08S 19.02E	Pakhuis Pass
HYMA29234	d15	07.09.1987	Western Cape	32.08S 19.02E	Pakhuis Pass
HYMA22083	d16	27.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA29235	d17	27.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA29236	d18	16.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA22085	d19	03.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA29237	d20	03.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA22086	d21	25.09.1987	Western Cape	NA	Namaqualand (Grid 2917 DB)
HYMA22087	d22	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu
HYMA29238	d23	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu
HYMA29239	d24	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu
HYMA29240	d25	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu
HYMA29241	d26	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu

HYMA29231	d27	17.11.1982	Western Cape	33.48S 20.12E	Ouberg Pass, 24 km North East of Montagu

Table S4.3. Collection information of *Megachile maxillosa* bee specimens used for pollen collection in this study. Available information regarding the specimen collection, such as date, province, GPS coordinates and collection locality are given for each sample. All bee specimens are housed at the National Insect Collection, Agricultural Research Council, South Africa.

Bee collection identifier	Pollen sample identifier	Bee collection date	Province	GPS	Bee collection description
HYMA06176	c1	10.1979	Limpopo	NA	Mogoto Reserve near Zebediela
HYMA06202	c2	11.1976	North West	NA	Buffelspoort near Rustenburg
HYMA06232	c3	24.10.1975	Northern Cape	NA	Upington
HYMA21729	c4	19.10.1980	Gauteng	NA	Pumulani
HYMA21731	c5	19.12.2003	KwaZulu-Natal	NA	Bisley Nature Reserve, Pietermaritzburg
HYMA06223	c6	17.11.1972	Limpopo	NA	Rooiberg, east of Thabazimbi
HYMA05911	c7	25.02.1993	Free State	27.40S 25.45E	Sandveld Nature Reserve
HYMA06195	c8	22.10.1914	Free State	NA	Bloemfontein
HYMA06194	c9	17.10.1914	Free State	NA	Bloemfontein
HYMA06222	c10	19.01.1984	Mpumalanga	24.59S 31.55E	Skukuza, Kruger National Park
HYMA21737	c11	22-23.02.1993	Free State	27.15S 27.41E	Koppies Dam Nature Reserve
HYMA06226	c12	04.1975	North West	NA	Buffelspoort

HYMA06205/1	c13	31.03.1972	Limpopo	NA	Ellisras
HYMA06205/2	c14	31.03.1972	Limpopo	NA	Ellisras
HYMA06205/3	c15	31.03.1972	Limpopo	NA	Ellisras
HYMA06175	c16	01.11.1969	Transvaal[1]	NA	Breedsnek Pass
HYMA06184	c17	06.04.1962	Limpopo	NA	Ellisras
HYMA06176	c18	10.1979	Limpopo	NA	Mogoto Reserve near Zebediela
HYMA21741	c19	17.02.1981	Gauteng	25.24S 28.06E	Soutpan, Pretoria district
HYMA06308	c20	10.12.1978	Northern Cape	NA	Olifantshoek
HYMA21742	c21	22-23.02.1993	Free State	27.15S 27.41E	Koppies Dam Nature Reserve
HYMA06179	c22	14.03.1969	Eastern Cape	NA	Graaff Reinet
HYMA06224	c23	02.12.1981	Limpopo	24.13S 29.30E	Chuniespoort
HYMA06305/1	c24	18-29.10.1989	Northern Cape	26.25S 20.37E	Kalahari Gemsbok Park, Twee Rivieren
HYMA06305/2	c25	18-29.10.1989	Northern Cape	26.25S 20.37E	Kalahari Gemsbok Park, Twee Rivieren
HYMA06305/3	c26	18-29.10.1989	Northern Cape	26.25S 20.37E	Kalahari Gemsbok Park, Twee Rivieren

HYMA06200/1	c27	10.1978	Limpopo	NA	Nylsvley Nature Reserve
HYMA06335	c28	19.12.1987	Limpopo	23.45S 27.49E	D'Nyala Nature Reserve, Ellisras District
HYMA06200/2	c29	10.1978	Limpopo	NA	Nylsvley Nature Reserve
HYMA06200/3	c30	10.1978	Limpopo	NA	Nylsvley Nature Reserve
HYMA06181	c31	24-28.11.1980	Limpopo	24.37S 27.23E	Ben Alberts Nature Reserve, Thabazimbi
HYMA06200/4	c32	10.1978	Limpopo	NA	Nylsvley Nature Reserve

[1] The borders of South Africa's provinces have changed, and with it also the province names. Breedsnek Pass used to fall within one province, Transvaal, but now stretches through Gauteng and the North West Province. Without GPS coordinates, it was not possible to ascertain in which of the renamed provinces the sample was collected.

Table S4.4. Collection information of *Megachile felina* bee specimens used for pollen collection in this study. Available information regarding the specimen collection, such as date, province, GPS coordinates and collection locality are given for each sample. All bee specimens are housed at the National Insect Collection, Agricultural Research Council, South Africa.

Bee collection identifier	Pollen sample identifier	Bee collection date	Province	GPS	Bee collection description
HYMA06277/1	a1	25-26.10.1984	Mpumalanga	24.32S 30.47E	Blyderivierspoort Nature Reserve
HYMA06277/2	a2	25-26.10.1984	Mpumalanga	24.32S 30.47E	Blyderivierspoort Nature Reserve
HYMA06277/3	a3	25-26.10.1984	Mpumalanga	24.32S 30.47E	Blyderivierspoort Nature Reserve
HYMA06277/4	a4	25-26.10.1984	Mpumalanga	24.32S 30.47E	Blyderivierspoort Nature Reserve
HYMA06301	a5	09.03.1990	Limpopo	22.14S 29.59E	Near Beitbridge
HYMA06272	a6	08-12.12.1989	Limpopo	23.45S 27.49E	D'Nyala Nature Reserve, Ellisras District
HYMA06297/1	a7	07.03.1990	Limpopo	30.03S 22.23E	Messina Nature Reserve
HYMA06297/2	a8	07.03.1990	Limpopo	30.03S 22.23E	Messina Nature Reserve
HYMA06156	a9	10-11.12.1979	Limpopo	24.39S 28.42E	Nylsvlei Nature Reserve
HYMA06146	a10	12.1974	North West	NA	Mokopane
HYMA06464	a11	01.1987	Gauteng	25.24S 28.06E	Soutpan, Pretoria district
HYMA22025	a12	04.1970	Limpopo	NA	Letsitele

