The gut microbiomes of desert *Pachysoma* spp.

**MacLeay (Coleoptera: Scarabaeidae)**

by

Philippa Zena Nel Franzini

Submitted in partial fulfilment of the requirements for the degree

*Philosophiae Doctor* (Ph.D.) Genetics

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

November 2016

Supervisor: Prof. Don A. Cowan

Co-supervisors: Dr. Jean-Baptise Ramond; Dr. Sandra Ronca
Declaration

I, Philippa Zena Nel Franzini, declare that the thesis/dissertation, which I hereby submit for the degree *Philosophiae Doctor* Genetics (Ph.D.) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE:
Dedication

This thesis is dedicated to my late father. You were my inspiration and my greatest supporter. You are greatly missed.
Acknowledgments

Foremost, I would like to thank my family for their continued love and support throughout my studies. To my parents who supported me continuously in all my endeavours and who taught me that together we could accomplish anything. To my brother who inspired my love in science and nature as a child. To my husband who has been by my side throughout my Ph.D. and who I couldn’t have done this without.

I thank God for all of his blessings and for guiding me through my Ph.D.

I would like to thank the NRF for financial assistance during the Ph.D. study.

I thank the University of Pretoria, Department of Genetics, Centre for Microbial Ecology and Genomics (CMEG) and the Genomics Research Institute (GRI) for allowing me to pursue this study and for supplying financial support and infrastructure.

I owe a huge thank you to my supervisors Prof. Don Cowan, Dr. Jean-Baptise Ramond and Dr. Sandra Ronca for continuously guiding and supporting me. Furthermore, I would like to thank our collaborators Prof Clarke Scholtz and Dr Catherine Sole as well as the members of the Scarab Research Group for their assistance with collections of samples, dissections and support.

I would thank all of my fellow lab-mates at CMEG for your advice, support and friendship throughout this study. In particular I would like to thank Dr, O. Bezuidt, Mr M. van Goethem, Dr A. Valverde and Mr. R. Johnson for your continued assistance.

I especially want to thank Sandra, Storme and Olivier for their encouragement during these past four years.

Surendra Vikram your assistance with bioinformatics analysis was invaluable and you have taught me a great deal during this study.
Abstract

Microbial communities inhabit many environmental niches including the nutrient-rich gut systems of animals, where they are involved in a number of important processes. Insect gut microbiota may assist the host with several functions including synthesis of nutritional components lacking from the host diet and digestion of lignocellulosic materials. It is generally believed that the diet of the host plays an important role in the structure of the gut microbiome. Numerous studies have focused on insects feeding on lignocellulosic diets such as termites, as well as medically and agriculturally important insect species. Few studies have researched the gut microbiota of adult dung beetles. Most scarab beetle species feed on the liquid component of wet dung, whereas *Pachysoma* spp. may feed on lignocellulosic materials within their diet of dry dung, plant detritus or both. This feeding behaviour makes *Pachysoma* an ideal candidate for studying the role that diet has on gut microbiome assembly.

Plant detritus feeding *P. endroedyi* and the dry dung feeding *P. striatum* were collected from Namaqualand, South Africa. The mid- and hindgut of each individual were dissected and mDNA extracted using a phenol-chloroform method. Amplicon sequencing of the bacterial 16S rRNA gene and the fungal ITS region was used to determine inter- and intra-specific differences in microbial community structures. Shotgun sequencing of the entire gut metagenome was carried out on mDNA extracted from whole gut samples. Shotgun sequencing was used for both taxonomic and functional annotation of the *Pachysoma* gut microbiomes.

Both amplicon and shotgun sequencing detected substantial differences in bacterial and fungal diversity between the two *Pachysoma* species. Amplicon sequencing showed the number of bacterial phyla ranged from 6-11 and 4-7 (total 14 phyla) for *P. endroedyi* and *P. striatum*, respectively. Furthermore, a minimal core microbiome was detected with only 2.57% of the bacterial OTUs shared between the two *Pachysoma* species studied. Large intra-specific variations were also noted within both *Pachysoma* species. Fungal communities could not be detected in the gut of *P. endroedyi*, while only two fungal phyla were detected *P.
*Pachysoma striatum* gut samples. Metagenome shotgun sequencing detected a greater bacterial diversity (total of 39 phyla) than the 16S rRNA gene amplicon study, although large differences were noted between the two species. Furthermore, shotgun sequencing demonstrated that fungal communities were present in the guts of both *Pachysoma* species. Archaea, viruses and other eukaryotic microorganisms were also present in the gut metagenomes of both *Pachysoma* species.

The functional capacity of the *Pachysoma* spp. gut microbiomes was analysed using shotgun sequencing. Both species had the genetic capacity to degrade cellulose and hemicellulose but not lignin, supporting the suggestion that *P. striatum* feeds on plant material in the dry dung. Furthermore, the functional capacity of the microbiomes of both *Pachysoma* species were comparable, suggesting the ability for both species to feed on either dry dung or plant detritus. The similarity of the functional profiles of the two *Pachysoma* species suggests the existence of a functional rather than phylogenetic core microbiome.

This primary study has successfully characterised the phylogenetic and functional profiles of the gut microbiomes of two *Pachysoma* species feeding on different substrates. However, it is still unclear if diet is the primary driver in gut microbiome assembly.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOSIM</td>
<td>Analysis of Similarities</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>BP</td>
<td>Base pair</td>
</tr>
<tr>
<td>CAZy</td>
<td>Carbohydrate-Active enZYme</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge coupled device</td>
</tr>
<tr>
<td>COG</td>
<td>Clusters of Orthologous Groups</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyl trimethylammonium bromide</td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>DictDB</td>
<td>Dictyoptera gut microbiota reference database</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribose nucleoside triphosphates</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>GH</td>
<td>Glycoside hydrolase</td>
</tr>
<tr>
<td>KAAS</td>
<td>KEGG Automatic Annotation System</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>mDNA</td>
<td>Metagenomic DNA</td>
</tr>
<tr>
<td>MEGAN</td>
<td>MEtaGenome ANalyzer</td>
</tr>
<tr>
<td>MG-RAST</td>
<td>Metagenomics RAST server</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation Sequencing</td>
</tr>
<tr>
<td>nMDS</td>
<td>Non-Metric Multi-Dimensional Scaling</td>
</tr>
<tr>
<td>NR</td>
<td>Non-redundant</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>OTU</td>
<td>Operational taxonomic unit</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
</tbody>
</table>

© University of Pretoria
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pH</td>
<td>Potential of hydrogen</td>
</tr>
<tr>
<td>PM</td>
<td>Peritrophic membrane</td>
</tr>
<tr>
<td>RDP</td>
<td>Ribosomal Database Project</td>
</tr>
<tr>
<td>TRFLP</td>
<td>Terminal Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
Table of Figures