HYMA22045	a13	24.01.1985	Limpopo	22.36S 31.17E	Machayi Pan 418 m, Kruger National Park
HYMA29229	a14	24.01.1985	Limpopo	22.36S 31.17E	Machayi Pan 418 m, Kruger National Park
HYMA29230	a15	20-24-01.1985	Limpopo	22.26S 31.12E	Pafuri 264 m, Kruger National Park
HYMA22063	a16	20-24.01.1985	Limpopo	22.26S 31.12E	Pafuri 264 m, Kruger National Park
HYMA22076	a17	12.01.1966	Limpopo	NA	Shipudza (Grid 2230 BD), Kruger National Park

Table S4.5. Collection information of *Megachile niveofasciata* bee specimens used for pollen collection in this study. Available information regarding the specimen collection, such as date, province, GPS coordinates and collection locality are given for each sample. All bee specimens are housed at the National Insect Collection, Agricultural Research Council, South Africa.

Bee collection identifier	Pollen sample identifier	Bee collection date	Province	GPS	Bee collection description
HYMA29242	f1	07.10.1987	Northern Cape	NA	Richtersveld National Park, Road Khubus-Ochta near Vyfsusters-Mt (Grid 2816 BB)
HYMA29243	f2	07.10.1987	Northern Cape	NA	Richtersveld National Park, Road Khubus-Ochta near Vyfsusters-Mt (Grid 2816 BB)
HYMA29244	f3	02.02.2000	Northern Cape	29.04S 19.24E	8 km North East of Pofadder
HYMA29245	f4	02.02.2000	Northern Cape	29.04S 19.24E	8 km North East of Pofadder
HYMA06323/1	f5	09.12.1990	Western Cape	33.16S 19.43E	Verlorenvlei near Ceres
HYMA06323/2	f6	09.12.1990	Western Cape	33.16S 19.43E	Verlorenvlei near Ceres
HYMA06323/3	f7	09.12.1990	Western Cape	33.16S 19.43E	Verlorenvlei near Ceres
HYMA06323/4	f8	09.12.1990	Western Cape	33.16S 19.43E	Verlorenvlei near Ceres

HYMA06352	f9	17.11.1984	Western Cape	31.59S 19.14E	Doringbos
HYMA06365	f10	02.11.1992	Western Cape	29.27S 17.03E	Kwakanap Road on Kleinsee Road 20 km, South East of Port Nolloth