- Figure 1.1: The global distribution of deserts measured against the terrestrial surface temperature (taken from Makhalanyane et al. (2015)). .......................................................... 2
- Figure 1.2: Typical vegetation within Namaqualand (Photo courtesy of H. de Klerk). .... 4
- Figure 1.3: Basic structure of the larval scarab gut (adapted from Huang et al. (2010)). ........................................................................................................................................... 14
- Figure 2.1: *Pachysoma* samples collected for the present study: a) *P. endroedyi* and b) *P. striatum* (Courtesy of Hennie de Klerk). ................................................................................................................ 48
- Figure 2.2: Collection of *Pachysoma* samples: a) Observation of a burrow; b) a typical *Pachysoma* spp. burrow with the food-source (plant detritus) and a beetle inside (Courtesy of Hennie de Klerk). ................................................................................................................ 48
- Figure 2.3: Aerial view of the Western coast of South Africa with the Namaqualand Desert collection site indicated by the yellow marker and shown in the insert (Insert map data: Google, AfriGIS (Pty) Ltd, 2016). ........................................................................................................................................... 49
- Figure 2.4: Burrow being examined surrounded by typical vegetation at the collection site (Photo courtesy of H. de Klerk)................................................................................................................ 49
- Figure 2.5: Dissected *P. striatum* showing H: hindgut, M: Midgut and P (square): pyloris (Foregut not shown). ........................................................................................................................................... 51
- Figure 3.1: Cytochrome oxidase I gene Parsimony tree phylogeny of 13 *Pachysoma* spp. Branch colours indicate the diet of the *Pachysoma* spp.: dung (brown), plant detritus (green), polyphagous (blue) and unknown (black). This phylogentic tree was adapted, with permission, from (Sole et al., 2005) and the dietary information taken from both (Harrison et al., 2003) and personal observations by Prof C. Scholtz. Two species of *Scarabaeus*, (*S. proboscideus* and *S. rugosus*) which is the sister-genus to *Pachysoma* and a typical wet-dung-feeder, were used as outgroups. Numbers to the right of the tree indicate the three *Pachysoma* lineages. The two species considered in this study, *P. endroedyi* and *P. striatum*, are indicated with stars (Adapted from Sole et al. (2005)). ................................................................................................................ 64
Figure 3.2: Rarefactions curves showing gut microbial community richness of all Pachysoma individuals for bacterial 16S rRNA gene amplicon data of: a) P. endroedyi, b) P. striatum; and c) fungal ITS region amplicon data of P. striatum. ................................. 67

Figure 3.3: nMDS ordination plot based on Bray-Curtis distance matrices of bacterial 16S rRNA gene pyrosequencing data for P. endroedyi and P. striatum individuals. A stress value of less than 0.1 represents a high quality ordination. Pachysoma endroedyi and P. striatum are represented by green and inverted brown triangles, respectively. ......................... 69

Figure 3.4: Venn diagrams showing distribution of bacterial OTUs between (a) P. endroedyi and (b) P. striatum individuals based on the 16S rRNA gene pyrosequencing analysis. Shared OTUs are shown in bold. Numerical labels are shown for each individual. 71

Figure 3.5: a) Venn diagram comparing the distribution of fungal OTUs between P. striatum individuals based on the ITS region pyrosequencing analysis. Shared OTUs are shown in bold. Numerical labels are given for each individual; b) Relative abundance of fungal phyla in five P. striatum individuals based on ITS rRNA gene region pyrosequencing analysis at a 97% identity threshold. .................................................................................................. 72

Figure 3.6: Comparison of inter-specific differences in relative abundance of bacterial phyla in the gut of two Pachysoma spp., P. endroedyi and P. striatum, based on 16S rRNA gene pyrosequencing analysis. ................................................................................................................. 75

Figure 4.1: Amplification of the a) 18S rRNA and b) 16S rRNA genes from cockroach gut samples subjected to all four host DNA reduction methods. L: molecular weight KAPA Universal Ladder, C: no treatment, 1: removal of PM, 2: filtration, 3: minimal homogenisation and 4: low-speed centrifugation. Escherichia coli cells were used as the positive control, therefore, no amplification was expected in gel (a). ................................................................. 93

Figure 4.2: Rarefaction curves of gut microbial community richness of a) P. endroedyi and b) P. striatum gut samples. Eukaryotic contigs derived from Metazoa and Streptophyta were removed before analysis. ............................................................................................................. 96
Figure 4.3: Taxonomic affiliation of shotgun sequencing reads assigned using MG-RAST (a and b), and MEGAN (c and d) for *P. endroedyi* (a and c) and *P. striatum* (b and d) gut metagenomes. .............................................................................................................................................. 98

Figure 4.4: Diversity and abundance of bacterial phyla in the gut samples of *P. endroedyi* (green) and *P. striatum* (brown) using shotgun (Illumina Hiseq) analysed using MEGAN.. 100

Figure 4.5: Functional assignments of microbial-derived ORFs from both *Pachysoma* species to categories of COGs generated by MEGAN. ................................................................. 102

Figure 4.7: Summary of the plant cell wall degrading enzymes present in the *P. endroedyi* and *P. striatum* gut metagenomes. Only hemicellulases present in at least one of the metagenomes are listed. Enzymes are highlighted according to the metagenome in which they were detected: *P. striatum* (brown) or both *Pachysoma* metagenomes (yellow). ....... 111

Figure 4.8: Adapted KEGG nitrogen metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes. ........................................................................................................ 113

Figure 4.9: Adapted KEGG peptidoglycan biosynthesis reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes. ........................................................................................................ 115

Figure 4.10: Abundance and diversity of genes encoding energy metabolism pathways (KEGG) for *P. endroedyi* (green) and *P. striatum* (brown) gut microbial metagenomes..... 116

Figure 4.11: Abundance and diversity of genes encoding carbohydrate metabolism pathways (KEGG) for *P. endroedyi* (green) and *P. striatum* (brown) gut microbial metagenomes. .................................................................................................................. 117

Figure 4.12: Adapted KEGG pyridoxine (vitamin B6) metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes. .................................................................................. 118

© University of Pretoria
Figure 4.13: Adapted KEGG Ascorbate (Vitamin C) and Aldarate metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes. .............................................. 119

Figure 4.6: Distribution of glycosyl hydrolase (GH) families in the gut metagenomes of *P. endroedyi* and *P. striatum* annotated using hmmer. ................................................................. 104
List of Tables

Table 2.1: List of chemical reagents .................................................................................. 46
Table 2.2: Enzymes used in this study ............................................................................... 46
Table 2.3: Compositions of buffers and solutions used in this study ......................... 47
Table 3.1: Values for sequence reads, OTUs, phyla and diversity indices for bacterial and fungal gut communities of *P. endroedyi* and *P. striatum* individuals .................... 66
Table 3.2: Phylogenetic classification of the most abundant bacterial genera in the gut samples of *P. endroedyi* and *P. striatum*: i.e., representing >2% reads ...................... 77
Table 4.1: Overview of methods tested for efficiency in reducing host DNA contamination during mDNA extraction ................................................................. 91
Table 4.2: Nanodrop readings of mDNA yields and purity ratios for samples treated with four host DNA reduction methods ................................................................. 92
Table 4.3: Nanodrop readings of mDNA yields and purity values for replicate *P. endroedyi* and *P. striatum* gut samples sent for shotgun sequencing ................................. 94
Table 4.4: Sequencing metadata and assembly metrics for shotgun sequencing of the *P. endroedyi* and *P. striatum* gut microbiomes using CLC Genomics Workbench, MG-RAST and MEGAN .......................................................... 95
Table 4.5: Glycosyl hydrolase families detected in the gut metagenomes of *P. endroedyi* and *P. striatum* and the associated KEGG E.C. assignments were applicable .......................... 105
# Table of Contents

Declaration ............................................................................................................................ II  
Dedication ............................................................................................................................ III  
Acknowledgments ................................................................................................................ IV  
Abstract ................................................................................................................................ V  
List of Abbreviations ............................................................................................................ VII  
Table of Figures ................................................................................................................... IX  
List of Tables ...................................................................................................................... XIII  
Chapter 1: Literature Review .............................................................................................. 1  
  1.1 Desert Environments ................................................................................................. 1  
  1.1.1 The Desert Biome ............................................................................................... 1  
  1.1.2 Namaqualand: Study Site .................................................................................. 2  
  1.2 Desert Insect Communities: A Brief Overview ....................................................... 4  
  1.2.1 Desert Dung Beetles ......................................................................................... 5  
  1.2.2 The Genus *Pachysoma* .................................................................................. 5  
  1.3 Diets of Insects ......................................................................................................... 6  
  1.3.1 The Nutritional Quality of the *Pachysoma* Diet .............................................. 7  
  1.3.2 An Overview of Plant Cell Wall Degradation .................................................… 11  
  1.4 Structure of the Insect Gut ....................................................................................... 12  
  1.5 Insect Gut Microbial Communities ......................................................................... 15  
  1.5.1 Microbial Colonization of the Gut Environment .............................................. 15  
  1.5.2 Variations in Gut Microbial Communities ....................................................... 18  
  1.5.3 Gut Microbial Symbionts and the Relationship between Host Diet and the Gut Microbiome ........................................................................................................ 19  
  1.5.4 Previous Research on Scarab Beetle Gut Microbiomes .................................... 21  
  1.6 Sequence-Based Methods in Microbial Ecology ..................................................... 22
3.2.2 Sequencing Outputs and Diversity Indices of the Bacterial 16S rRNA Gene and Fungal ITS Region of the Pachysoma Gut Microbiome ............................ 65

3.2.3 Inter-specific Variations in Bacterial and Fungal Pachysoma Gut Communities ......................................................................................... 68

3.2.4 Intra-specific Variation of Pachysoma Gut Microbial Communities ................. 70

3.2.5 Phylogenetic Diversity of Bacterial and Fungal Pachysoma Gut Communities ......................................................................................... 74

3.3 Conclusion ........................................................................................................... 78

3.4 References ........................................................................................................... 79

Chapter 4: Shotgun Metagenomic Sequencing of the Gut Microbiomes of Pachysoma endroedyi and Pachysoma striatum ................................................................. 88

4.1 Introduction ........................................................................................................... 88

4.2 Results and Discussion ........................................................................................ 89

4.2.1 Comparison of Four Previously Published Host DNA Reduction Methods . ......................................................................................... 89

4.2.2 Shotgun Sequencing of the Gut Microbiomes of P. endroedyi and P. striatum ......................................................................................... 94

4.2.3 Community Composition of the Pachysoma Metagenome ......................... 96

4.2.4 Functional Profiling of the Metagenomes of P. endroedyi and P. striatum ......................................................................................... 101

4.3 Conclusion ........................................................................................................... 120

4.4 References ........................................................................................................... 120

Chapter 5: Concluding Remarks ....................................................................................... 129

5.1 Studying the Pachysoma gut microbiome ....................................................... 129

5.2 Limitations of this study ..................................................................................... 131

5.3 Future study prospects ....................................................................................... 133

5.4 References ........................................................................................................... 135
Chapter 1: Literature Review

1.1 Desert Environments

1.1.1 The Desert Biome

Deserts make up approximately a third of the global terrestrial biome (Laity, 2009). Deserts are located globally (Figure 1.1), with 57.2% of the African continent considered arid (Peel et al., 2007), including the world’s largest desert (the Sahara Desert (Harris, 2003)), while the only continent absent of any known deserts is Europe (Harris, 2003).

A desert can be defined as an environment where the ratio of precipitation to potential evaporation (P/PET) is less than 1 (Unep, 1992). Using this definition, deserts can further be defined according to four categories of aridity, namely dry sub-arid (0.5≤P/PET<0.65), semi-arid (0.2≤P/PET<0.5), arid (0.05≤P/PET<0.2), and hyperarid (P/PET<0.05). Although most deserts are known as being extremely hot, polar deserts, such as those in Antarctica, are known for their exceptionally cold climates (Cowan and Tow, 2004). For this reason deserts are further classed as hot (>18°C) and cold deserts (<18°C) (Peel et al., 2007).
Figure 1.1: The global distribution of deserts measured against the terrestrial surface temperature (taken from Makhalanyane et al. (2015)).

1.1.2 Namaqualand: Study Site

The name Namaqualand is derived from the Khoe-speaking Nama people who originally inhabited the land along with the San people (Wisborg and Rohde, 2005, Benjaminsen et al., 2006). Namaqualand is a semi-arid coastal desert forming part of the Succulent Karoo biome in the Northern Cape Province of South Africa (Cowling et al., 1999, Benjaminsen et al., 2006, Desmet, 2007). Various studies report the approximate size of Namaqualand to vary from 45000-52600km² (Cowling et al., 1999, Wisborg and Rohde, 2005, Benjaminsen et al., 2006, Desmet, 2007, May and Lahiff, 2007). The Gariep River separates Namaqualand from Namibia, while the Olifants River and Bokkeveld escarpment form the southern boundary with vast Bushman planes surrounding the eastern side of the desert (Desmet, 2007). The Atlantic Ocean forms the western border of the desert (Benjaminsen et al., 2006, Desmet, 2007).

The climate of Namaqualand varies according to region. Winter rainfall occurs towards the western coastal areas with summer rainfall in the interior regions (Desmet, 2007).
annual rainfall in the region is relatively low at approximately 50-400 mm (Cowling et al., 1999, Desmet, 2007). Namaqualand is relatively cold due to the south-westerly sea breeze, keeping maximum average summer temperatures below 30°C (Desmet, 2007). In contrast, winter temperatures typically exceed 35°C (Cowling et al., 1999, Desmet, 2007). Due to the coastal location of the Namaqualand, supplemented by high humidity with low temperatures at night, desert fog events and heavy dew are common occurrences, particularly in winter months (Cowling et al., 1999, Desmet and Cowling, 1999, Desmet, 2007).