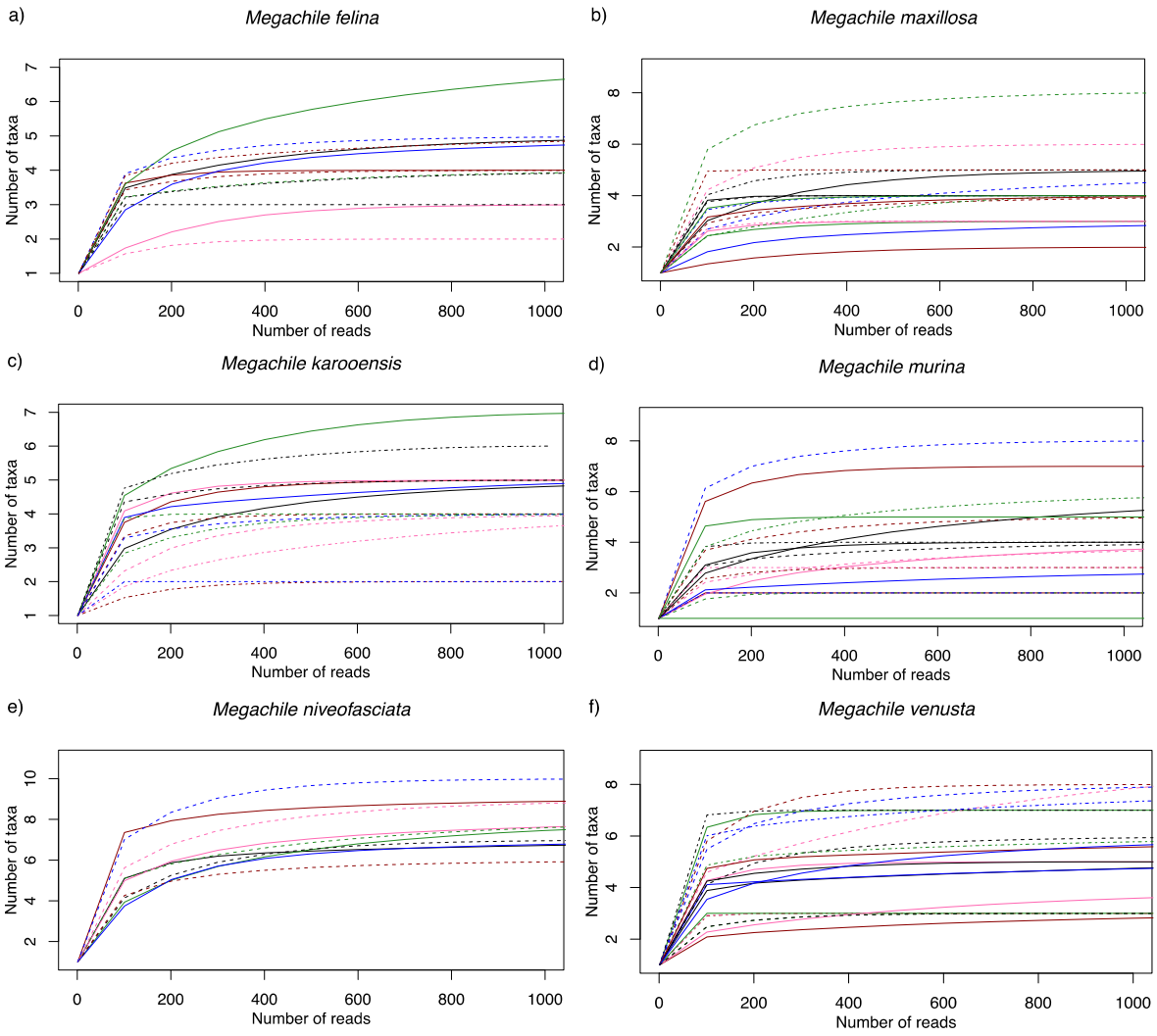

Figure S4.1. Rarefaction curves for each bee species to determine whether sequence saturation has been reached with the amount of reads sequenced per pollen sample. The lines for all samples collected from specimens from all six of the bee species flattened when 1,000 sequence reads were reached, indicating that all possible plant families represented in the pollen samples have been identified. a) and b) represents bee species occurring in the Savanna biome, c) and d) represents bee species occurring exclusively in the Succulent Karoo biome, and e) and f) represents bee species occurring all over South Africa (widespread group).

Table S4.6. Uncorrected significance (p) values between the different groups compared in ANOSIM.

	M. felina	*M. karooensis*	*M. maxillosa*	*M. murina*	*M. venusta*	*M. niveofasciata*
M. felina	-	0,514	0,0275	0,07499	0,0001	0,0002
M. karooensis	0,514	-	0,3291	0,605	0,0001	0,0016
M. maxillosa	0,0275	0,3291	-	0,2081	0,0001	0,0001
M. murina	0,07499	0,605	0,2081	-	0,0001	0,0004
M. venusta	0,0001	0,0001	0,0001	0,0001	-	0,0003
M. niveofasciata	0,0002	0,0016	0,0001	0,0001	0,0003	-

Table S4.7. A list of plant species and families on which visits from the six studied bee species have been recorded.

Bee species	Plant species visited	Plant family visited	Additional information and references
Megachile karooensis	*Anchusa capensis*	Boraginaceae	Eardley and Urban, 2010
	Aspalathus linearis	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus pulicifolia	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus spinescens	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus sp.	Fabaceae	Eardley, 2012
	Blepharis extenuata	Acanthaceae	Eardley and Urban, 2010; Eardley, 2012
	Disa filicornis	Orchidaceae	Eardley and Urban, 2010; Eardley, 2012
	Hermannia trifurca	Malvaceae	Eardley and Urban, 2010
	Hermbstaedtia glauca	Amaranthaceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia pungens	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia sericea	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia spinescens	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Moraea tripetala	Iridaceae	Eardley and Urban, 2010; Eardley, 2012
	Nemesia sp.	Scrophulariaceae	Eardley and Urban, 2010; Eardley, 2012
	Pelargonium capitatum	Geraniaceae	Gess and Gess, 2014
	Pelargonium sp.	Geraniaceae	Eardley and Urban, 2010
	Polygala virgata	Polygalaceae	Eardley and Urban, 2010
	Polymita albiflora	Aizoaceae	Eardley and Urban, 2010; Eardley, 2012
	Pteronia incana	Asteraceae	Eardley and Urban, 2010; Eardley, 2012
	Wiborgia monoptera	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Zygophyllum meyeri (*Zygophyllum foetidum*)	Zygophyllaceae	Eardley and Urban, 2010; Eardley, 2012
	Zygophyllum retrofractum	Zygophyllaceae	Eardley and Urban, 2010; Eardley, 2012
	-	Fabaceae	Particularly Papilionaceae, Eardley, 2012
Megachile murina	*Albuca sp.*	Asparagaceae	Eardley and Urban, 2010
	Anchusa capensis	Boraginaceae	Eardley and Urban, 2010
	Aridaria brevicarpa	Aizoaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus chortophila	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus linearis	Fabaceae	Eardley and Urban, 2010; Eardley, 2012

	Aspalathus pulicifolia	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus spinescens	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Aspalathus sp.	Fabaceae	Gess and Gess, 2014
	Ballota africana	Lamiaceae	Eardley and Urban, 2010
	Chrysanthemum sp.	Asteraceae	Eardley, 2012
	Hermannia disermifolia	Malvaceae	Eardley and Urban, 2010
	Herrea sp.	Aizoaceae	Eardley and Urban, 2010
	Hirpicium alienatum	Asteraceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia pungens	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia sericea	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Lebeckia simsiana	Fabaceae	Eardley and Urban, 2010
	Lebeckia sp.	Fabaceae	Eardley, 2012
	Lotononis bainesii	Fabaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	Polygala virgata	Polygalaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	Polygala sp.	Polygalaceae	Gess and Gess, 2014
	Polymita albiflora	Aizoaceae	Eardley and Urban, 2010
	Prenia pallens	Aizoaceae	Eardley and Urban, 2010
	Sarcocaulon crassicaule	Geraniaceae	Eardley and Urban, 2010; Eardley, 2012
	Stachys aurea	Lamiaceae	Single specimen, Eardley and Urban, 2010; Gess and Gess, 2014
	Wahlenbergia sp.	Campanulaceae	Eardley, 2012
	Wiborgia monoptera	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Wiborgia sp.	Fabaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	Zygophyllum divaricatum	Zygophyllaceae	Eardley and Urban, 2010
	Zygophyllum meyeri (*Zygophyllum foetidum*)	Zygophyllaceae	Eardley and Urban, 2010; Eardley, 2012
	-	Fabaceae	Close association, Papilionaceae Eardley and Urban, 2010; Gess and Gess, 2014
	-	Lamiaceae	Single female, Gess and Gess, 2014
Megachile felina	*Cleome angustifolia*	Cleomaceae	Eardley and Urban, 2010
	Crotalaria argyraea	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Crotalaria podocarpa	Fabaceae	Eardley and Urban, 2010
	Crotalaria sp.	Fabaceae	Gess and Gess, 2014
	Eulophia streptopetala	Orchidaceae	Eardley and Urban, 2010
	Otoptera burchellii	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Tephrosia oxygona	Fabaceae	Eardley and Urban, 2010; Gess and Gess, 2014

	-	Asteraceae	Eardley and Urban, 2010; Eardley, 2012
Megachile maxillosa	*Acacia horrida*	Fabaceae	Eardley, 2012
	Acacia karroo	Fabaceae	Eardley and Urban, 2010
	Acacia nilotica	Fabaceae	Eardley, 2012
	Acacia senegal	Fabaceae	Eardley and Urban, 2010
	Adenolobus pechuelii	Fabaceae	Eardley and Urban, 2010
	Asclepias buchenaviana	Asclepiadaceae	Eardley and Urban, 2010; Eardley, 2012
	Blepharis capensis	Acanthaceae	Eardley and Urban, 2010
	Blepharis sp.	Acanthaceae	Gess and Gess, 2014
	Cleome elegantissima	Cleomaceae	Eardley and Urban, 2010
	Cleome suffruticosa	Cleomaceae	Eardley and Urban, 2010
	Cleome sp.	Cleomaceae	Gess and Gess, 2014
	Crotalaria argyraea	Fabaceae	Eardley and Urban, 2010
	Crotalaria dinteri	Fabaceae	Eardley and Urban, 2010
	Crotalaria podocarpa	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Crotolaria virgultalis	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	Crotalaria sp.	Fabaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	Gomphocarpus filiformis	Asclepiadaceae	Gess and Gess, 2014
	Monechma genistifolium	Acanthaceae	Eardley and Urban, 2010
	Monechma mollissimum	Acanthaceae	Gess and Gess, 2014
	Monechma spartioides	Acanthaceae	Eardley and Urban, 2010
	Monechma sp.	Acanthaceae	Eardley and Urban, 2010; Eardley, 2012; Gess and Gess, 2014
	Polygala leptophylla	Polygalaceae	Eardley and Urban, 2010
	Sesamum triphyllum	Pedaliaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	Sesamum sp.	Pedaliaceae	Eardley and Urban, 2010; Eardley, 2012; Gess and Gess, 2014
	Tephrosia oxygona	Fabaceae	Eardley and Urban, 2010; Gess and Gess, 2014
	-	Acanthaceae	Gess and Gess, 2014
	-	Apocynaceae	Gess and Gess, 2014
	-	Asteraceae	Single male specimen, Gess and Gess, 2014
	-	Brassicaceae	Gess and Gess, 2014
	-	Fabaceae	Gess and Gess, 2014
	-	Pedaliaceae	Gess and Gess, 2014
	-	Polygalaceae	Gess and Gess, 2014
Megachile niveofasciata	*Asclepias buchenaviana*	Asclepiadaceae	Eardley and Urban, 2010
	Aspalathus chortophila	Fabaceae	Eardley and Urban, 2010

	Berkheya sp.	Asteraceae	Eardley, 2012
	Cleome paxii	Cleomaceae	Eardley and Urban, 2010
	Geigeria sp.	Asteraceae	Eardley and Urban, 2010
	Hermannia modesta	Malvaceae	Eardley and Urban, 2010
	Lebeckia multiflora	Fabaceae	Eardley and Urban, 2010
	Lessertia macrostachya	Fabaceae	Eardley and Urban, 2010
	Limeum fenestratum	Molluginaceae	Eardley and Urban, 2010
	Maerua gilgii	Capparaceae	Eardley and Urban, 2010
	Maerua schinzii	Capparaceae	Gess and Gess, 2014
	Pentzia sphaerocephala	Asteraceae	Eardley and Urban, 2010
	Polymita albiflora	Aizoaceae	Eardley and Urban, 2010; Eardley, 2012
	Prenia pallens	Aizoaceae	Eardley and Urban, 2010
	Psilocaulon salicornioides	Aizoaceae	Eardley and Urban, 2010
	Sesamum triphyllum	Pedaliaceae	Eardley and Urban, 2010
	Sisyndite spartea	Zygophyllaceae	Eardley, 2012
	Zygophyllum simplex	Zygophyllaceae	Eardley and Urban, 2010
	Wiborgia sp.	Fabaceae	Eardley and Urban, 2010; Eardley, 2012
	-	Mesembryanthemaceae	Eardley and Urban, 2010
Megachile venusta	*Acacia caffra*	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Acacia karroo	Fabaceae	Eardley and Urban, 2010
	Allium sp.	Amaryllidaceae	Eardley and Urban, 2010; Eardley, 2013
	Aptosimum procumbens	Scrophulariaceae	Eardley and Urban, 2010; Eardley, 2013
	Aspalathus subtingens	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Bergia glomerata	Elatinaceae	Eardley and Urban, 2010
	Bulbine frutescens	Asphodelaceae	Eardley and Urban, 2010
	Eucalyptus sp.	Myrtaceae	Eardley and Urban, 2010; Eardley, 2013
	Foeniculum vulgare	Apiaceae	Eardley and Urban, 2010
	Grewia occidentalis	Malvaceae	Eardley and Urban, 2010; Eardley, 2013
	Lycium sp.	Solanaceae	Eardley and Urban, 2010; Eardley, 2013
	Medicago sativa	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Melolobium candicans	Fabaceae	Eardley and Urban, 2010
	Polygala pinifolia	Polygalaceae	Eardley and Urban, 2010; Eardley, 2013
	Portulaca sp.	Portulacaceae	Eardley and Urban, 2010; Eardley, 2013
	Prosopis chilensis	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Rhus sp.	Anacardiaceae	Eardley and Urban, 2010; Eardley, 2013