Several bioregions (biophysical regions which are classified according to physical environment, climate and flora), are found within this arid ecosystem (Desmet, 2007). Namaqualand covers approximately a quarter of the Succulent Karoo (Desmet, 2007), which is listed as one of the global 200 priority ecoregions (Olson and Dinerstein, 2002) as well as a biodiversity hotspot (Myers et al., 2000). Namaqualand boasts more than 3500 plant species in 135 families (Desmet, 2007). Approximately 25% of the Namaqualand flora is endemic to this region, although this value is estimated to actually be between 40 and 50% (Cowling et al., 1999, Desmet, 2007). The Namaqualand uplands and central Namaqualand coast harbour 1109 and 432 plant species, respectively, 71 and 74 of which are on the Red Data List of endangered species (SKEP, 2008). The most diverse plant families within Namaqualand include Aizoaceae (ice plants), Asteraceae (Aster family) and Iridaceae (Iris family) all consisting of over 200 species (Desmet, 2007). Figure 1.2 shows an example of the typical vegetation which is found in parts of Namaqualand.

Anthropogenic activities impact desert ecosystems, with the most prominent being commercial livestock farming (Benjaminsen et al., 2006, O’Farrell et al., 2011) and mining activities, both of which are prominent in this area (Benjaminsen et al., 2006, May and Lahiff, 2007). The Namaqualand area is considered to be a global conservation priority although only a small portion of the region and the associated vegetation is currently protected (Desmet, 2007, SKEP, 2008).
1.2 Desert Insect Communities: A Brief Overview

The fauna and flora able to survive in extreme desert environments are typically found in low abundance, although diversity may be high (Harris, 2003). Organisms that manage to survive in desert environments benefit in several ways including exploitation of available resources (e.g. food-sources) and reduced competition (Wharton, 2002). Insects are amongst the most abundant members of the kingdom Animalia found in desert ecosystems due to their ability to adapt to and survive under extreme conditions (Wharton, 2002, Harris, 2003).

The main challenges affecting survival in desert ecosystems are the extreme heat and lack of water (Harris, 2003, Allaby, 2006). Desert fauna are largely dependent on the available plants as a valuable water and nutritional source (Harris, 2003, Allaby, 2006). Therefore, the reduced ability of plants to adapt to desert environments in turn affectively reduces the number of animals that can successfully inhabit these areas (Allaby, 2006).

1.2.1 Desert Dung Beetles

Coleoptera (commonly known as beetles) is the largest of 30 orders within the class Insecta (Roskov et al., 2016). The subfamily Scarabaeinae (scarab beetles) encompasses more than 225 genera consisting of 5700 valid species of dung beetles (Scholtz et al., 2009), with 670 known species in Southern Africa (Mlambo et al., 2015). Scarabaeinae consists of 11 tribes, namely: Ateuchini, Deltochilini, Eucraniini, Gymnopleurini, Oniticellini, Onitini, Coprini, Onthophagini, Phanaeini, Scarabaeini, and Sisyphini (Bouchard et al., 2011). Contradictory to what the name may suggest, species of dung beetles may also feed on carrion, fruit, fungi and plant detritus (Scholtz et al., 2009, Mlambo et al., 2015). This study will focus on the genus *Pachysoma* MacLeay (1821) which forms part of the tribe Scarabaeini (Bouchard et al., 2011).

1.2.2 The Genus *Pachysoma*

The taxonomic level of *Pachysoma* has been consistently disputed. *Pachysoma* was originally a generic classification (Ferreira, 1953) before being synonymized with its closest relative *Scarabaeus* Linnaeus, 1758 (Mostert and Holm, 1982). Later studies determined
*Pachysoma* to be a subgenus of *Scarabaeus* (Harrison and Philips, 2003, Forgie et al., 2005). However, *Pachysoma* currently retains full generic status through collective morphometric and genetic evidence (Forgie et al., 2006, Sole et al., 2007, Mlambo et al., 2015).

The genus *Pachysoma* (Scarabaeoidea: Scarabaeidae) consists of 13 species endemic to the south-western African coastal desert (Cape Town, South Africa to Walvis Bay, Namibia) (Harrison et al., 2003). While Scarabaeus are distributed around the world in a variety of habitat types, distributions of individual desert *Pachysoma* species are usually restricted (Harrison et al., 2003).

*Pachysoma* species differ most significantly from other dung beetles in their diet and foraging behaviour. While most insects of the same genus feed on similar diets, *Pachysoma* species vary in the dietary materials on which they feed: plant detritus, dung or both (Harrison et al., 2003). Furthermore, *Scarabaeus* species typically feed on wet dung from herbivorous mammals. They form balls from the dung and roll them backwards to a suitable area for burial below the surface (Doube, 1990). *Pachysoma* deviate from this behaviour by feeding on dry dung and plant detritus. *Pachysoma* dig a burrow once suitable food has been located, repeatedly foraging and dragging the food forwards to the burrow (Scholtz, 1989). It is hypothesized that *Pachysoma* have probably adapted to feeding on dry dung due to the unavailability of the wet counterpart and the arid environments in which these species are found (Scholtz et al., 2004). In order to exploit this resource, a change in foraging behaviour was necessary as dry dung and detritus cannot be rolled into a ball, thereby modifying the behaviour of *Pachysoma* to dragging small fragments to their burrows (Scholtz et al., 2004).

### 1.3 Diets of Insects

Insects feed on a wide array of substrates including plant-based diets (e.g. wood, detritus, leaves and roots), nectar, pollen, sap, fungi, fur, feathers, skin, soil, blood, carion and other insects (Brauman, 2000, Douglas, 2009). The chemical composition of each food-source differs, with some dietary components easier to digest than others (Karasov et al., 2011).
Sugars, proteins and lipids are rapidly digested, whereas chitin, lignocelluloses and insoluble starches are harder to digest (Karasov et al., 2011). The composition of refractory and non-refractory chemicals and materials in a variety of dietary substrates is reviewed by Karasov et al. (2011).

Animals feeding on diets high in recalcitrant materials may either assimilate the less recalcitrant compounds in the diet, avoiding indigestible materials or may have adapted to feed on and utilise these recalcitrant materials (Abe and Higashi, 1991). *Pachysoma* species are believed to have adapted to feed on and utilize recalcitrant materials in their diets (Holter and Scholtz, 2011, 2013).

1.3.1 The Nutritional Quality of the *Pachysoma* Diet

All species of *Pachysoma* appear to feed on and utilise refractive materials within their respective diets (dung or plant detritus). Four *Pachysoma* species are known to primarily feed on plant detritus (*P. hippacrates*, *P. glentoni*, *P. schinzi* and *P. endroedyi* (C. Scholtz Pers Comm; Harrison et al. (2003))). Plant detritus (otherwise known as plant litter), refers to decomposing plant matter. The chemical and material composition of plant detritus differs according to plant species (Hättenschwiler et al., 2008). For this reason, a brief overview of the general composition of plant detritus will be given.

Plant detritus consists of primary (plant residues) and secondary (microbial residues and exudates) resources (Kögel-Knabner, 2002). Plant detritus can be composed of two types of plant tissues: namely parenchymatic and woody tissues (Kögel-Knabner, 2002). Polysaccharides (and lignin when relevant) are the most abundant organic compounds of plant detritus (Kögel-Knabner, 2002), forming approximately 90% of the plant cell wall (Doblin et al., 2010). Plant cell wall polysaccharides include cellulose, xylan, mannose, glucomannans, galactan and pectin (Kögel-Knabner, 2002, Hättenschwiler et al., 2008). Proteins account for approximately 10% of the plant cell wall (Doblin et al., 2010). Other components of plant cell walls include polyphenols, lipids and cutin (Kögel-Knabner, 2002, Hättenschwiler et al., 2008). Storage and non-refractory compounds of plant detritus, such as proteins, starches, pigments,
fructan and α-glycans, are easily digested and form an important resource for microorganisms (Kögel-Knabner, 2002, Karasov et al., 2011). Lignified plant material is considered to be nutrient deficient with low concentrations of nitrogen, amino acids, sterols and vitamin B (Dillon and Dillon, 2004, Brune, 2013). However, plant materials with low concentrations of lignin are a protein and nitrogen rich source (Brune, 2013).

The plant litter chemical composition varies, with high carbon concentrations (45.3 and 52.4% (Hättenschwiler et al., 2008)), while nitrogen concentrations (0.68-2.5%) vary according to plant species (Kemp et al., 2003, Hättenschwiler et al., 2008). Along with carbon and nitrogen, phosphorous, iron, aluminium, calcium, potassium, magnesium and manganese concentrations in plant detritus tend to vary over time (Kemp et al., 2003, Goya et al., 2008, Hättenschwiler et al., 2008). However, starch, nitrogen and phosphorous concentrations have been found to be significantly lower in the plant detritus when compared to the living plant (Hättenschwiler et al., 2008). Furthermore, seasonal changes affect the chemical composition of vegetation. Specifically, proteins and carbohydrates are moved to the root systems during the winter seasons, changing the nutritional content of the leafy structures typically fed on by animals (Sinclair, 1975).

Before consumption by insects, detritus undergoes microbial conditioning (Swift et al., 1979), where microorganisms initiate the degradation process and release a portion of the nutrient content. Microbial decomposition over time affects the composition of plant litter (Benner and Hodson, 1985). However, microbial conditioning offers several advantages including the detoxification of allelochemicals, breakdown of the cell wall and immobilization of nitrogen and phosphorous (Nalepa et al., 2001).

Plant detritus and faecal matter are considered to be comparable (Webb, 1977), with dung beetles having been able to routinely move between the two dietary substrates (Cambefort, 1991). An example of this may be P. benningesi, which is considered polyphagous, feeding on both dry dung pellets and plant detritus (Harrison et al., 2003) depending on availability of the respective substrate.
In comparison to the known plant detritus feeding *Pachysoma* species, six species are found to feed primarily on dry dung pellets (*P. aesculapius*, *P. garapinum*, *P. denticola*, *P. rotundigenus*, *P. rodriguesi* and *P. striatum*). Dung is excreted organic waste consisting of nutritional materials and chemicals not fully utilised by the animal (Mlambo et al., 2015). Dung can be divided into plant fragments (fibre), which have generally been subjected to mastication by the original animal, and a brown liquid/viscous component (Anderson and Coe, 1974).

In general, dung is composed of digestive juices, ash, albuminous substances, fats, carbohydrates, free living and host associated microorganisms, epithelial cells from the animal, minerals and vitamins (Anderson and Coe, 1974, Lodha, 1974, Müller, 1980, Nalepa et al., 2001, McDowell and Stewart, 2005, Mlambo et al., 2015, Holter, 2016). Furthermore, the chemical and material content, texture and volume of dung varies according to animal species from which the dung is expelled, respective feeding strategies/diets, rainfall and the season, feeding behaviours (i.e., grazing or browsing) as well as whether or not the animal is ruminant (Greenham, 1972, Arman and Hopcroft, 1975, Edwards, 1991, Leeming et al., 1996, Delve et al., 2001, Janecke and Smit, 2015). When larger undigested materials are excreted, the gut mucosa can be damaged resulting in the increase of epithelial cells, mucus and the presence of blood in the stool (Greenham, 1972).

Fibre in dung has been subjected to at least one round of previous digestion, altering the chemical and physical structure of components such as carbohydrates (Lodha, 1974). The concentration of refractive components of plant materials (fibre) in the dung varies across animal species and the ruminant nature of the animal, with cellulose ranging from 15-41%, hemicellulose varying from 7-32% and lignin from 10-44% of the respective faecal organic matter (Holter, 2016).

Nitrogen is a very important chemical component of dung, with greater concentrations in the dung than in the original substrates/vegetation, presumably due to microbial activity in the original animal's intestine (Anderson and Coe, 1974, Lodha, 1974). Concentrations of faecal nitrogen fluctuate according to animal species with nitrogen content higher in dung from
ruminants than non-ruminants (Holter, 2016). Faecal nitrogen is both available from undigested material in the dung or from microbial biomass in the dung and water-soluble nitrogenous compounds (Holter, 2016).

Nine of the ten essential amino acids for insects (Barbehenn et al., 1999, Nation, 2008), as well as some non-essential amino acids are found in dung, albeit sometimes in low concentrations (Müller, 1980, Mason et al., 1989, Rougon et al., 1990). Experimental methods may discriminate against the tenth essential amino acid (Tryptophan) due to strong acids used for protein hydrolysis degrading tryptophan (Edelhoch, 1967). A variety of sterols, including the animal sterol cholesterol, are available from dung of herbivorous and omnivorous animals (Leeming et al., 1996, Tyagi et al., 2008, Derrien et al., 2011), suggesting that the coprophagous insects acquire their essential sterols from their diet.

Dung also contains ash, a mineral residue of soil particles and cell wall silica which constitutes an average of 12 and 16% of the inorganic mass of non-ruminant and ruminant dung, respectively (Holter, 2016). The presence of ash is expected to lower food quality, reducing organic matter concentrations and therefore the availability of nutritional food to the insect (Holter, 2016).

A recent review by Holter (2016) has collected information on the nutritional composition of a large variety of herbivorous mammals to better understand the quality of food afforded to dung beetles. Adult dung beetles feeding on wet dung typically exhibit soft saprophagy, feeding on very small particles in the dung which are generally high in nutritional content (Holter et al., 2002), sifting out large plant fragments (Anderson and Coe, 1974). Holter (2016) hypothesised that it was likely that wet dung feeding beetles feed primarily on microbial biomass and the associated nutrients in the dung. However, coprophagous *Pachysoma* feed on dry dung (Harrison et al., 2003) with studies finding evidence that *Pachysoma* have adapted to feed on the refractive materials in both dung and plant detritus (Holter and Scholtz, 2011, 2013).
Little information is available on dry dung composition compared to that of wet dung (Holter, 2016). As dung ages, concentrations of water-soluble organic matter, hemicellulose and cellulose may decrease significantly, while lignin, ash and water-soluble protein concentrations are likely to remain constant or increase (Lodha, 1974, Holter, 2016). Concentrations of phosphorous also differ between wet and dry dung (McDowell and Stewart, 2005). Nitrogen concentrations are expected to decrease as the dung ages (Holter, 2016). 