	Tribulus sp.	Zygophyllaceae	Eardley and Urban, 2010; Eardley, 2013
	Trifolium sp.	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Vernonia sp.	Asteraceae	Eardley and Urban, 2010; Eardley, 2013
	Vicia faba	Fabaceae	Eardley and Urban, 2010; Eardley, 2013
	Zea mays	Poaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Amaranthaceae	Goosefoot family Chenopodiaceae species observed, Eardley and Urban, 2010; Eardley, 2013
	-	Asteraceae	Eardley and Urban, 2010; Eardley, 2013
	-	Brassicaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Fabaceae	Eardley and Urban, 2010
	-	Iridaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Liliaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Loganiaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Mesembryanthemaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Myrtaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Poaceae	Eardley and Urban, 2010; Eardley, 2013
	-	Solanaceae	Eardley and Urban, 2010; Eardley, 2013

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

5.1 Introduction

It is important to understand the interactions between pollinators and plants, not only because of the importance of pollination in agriculture, but also for biodiversity conservation purposes. Traditional methods of studying these interactions are time-consuming and inefficient, but more advanced molecular techniques have recently become more accessible and presented the opportunity to examine plant-pollinator interactions by genetic analyses.

This study refined the techniques to provide accurate DNA-based plant taxon identifications from pollen sampled directly from bee specimens housed in a historic insect collection. Sampling from an insect collection provides specimens that are spread across geographical space and a temporal line, giving depth to the samples considered in the analysis. Refining metabarcoding techniques to identify pollen from historic specimens meant that every step from sampling pollen, to the bioinformatics procedures needed to be carefully considered. These techniques could then be used to investigate plant-pollinator interactions in historic Megachilid bees by sequencing their pollen loads.

In this chapter, the key findings of this study are summarised and the implications of the research reviewed. Further, opportunities for future research are explored.

5.2 Aims and objectives

This study aimed to develop appropriate methods to accurately identify the plant origins of pollen sampled from historic bee specimens and to subsequently evaluate these methods. After development, the study aimed to use the methods to investigate differences in floral choice in long-tongued bees from the Succulent Karoo and the Savanna biomes of South Africa, comparing these regions to bees occurring across the country.

To meet these aims, the objectives of the study were to: i) Develop and improve the techniques used in DNA barcoding applied to historic bee specimens to be able to identify provenance in pollen, ii) Use the developed techniques to identify the plant origins of pollen sampled from bee specimens housed in a historic collection and iii) Investigate the floral choices of bees from different areas in South Africa (Succulent Karoo, Savanna and widespread).

5.3 Summary of main findings

The development of appropriate techniques to utilise genetic analyses for identification of pollen origins sampled from historic insect collection, and the application of these techniques, the first aim of the study, yielded the following results:

- The optimal method of DNA extraction for pollen sampled from historic specimens was to use the DNeasy® Plant Mini Kit (Qiagen) without disrupting the pollen with beads.
- The *rbcL* gene could not be amplified and sequenced with sufficient consistency and success from bee-collected-pollen samples to be used as a comparative locus for plant identification.
- ITS2 showed a considerably enhanced discriminatory power in plant identifications from pollen than ITS1.
- Approximately 250 reads were sufficient to identify all plant species with ITS1, but a maximum of two plant species could be detected using this barcode. Using ITS2, up to eight plant species could be detected within a single pollen load, and between 750 and 2,000 reads were sufficient to obtain this resolution.
- The sequence reference database used for classifications is not sufficiently representative of South African flora to identify pollen to species- or genus-level and therefore family-level classifications are recommended.
- Fungal presence in pollen samples was documented, with a specific importance of the genus *Malassezia*, likely transferred during handling of bee specimens.
- No relationship between the age of historic bee specimens and the success of pollen identification could be found.

The previously developed techniques were applied to investigate floral choice in bees from different areas within South Africa (the hyper-diverse Succulent Karoo, the Savanna and a widespread area) to address the remainder of the aims of the study. From these investigations, the following results were obtained:

- General linear models (GLM) indicated that there was a significant difference in the mean number of plant families and species visited between different geographic groups.
- Time since sampling of bees did not have a significant effect on the number of taxa detected in pollen loads.
- The bees from the Succulent Karoo and the Savanna did not differ significantly in the amount of plant families or species they visit.

- Bees from the widespread group showed a significant increase in plant families and species visited when compared to bees from the Succulent Karoo and Savanna groups, likely because they have access to a larger assortment of plants than bees in either of the other two groups. This increase could be attributed to significant differences between *M. niveofasciata* and all other bee species for plant families, as *M. venusta* did not differ significantly from any of the other bee species. On the mean number of plant species visited, both *M. niveofasciata* and *M. venusta* differed from all other bee species, with the exception of *M. venusta* not differing significantly from *M. felina*.
- Non-metric multidimensional scaling (NMDS) analyses showed that there is a small, but significant difference in the composition of pollen assemblages sampled from bees from different regional groups.
- The only group visiting Asteraceae on a noteworthy level was the widespread group of bees, which was surprising.
- Bees from the Succulent Karoo and the Savanna mainly visited Amaranthaceae, Poaceae, Fabaceae, and Polygonaceae families.

Sampling pollen loads from historic insect specimens occurring in different geographic regions allows biological questions to be answered, even though challenges remain.