$\text{CO}_2$ production from the dung rapidly increases for the first few days, decreasing as the dung ages and dries out (Anderson and Coe, 1974). $\text{CO}_2$ production is increased when moisture is re-introduced into dry dung (Anderson and Coe, 1974). Some \textit{Pachysoma} species (\textit{P. striatum}) dig burrows below the moisture line, and in so doing may reintroduce moisture into the detritus (Scholtz, 1989). Therefore, it is possible that dry dung has undergone some preconditioning by the involvement of microbial communities, leading to suggestions that \textit{Pachysoma} may feed on the microbial communities and not the refractive materials within the diet (C. Sole Pers. Comm.). However, no evidence of this has been found (Holter et al., 2009). Furthermore, \textit{Pachysoma} have adapted mouthparts and gut systems that suggest they have evolved to feed on cellulosic diets (Holter and Scholtz, 2011, 2013).

1.3.2 An Overview of Plant Cell Wall Degradation

Plant cell walls form a major component of the diet of both detritivorous and coprophagous \textit{Pachysoma} species. Lignin and cellulose are the most abundant natural polymers (Pérez et al., 2002). Cellulose and hemicelluloses are sugar-derived macromolecules, while lignin is an aromatic polymer derived from phenylpropanoids (Pérez et al., 2002). A brief overview of the degradation of lignin, hemicellulose and cellulose is provided, with a focus on microbial degradation.

The first process in plant cell wall degradation is depolarization of lignin (Calderón-Cortés et al., 2012). Lignin is a highly recalcitrant material that is able to withstand the effects of enzymes and chemicals (Ohkuma, 2003). As such, lignin protects other plant cell wall materials including cellulose (Ohkuma, 2003). Few known taxa of organisms are able to
successfully break-down lignin (Ohkuma, 2003). However, there is strong evidence that some insects feeding on healthy wood are able to digest lignin, presumably with the assistance of specific fungal gut symbionts (Geib et al., 2008). Enzymes involved in lignin degradation include laccases, ligninolytic peroxidases (lignin peroxidase, manganese peroxidase and versatile peroxidase), oxidases, dehydrogenases associated with mycelium, aryl-alcohol dehydrogenases (fungal origin) and quinone reductases (Guillén et al., 2005).

Degradation of hemicellulose is the second step in plant cell wall degradation (Calderón-Cortés et al., 2012). Hemicelluloses are non-cellulosic polysaccharides (excluding pectin and starch) forming part of the plant cell wall (Watanabe and Tokuda, 2010). The degradation of cellulose is the final step in plant cell wall degradation (Calderón-Cortés et al., 2012). Enzymes which digest cellulose are collectively termed cellulases and are divided into three classes of hydrolytic enzymes: endoglucanases, exoglucanases (cellobiohydrolases) and β-glucosidases (cellobiases) (Terra and Ferreira, 1994, Lynd et al., 2002, Gilbert, 2010, Watanabe and Tokuda, 2010).

1.4 Structure of the Insect Gut

The insect gut is a long muscular tube (Figure 1.3) extending from the insect’s mouth to the anus and is divided into the fore-, mid- and hindgut (Chapman, 1998, Brune, 2013). The foregut, consisting of the buccal cavity, pharynx, oesophagus and crop, begins at the oral cavity and extends to the proventriculus, a simple valve at the entrance of the midgut (Terra and Ferreira, 1994, Chapman, 1998). In several insect species, the proventriculus is a more complex structure, assisting with filtering of food or may have circular plates which grind food at entry into the midgut (Chapman, 1998). Food may be stored in the crop (Terra and Ferreira, 1994, Chapman, 1998) before moving into the midgut. The midgut is the longest gut segment, extending across most of the body cavity (Huang et al., 2010). A characteristic feature of the midgut is the presence of rings of caeca (Huang et al., 2010). Peristaltic movements of circular and longitudinal muscles aid the movement of food through the gut (Vallet-Gely et al.,
Digestions mainly takes place in the midgut where enzymes are secreted and nutrients are absorbed (Terra and Ferreira, 1994, Chapman, 1998). From the midgut, food moves into the hindgut, often via the pyloric sphincter (pylorus) which functions to restrict movement of food back into the midgut (Terra and Ferreira, 1994, Huang et al., 2010). The hindgut is made-up of the ileum and rectum (Terra and Ferreira, 1994, Chapman, 1998). The hindgut is responsible for nutrient and water absorption in some insect groups such as termites (Potrikus and Breznak, 1981, Terra and Ferreira, 1994). The hindgut of different insect species can differ, as some insects retain a straight gut, while others form large dilated compartment(s) commonly termed a fermentation chamber (Terra and Ferreira, 1994, Brune and Friedrich, 2000, Huang et al., 2010).

This fermentation chamber is where a large portion of the gut microbial community is found (Brune and Friedrich, 2000). For example, termites and scarab larvae support sac-like hindgut dilations (Brune and Friedrich, 2000, Huang et al., 2010). Certain insects such as families within Orthoptera do not have fermentation chambers but rather have other specialised hindgut features where the majority of microbes are found (Nation, 1983). Xylophagous species or those with extra-corporeal digestive symbionts generally do not have a fermentation chamber present within the hindgut (Watanabe and Tokuda, 2010).

The fore- and hindgut form from the embryonic ectoderm, resulting in the secretion of a cuticle (intima) (Chapman, 1998) to protect the gut. This lining is shed during ec dysis (i.e. the shedding of the cuticle). In contrast, the midgut is derived from endodermal cells (Chapman, 1998) and in most insect species creates a network of chitin and proteins called the peritrophic membrane (PM) which forms around the food bolus (Terra and Ferreira, 1994, Chapman, 1998, Shao et al., 2001, Terra, 2001, Wang and Granados, 2001). The PM forms a layer between the gut epithelium and the ingested food particles (Chapman, 1998). The PM also acts to concentrate dietary substrates and the necessary digestive enzymes. The formation of the PM results in the division of the midgut into the endo- and ectoperitrophic spaces (Terra and Ferreira, 1994, Chapman, 1998). The PM can be categorised into two types according to
region of the midgut in which it is formed. Type I forms across the entire midgut, typically when specific food is ingested. The Type II PM only forms in a specific area within the anterior region of the midgut, regardless of the presence of food (Lehane, 1997, Terra, 2001). Small pores throughout the PM allow enzymes and small particles to move between the endo- and ectoperitrophic spaces while confining the majority of microorganisms to the endoperitrophic space (Spence and Kawata, 1993, Barbehenn and Martin, 1995, Chapman, 1998, Edwards and Jacobs-Lorena, 2000).

Redox potential, pH, and oxygen and hydrogen concentrations may vary according to the specific gut segment (Brune and Friedrich, 2000, Brune, 2013). Furthermore, gut pH varies according to insect species (Terra and Ferreira, 1994). pH may range from acidic to alkaline varying across insect species and the respective gut segment (Terra and Ferreira, 1994).

The Malpighian tubules are the excretory organs in insects, extending from the hindgut (the pylorus is the site of attachment) into the body cavity (Terra and Ferreira, 1994, Chapman, 1998, Engel and Moran, 2013). The placement and number of Malpighian tubules may differ according to insect species as has been noted in ants (Cook and Davidson, 2006). These organs absorb waste products which are transported to the anterior hindgut, where nitrogenous and food waste accumulates (Engel and Moran, 2013).

![Figure 1.3: Basic structure of the larval scarab gut (adapted from Huang et al. (2010)).](image-url)
1.5 Insect Gut Microbial Communities

1.5.1 Microbial Colonization of the Gut Environment

The insect gut is an extreme environment, with unstable conditions making colonisation by microbial organisms difficult. Firstly, microorganisms need to persist against harsh physiochemical conditions, digestive processes and host immune defence (Douglas, 2015). Secondly, insects undergo several events in which the gut is shed, affecting the in situ microbial communities (Broderick and Lemaitre, 2012, Engel and Moran, 2013, Douglas, 2015). In insect groups producing a PM, microorganisms are unavoidably expelled during regular shedding of this lining (Engel and Moran, 2013). Likewise, larvae moult several times during development, at which time the lining of the fore- and hindgut is shed, along with it any inhabiting microorganisms (Douglas, 2011, Broderick and Lemaitre, 2012, Engel and Moran, 2013). This results in variable gut microbiomes across the different life stages in holometabolous insects (Engel and Moran, 2013). However, moulting events stop once the insect is an adult, providing a more stable environment for colonisation of the fore- and hindgut (Engel and Moran, 2013).

Microorganisms found in the insect gut environment are typically either resident (naturally occurring), successfully inhabiting the gut environment, or transient, acquired through contact with the environment or ingested with the food (De Vries et al., 2001, Bright and Bulgheresi, 2010). While a small number of transient microorganisms may be harmful (pathogenic) to the host, the majority have little influence on the host, typically being removed from the gut (Douglas, 2009). Some ingested microorganisms are beneficial to the host, as a food-source, assisting with digestion of the ingested food and supplementing the nutritional needs of the host (Douglas, 2009). Transient microorganisms may be prevented from colonising the insect gut by the presence of the resident microflora (Veivers et al., 1982). However a small number of bacteria are capable of persisting in the gut (Vallet-Gely et al., 2008).
Resident microorganisms inherited by the host can either be obligate or facultative. Obligate or primary microorganisms are evolutionarily ancient, living in bacteriocytes (specialised host cells) and are expected to assist the host nutritionally (Dale and Moran, 2006, Moran et al., 2008, Ferrari and Vavre, 2011). However, they may not necessarily be required by the host (Moran et al., 2008). Obligate microorganisms typically have reduced genome size compared to their free-living counterparts (Dale and Moran, 2006, Nakabachi et al., 2006, McCutcheon and Moran, 2012, Bennett and Moran, 2013), retaining genes necessary for interactions with the host (Dale and Moran, 2006, Bennett and Moran, 2013). Facultative microorganisms (secondary microorganisms) are acquired through maternal horizontal transmission and are further categorised as mutualistic or parasitic (Dale and Moran, 2006, Moran et al., 2008, Bright and Bulgheresi, 2010). Facultative microorganisms either invade the tissues and cells of new insect host species (Dale and Moran, 2006) or reside in the hemocoel (Fukatsu et al., 2000). Distribution of facultative mutualists (and the respective effect on the host) is typically influenced by environmental factors, such as location and temperature (Tsuchida et al., 2002, Ferrari and Vavre, 2011).

Gut microorganisms can have both a physiological and evolutionary effect on the host (Douglas, 2011). For example, microbial composition, abundance and activity may impact the physiological welfare of the host (Douglas, 2011). As such, microorganisms may have a negative effect on the host, impacting host growth and metabolism (Broderick and Lemaitre, 2012) and influencing host behaviour (Hosokawa et al., 2008, Sharon et al., 2010, Najarro et al., 2015, Wittman and Fedorka, 2015). The host is able to control microbial (pathogenic) colonisation using two different mechanisms: resistance and tolerance (Schneider and Ayres, 2008). Resistance limits pathogen/microbial colonisation, whereas tolerance reduces the impact that pathogen/microbial colonisation would have on the host (Schneider and Ayres, 2008). Therefore, tolerance mechanisms are proportional to the population size of microorganisms inhabiting the gut of the respective host, with resistance mechanisms preferred in cases of low density populations (Engel and Moran, 2013). Different mechanisms
have been adapted by the host. For example, the host may produce enzymes (e.g. lysozymes) that damage bacterial cell walls, thereby decreasing abundance of microorganisms in either the whole or particular sections of the gut (Daffre et al., 1994, Fujita et al., 2001, Engel and Moran, 2013, Nayduch and Joyner, 2013). The resident microbiota may also assist in reducing colonization of other microorganisms in the gut (Vollaard and Clasener, 1994, Dillon and Dillon, 2004). In contrast, physical features such as the gut linings (e.g. PM) protecting the epithelium restrict access of the microorganisms to the insect itself, allowing the host to tolerate microbial colonisation of the gut (Engel and Moran, 2013).

The types of microorganisms commonly found within the insect gut include, bacteria (Ahn et al., 2012, Boissière et al., 2012, He et al., 2013, Scully et al., 2013, Dietrich et al., 2014, Gauthier et al., 2015), eukaryotes (including fungi)(Suh et al., 2004, Nguyen et al., 2007, Scully et al., 2013, Santana et al., 2015), archaea (Egert et al., 2003, Lemke et al., 2003, Santana et al., 2015) and viruses (Jia et al., 2013, Liu et al., 2013, Chandler et al., 2015).

Bacteria are typically the most abundant microorganisms in the guts of the majority of insect species (Engel and Moran, 2013). The diversity of bacterial taxa within insect guts is far less than mammals with bacterial operational taxonomic units (OTUs) within the majority of insect guts being substantially lower than in mammalian gut systems (Dethlefsen et al., 2008, Boissière et al., 2012, Andongma et al., 2015, Diouf et al., 2015). These differences are likely due to the insect gut environment being affected by moults and other perturbation events as well as the shorter life-span of insects when compared to mammals (Broderick and Lemaitre, 2012). The bacterial phyla Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria are commonly abundant in insect guts (Jones et al., 2013, Yun et al., 2014). Archaea (particularly Thermoplasmatales, Halobacteriales and methanogens) have been recorded in insects such as termites, scarab beetles and cockroaches (Egert et al., 2003, Lemke et al., 2003). Protists, mainly of the phyla Metamonada, Parabasalia and Preaxostyla, appear specific to lower termites and wood-feeding cockroaches (Hongoh, 2010, Douglas, 2015). Fungi are not the only microbial eukaryotes reported in insect guts. Anaerobic ciliates (Clevelandellida) appear
1.5.2 Variations in Gut Microbial Communities

Variations in gut microbiome assembly are evident within an individual insect and between insects. Within an individual, gut microbial communities may differ axially across the different gut segments. The majority of microorganisms colonising the foregut are found in the crop (when present) (Köhler et al., 2012, Schauer et al., 2012). As the crop functions as a temporary storage space of food, microorganisms are expected to be removed regularly and move into the midgut (Douglas, 2015). In the midgut microbial colonisation is complicated by the secretion of digestive enzymes and a high immune response which may result in inhibited growth or death of the microbial population (Vallet-Gely et al., 2008, Douglas, 2015). Microbial communities successfully colonising the midgut are presumably able to withstand extremes of pH, ionic strength, redox potential and digestive enzymes of this gut segment (Vallet-Gely et al., 2008). Microbial colonisation of the midgut is further complicated when a PM is produced as penetration of this membrane by microorganisms is difficult, resulting in the majority of microbial symbionts being removed by peristalsis with the food (Douglas, 2015). However, insect or microbial chitinases may facilitate the movement of microorganisms across the PM (Tsai et al., 2001, Dostálová and Volf, 2012).