5.4 Challenges

At the commencement of the study the only published pollen metabarcoding literature available was that of Valentini *et al*. (2010). The only other published study using pollen directly from pollinators that used genetic methods to identify the plant origins of pollen was that of Wilson *et al*. (2010). In their research Sanger sequencing was used to distinguish between the mixed origins of pollen samples, and not NGS as in metabarcoding studies. Working with very little literature guidance at the start resulted in a trial-and-error approach to the study. During the course of the study, a surge of pollen metabarcoding research was published; all examining fresh pollen in larger amounts in either honey, air, or bee nests (Hawkins *et al*., 2015; Keller *et al*., 2015; Kraaijeveld *et al*., 2015; McFrederick and Rehan, 2016; Richardson *et al*., 2015a, 2015b; Sickel *et al*., 2015). The DNA extraction of pollen is a notoriously difficult task (Bell *et al*., 2016), and working with limited pollen quantities and old samples complicated matters further.

As bee specimens were selected from the insect collection, and not all bees were carrying pollen loads, the number of viable specimens was reduced by approximately 75%. When bee specimens did have pollen loads, the quantities were often extremely low. DNA extraction from all pollen loads, regardless of quantity, produced immeasurable concentrations,

complicating all further steps in the procedure. Nonetheless, successful sequencing results were obtained even under these challenging conditions.

Sequenced products needed to be compared against a sequence reference database, because a simple BLAST search is not the best method of classification (Altschul *et al*., 1990). Most DNA metabarcoding studies, however, used this tool. That meant that a new reference database had to be created to conform to standards required by bioinformatics classification programs used, such as QIIME and RDP. Sequence data harvested from public databases such as NCBI are not always reliable (Harris, 2003) and this leads to incorrect identifications made in the classification procedure. The study published by Sickel *et al*. (2015) provided a sequence reference database for the ITS2 region that is structure-validated which largely overcame this problem. Despite this improvement, however, the lack of representation of South African flora in the ITS2 database lead to lower-level taxonomic classification being unachievable. It has been shown that multi-locus analyses in identification are advantageous (Richardson *et al*., 2015a) and this study aimed to include the plastid gene, *rbcL*. PCR and NGS of this gene were unsuccessful and could not be used in further analyses. This highlights the differences between working with fresh and historic samples. Making plant identifications up to family level with pollen metabarcoding is in essence comparable to using microscopic methods. Microscopic palynology is accurate to family level and would be cheaper to perform than metabarcoding. It would also be more time-consuming, however, and in areas of the world where sequence reference databases are more complete, identifications using metabarcoding can be accurately made to species level (Keller *et al.* 2015).

Limited viable specimen numbers within the selected species posed some challenges of its own. Selected specimens sometimes turned out to be all from the same date or the same sampling locality, which specifically posed a problem when analysing data from bees with widespread occurrance across South Africa. Due to some sampling sites of one bee species from the widespread group originating from one location, results cannot easily be generalised across the area. This was overcome by selecting two bee species in each geographic group to correct the potential sampling bias.

5.5 Future possibilities

There are many research opportunities stemming from this study. The development of a technique that can confidently identify the plant origins of pollen using historic insect collections as a sampling base is a big step forward in pollination research. Using historic collections could be invaluable in studying how plant communities, and plant-pollinator interactions change through space and time. This could shed light on how these communities

react to climate change or habitat fragmentation. The techniques developed for use on insect collections can easily be transferred to other animal pollinator collections; although pollen retrieval from the pollinator body may need some adjustment. Given some of the limitations in this study, much more research also needs to be done to incorporate historic specimen collections into present-day sampling efforts, and *vice versa*. Doing this will increase the overall number of available samples. This study serves only as an initial exploration of possibilities for utilising pollen metabarcoding of historic collections.

There is currently a large research project which is being conducted by the African Centre for DNA Barcoding (ACDB, www.acdb.co.za) to barcode the plants of Southern Africa (Lahaye *et al*., 2008). The International Barcode of Life (iBoL) barcode standards for plants are followed, which includes the selection of the official barcode genes to target. The recommended barcodes, *rbcL* and *matK* (CBOL Plant Working Group, 2009), are sequenced for plants being officially barcoded in South Africa. As mentioned previously, South Africa has incredible plant diversity, and it would take years to create a comprehensive reference sequence database for all plants. The ITS2 sequence database for South African plants is very small, and additional research into barcoding plants with ITS2 as well will greatly aid in the proliferation of this database, and improve sequence classifications for pollen barcoding purposes.

5.6 Final comments and summary conclusions

To the best of my knowledge, this study was the first to investigate plant-pollinator interactions using DNA barcoding combined with NGS in Africa, and the first of its kind on indigenous bee species in *Megachile* (*M. venusta, M. niveofasciata, M. karooensis, M. murina, M. felina,* and *M. maxillosa*). This study shows that historic collections can be immensely useful in the study of plant-pollinator interactions. Pollen for DNA metabarcoding can be sampled from pre-identified pollinators in collections. The identification of these pollinators and the maintenance of collections by taxonomists is an integral part of this interdisciplinary field. The importance of a sequence reference library that is representative of the flora that is being examined cannot be overstated. Even though the sequence reference databases do not cover nearly all of the floral diversity in South Africa, this is not the case in many other countries with lower floral diversity. The application of metabarcoding pollen sampled from specimens housed in collections in these countries could add value to the study of the plant-pollinator interactions of their endemic pollinators. The challenges of working with older specimens - that sometimes have very limited pollen quantities - should be taken into account when considering a study with collection specimens as samples. Despite

difficulties, pollen metabarcoding of samples obtained from historic pollinator specimens creates opportunities for novel new areas of research.

5.7 References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Bell, K., de Vere, N., Keller, A., Richardson, R., Gous, A., Burgess, K., *et al*., 2016. Pollen DNA barcoding: current applications and future prospects. Genome 59. doi:10.1139/gen-2015-0200.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 106, 12794–12797. doi:10.1073/pnas.0905845106

Harris, D.J., 2003. Can you bank on GenBank? Trends Ecol. Evol. 18, 317–319. doi:10.1016/S0169-5347(03)00128-9.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., *et al*., 2015. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10, e0134735. doi:10.1371/journal.pone.0134735.

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., *et al*., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 17, 558–566. doi:10.1111/plb.12251.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., *et al*., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol. Ecol. Res. 15, 8–16. doi:10.1111/1755-0998.12288.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., *et al*., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. U.S.A. 105, 2923–2928.

McFrederick, Q.S., Rehan, S.M., 2016. Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. Mol. Ecol. 25, 2302–2311. doi:10.1111/mec.13608.

Richardson, R.T., Lin, C.-H., Quijia, J.O., Riusech, N.S., Goodell, K., Johnson, R.M., 2015a. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl. Plant Sci. 3, 1500043. doi:10.3732/apps.1500043.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015b. Application of ITS2 metabarcoding to determine the provenance of pollen collected by

honey bees in an agroecosystem. Appl. Plant Sci. 3, 1400066. doi:10.3732/apps.1400066.

Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., *et al*., 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol. 15. doi:10.1186/s12898-015-0051-y.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA barcoding for honey biodiversity. Diversity 2, 610–617. doi:10.3390/d2040610.

Wilson, E.E., Sidhu, C.S., LeVan, K.E., Holway, D.A., 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol. Ecol. 19, 4823–4829. doi:10.1111/j.1365-294X.2010.04849.x.

APPENDIX: LAYOUT OF R SCRIPTS USED

In this appendix, a layout of the scripts used in R (R Core Team, 2016) is presented. Scripts are presented in the order they were used in Chapter 3 and Chapter 4. The script for producing rarefaction curves was used in both chapters but is only specified once. R code is presented in a smaller, different font and its function is given with a preceding hash (#) in the line above. Where data files should be loaded in R, only "data" is entered into the script syntax as file names may vary. Some packages are required to execute the scripts given in this appendix and should be loaded with "library(*package_name*)":

- *vegan* (Oksanen *et al*., 2016)
- *ggplot2* (Wickham, 2009)
- *reshape2* (Wickham, 2007)
- *plyr* (Wickham, 2011)
- *scales* (Wickham, 2016)

Rarefaction curves

#Input data

```
its1 = read.table(data, sep="¥t", header=T, row.names = 1)
```

#This defines the properties of the graph, such as line colour, line type, line width etc.

```
col <- c('black', 'darkred', 'forestgreen', 'hotpink', 'blue')
lty <- c('solid', 'dashed', 'dotdash')
lwd <- c(1, 2)
pars <- expand.grid(col = col, lty = lty, lwd = lwd, stringsAsFactors = FALSE)
head(pars)
```

#Here the column width of the data is defined.

```
with(pars[1:22, ],
```

#The *rarecurve* function draws the graph with specific parameters defined.

```
rarecurve(its1, step = 1, ylab = '', xlab = '', xlim = c(0,250), ylim = c(1,2.5), col =
col, lty = lty, label = FALSE))
title(main = substitute(paste(italic('ITS2'))))
title(ylab = 'Number of taxa', line= 2)
title(xlab = 'Number of reads', line = 2)
```

```
text(x = 100, y = 87, cex = 2)
```

Spearman rank-order correlation

#Input data

```
data_corr <- read.csv(data, header = TRUE)
```

#Subset the data for the correlation test

```
its1_sub <- subset(data_corr, Data == "its1", select = c("Year","Perc"))
its2_sub <- subset(data_corr, Data == "its2", select = c("Year","Perc"))
mala_sub <- subset(data_corr, Data == "mala", select = c("Year","Perc"))
```

#Execute correlation tests for each subset

```
cor.test(its1_sub$Year, its1_sub$Perc, method  = "spearman", conf.level = 0.95)
cor.test(its2_sub$Year, its2_sub$Perc, method  = "spearman", conf.level = 0.95)
cor.test(mala_sub$Year, mala_sub$Perc, method  = "spearman", conf.level = 0.95)
```

Stacked bar graph plotted with individual samples

#Input data and provide sample names

```
data$sample <- factor(data $ sample, levels = c('A1', 'A2', 'A3', 'A4', 'A5', 'A7', 'A8', 'A9', 'A10', 'A11','A12', 'A13', 'A14', 'A15', 'A16', 'A17', 'A18', 'A19', 'A20', 'A21', 'A22'))
```

#This defines how the bars should be stacked (each bar should be proportioned relative to the amount of reads for that specific sample)

```
bar_graph_2 <- ggplot(aes(x = sample, weight = reads, fill = plant), data = data)
```

#The following commands define the properties of the graph and legend. Each taxon is assigned its own colour and the name of that taxon linked to its colour.

```
bar_graph_2 + geom_bar(position = 'fill') +
theme(legend.text = element_text(face = 'italic')) +
labs(x ='Sample', y = 'Proportion') +
scale_fill_manual(values=c('#a6cee3', '#1f78b4', '#b2df8a', '#33a02c', '#fb9a99', '#e31a1c', '#fdbf6f', '#ff7f00', '#cab2d6', '#6a3d9a', '#ffff99', '#b15928', '#FF0000', '#BF3F00', '#7F7F00', '#3FBF00', 'brown', 'blue', 'black', 'green', 'purple', 'white', 'red', 'yellow', 'pink')

name = 'Plant'
```

```
breaks = c('Helianthus annuus','Pteris vittata','Epacris microphylla','Neolitsea confertifolia', 'Pseudostachyum polymorphum','Pirula salina','Magnolia kwangtungensis','Sisymbrium irio', 'Astragalus membranaceus','Caulerpa webbiana','Macrothamnium leptohymenioides', 'Alternanthera sp. XF30','Pleuropterus multiflorus','Tetrastigma glycosmoides', 'Dianthus masmenaeus','Magnolia grandiflora', 'Pinus fenzeliana','Pteris tremula','Ceratodon purpureus','Acrostichum speciosum', 'Valeriana edulis','Alternanthera env. Sample','Pteris ensiformis','Lepidotrichilia volkensii','Tripteris microcarpa')

labels = c('Helianthus annuus','Pteris vittata','Epacris microphylla','Neolitsea confertifolia', 'Pseudostachyum polymorphum','Pirula salina','Magnolia kwangtungensis','Sisymbrium irio', 'Astragalus membranaceus','Caulerpa webbiana','Macrothamnium leptohymenioides', 'Alternanthera sp. XF30','Pleuropterus multiflorus','Tetrastigma glycosmoides', 'Dianthus masmenaeus','Magnolia grandiflora', 'Pinus fenzeliana','Pteris tremula','Ceratodon purpureus','Acrostichum speciosum', 'Valeriana edulis','Alternanthera env. Sample','Pteris ensiformis','Lepidotrichilia volkensii','Tripteris microcarpa'))
```

Stacked bar graph with* Megachile venusta *sample data combined

#Input data

```
bar_graph_1 <- read.delim(data,row.names = 1,check.names = FALSE)
```

#This defines how the bar should be stacked (the bar should be proportioned relative to the amount of total detection counts per plant taxon)

```
bar_graph_1 <- ggplot(aes(x = sample, weight = reads, fill = plant), data = data)
```

#The following commands define the properties of the graph and legend. Each taxon is assigned its own colour and the name of that taxon linked to its colour.

```
bar_graph_1 + geom_bar(position = 'fill', width = .3) +
theme(legend.text = element_text(face = 'italic')) +
labs(x = expression(italic(Megachile~venusta)~floral~representation), y = 'Proportion') +
scale_fill_manual(values=c('#a6cee3', '#1f78b4', '#b2df8a', '#33a02c', '#fb9a99', '#e31a1c', '#fdbf6f', '#ff7f00', '#cab2d6', '#6a3d9a', '#ffff99', '#b15928', '#FF0000', '#BF3F00', '#7F7F00', '#3FBF00', 'brown', 'blue', 'black', 'green', 'purple', 'white', 'red', 'yellow', 'pink')

name = 'Plant'
```

```
breaks = c('Helianthus annuus','Pteris vittata','Epacris microphylla','Neolitsea confertifolia', 'Pseudostachyum polymorphum','Pirula salina','Magnolia kwangtungensis','Sisymbrium irio', 'Astragalus membranaceus','Caulerpa webbiana','Macrothamnium leptohymenioides', 'Alternanthera sp. XF30','Pleuropterus multiflorus','Tetrastigma glycosmoides', 'Dianthus masmenaeus','Magnolia grandiflora', 'Pinus fenzeliana','Pteris tremula','Ceratodon purpureus','Acrostichum speciosum', 'Valeriana edulis','Alternanthera env. Sample','Pteris ensiformis','Lepidotrichilia volkensii','Tripteris microcarpa')

labels = c('Helianthus annuus','Pteris vittata','Epacris microphylla','Neolitsea confertifolia', 'Pseudostachyum polymorphum','Pirula salina','Magnolia kwangtungensis','Sisymbrium irio', 'Astragalus membranaceus','Caulerpa webbiana','Macrothamnium leptohymenioides', 'Alternanthera sp. XF30','Pleuropterus multiflorus','Tetrastigma glycosmoides', 'Dianthus masmenaeus','Magnolia grandiflora', 'Pinus fenzeliana','Pteris tremula','Ceratodon purpureus','Acrostichum speciosum', 'Valeriana edulis','Alternanthera env. Sample','Pteris ensiformis','Lepidotrichilia volkensii','Tripteris microcarpa'))
```

Stacked bar graphs to represent detection count data for three different geographic regions

#Input data

```
data <- read.delim(data, header = TRUE)
```

#This defines how the bar should be stacked (bars should be proportioned relative to the amount of total detection counts per plant taxon)

```
single_bar <- ggplot(data=data, aes(y = num_per, fill=plant, x = bar))
```

#The following commands define the properties of the graph and legend.

```
single_bar <- single_bar + geom_bar(stat ='identity', width = 0.2)
single_bar <- single_bar + scale_fill_manual(values=c('#a6cee3', 'yellow', 'red', 'ivory3', 'seagreen', 'darkgray', 'cadetblue', 'darkgoldenrod3', 'gray26', 'darkorange', 'coral2', '#ffff99', 'deeppink', 'white', 'chocolate', '#7F7F00', 'deepskyblue', '#1F78B4','blue', 'black', 'green', 'purple', 'burlywood1', '#CAB2D6', 'aquamarine', 'pink', 'bisque4', 'lightcyan', '#33A02C', 'brown', 'darkolivegreen3', 'darkslateblue'))
single_bar <- single_bar + labs(title = '')
single_bar <- single_bar + guides(fill=guide_legend(title='Plant family'))
single_bar <- single_bar + xlab('')
```

```
single_bar <- single_bar + ylab('Percentage of plant family represented in biome') +
theme_bw()
single_bar <- single_bar + coord_fixed(ratio = 0.02)
```

References

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*., 2016. vegan: Community ecology package. R package version 2.3-4.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Wickham, H., 2016. scales: Scale functions for visualization. R package version 0.4.0.

Wickham, H., 2011. The split-apply-combine strategy for data analysis. J. Stat. Softw. 40, 1–29.

Wickham, H., 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.

Wickham, H., 2007. Reshaping data with the reshape package. J. Stat. Softw. 21, 1–20.

APPENDIX B: INVITED REVIEW AS CO-AUTHOR WITH PRESENTERS FROM "POLLEN BARCODING" SESSION AT 6TH INTERNATIONAL BARCODE OF LIFE CONFERENCE

The review article I co-authored was published in the journal *Genome* on 13 April 2016. It formed part of a special issue entitled "Barcodes to Biomes". The published paper is available with free access at dx.doi.org/10.1139/gen-2015-0200.