Microbial populations are most dense in the hindgut (particularly the ileum) where digestive enzymes and desiccation stress are negligible (Douglas, 2015). The high nutritional content (in the form of ions and metabolites in the Malpighian waste products) of the hindgut also promotes microbial growth (Douglas, 2015). Microbial community structure may also differ radially across the gut segment, i.e. between the gut wall and gut lumen (Köhler et al., 2012). Gut wall microorganisms persist in the gut, increasing their interactions with the host compared to gut lumen microbiota, which are typically removed with the food (Douglas, 2015).
Variations between insects may include large intra-specific differences between bacterial communities of insects of the same species (Ahn et al., 2012, Boissière et al., 2012, Osei-Poku et al., 2012, Montagna et al., 2015). Such variations may be driven by host genetics (Douglas, 2015) although non-genetic drivers have also been identified (Chandler et al., 2011, Osei-Poku et al., 2012, Wong et al., 2013). Intra-specific microbiome assembly may therefore not be driven only by deterministic processes but may be influenced by stochastic events including ingestion of food-based microorganisms, relationships with other microorganisms in the gut and the gut wall attachment site (Douglas, 2015).

Interactions between different microorganisms successfully colonising the gut can be either negative (antagonistic) or positive in nature. Negative interactions (competition or amensalism) influence the co-occurrence of microbial species and increase host-microbial variation, while positive (commensalism or mutualism) interactions often result in a persistent core microbiome (Wong et al., 2013). A core microbiome can be defined as “Organisms common across microbiomes hypothesized to play a key role in ecosystem function within a habitat.” (Shade and Handelsman, 2012, p. 5). Different models of core microbiomes have been defined for human microbiomes (Hamady and Knight, 2009), ranging from “substantial” (large number of OTUs shared across all individuals), “subpopulation” (OTUs only shared amongst individuals of same subpopulation), “gradient”, “minimal” and “no core”. The persistence of a core microbiome may indicate several scenarios: direct transmission of microbial communities between hosts, selectivity of microbial colonisation displayed by the host or the adaptive quality of microorganisms to successfully inhabit the gut (Engel and Moran, 2013).

1.5.3 Gut Microbial Symbionts and the Relationship between Host Diet and the Gut Microbiome

Symbiosis is any association between different species i.e. microorganisms (mutualist, commensal or parasite) that persistently associate with an insect host (Zook, 1998, Dillon and
Dillon, 2004, Douglas, 2007). It is important to note, however, that microorganisms may relate differently to different hosts and under different conditions, depending on genotype and the physiochemical conditions (i.e., a microorganism may be symbiotic in one situation and pathogenic in another) (Dillon and Charnley, 2002, Ffrench-Constant et al., 2003, Douglas, 2007, Chandler et al., 2008).

Gut microbial symbionts have been studied in an array of insect species and have been reported to assist the host in various ways including production of components of cohesion pheromones (Dillon et al., 2002), protection against parasites/pathogens and other harmful agents (Kaltenpoth et al., 2005, Kaltenpoth, 2009, Douglas, 2011), changing host body colour (Tsuchida et al., 2010), mediating thermal tolerance (Dunbar et al., 2007), contributing to fungus-garden health (Currie et al., 1999, Currie et al., 2003, Kaltenpoth, 2009) and nutrition (Douglas, 2009).

A number of insects are unable to synthesize certain essential nutritional elements including amino acids (Wilson et al., 2010, Douglas, 2013), certain vitamins linked to metabolic enzyme functioning (Douglas, 2013) and sterols (Zdobnov et al., 2002, Behmer and Nes, 2003, Janson et al., 2009). Insects rely on their food-source for acquisition of such nutrients, although some insects feed on diets deficient in these essential components (Douglas, 2013), making it necessary to find alternative methods for acquiring these nutrients.

One such alternative is the acquisition of microbial symbionts that assist the host in exploiting nutrient deficient dietary lifestyles (Akman et al., 2002, Wu et al., 2006). For example, blood-feeding insects such as mosquitoes may be deficient in B vitamins, those feeding on plant sap may lack a source of essential amino acids, and sound-wood feeders are deprived of nitrogen (Douglas, 2009, Douglas, 2011). Microbial symbionts may assist the host in degradation of plant cell walls (Park et al., 2007, Morales-Jiménez et al., 2009, Calderón-Cortés et al., 2012), nutrient, amino acid and vitamin production (Nakabachi and Ishikawa, 1999, Akman et al., 2002, Wu et al., 2006), nitrogen fixation (Nardi et al., 2002, Morales-Jiménez et al., 2009) and production/source of sterols (Janson et al., 2009). Multiple gut
microbial symbionts may collaborate to provide complete nutritional support for the host (McCutcheon and Moran, 2010).

Diet has been shown to have a pronounced effect on gut microbiome assembly. For example, pronounced differences are noted between wild-type insects feeding on natural diets and lab-reared insects fed on artificial diets (Chandler et al., 2011). Furthermore, microbiomes are altered when the major nutritional class of the food (or its concentration) is altered (i.e. the amount of protein, lipid, sugar and/or fibre is altered) (Huang et al., 2013). In the well-studied Drosophila model, wild fly microbiota show greater similarity between different Drosophila species feeding on the same substrate than closely related Drosophila species feeding on different diets (Chandler et al., 2011).

A core microbiome may still persist regardless of the influence of host diet (Huang et al., 2013, Schauer et al., 2014). Physiochemical gut conditions tend to favour particular microbial species and it is therefore expected that gut microbial communities should not be random even when some microbial species are acquired from the environment/diet (Engel and Moran, 2013).

1.5.4 Previous Research on Scarab Beetle Gut Microbiomes

The author is unaware of any previously published studies on the gut microbiomics of dry detritus feeding insects. Very few studies have investigated the gut microbiomes of scarab beetles, and most have concentrated on the larval life stage (Egert et al., 2003, Lemke et al., 2003, Egert et al., 2005, Zhang and Jackson, 2008, Andert et al., 2010, Huang et al., 2010, Huang et al., 2012, Huang and Zhang, 2013). Based on studies of other insect classes, it is expected that the larval gut microbiomes will be different to those of the respective adult scarab beetles (Engel and Moran, 2013).

Previous studies on scarab beetle gut microbiomes are largely restricted to the use of clone libraries and Terminal Restriction Fragment Length Polymorphism (TRFLP)/ Denaturing gradient gel electrophoresis (DGGE) analyses and not newer sequencing technologies (Egert

The midgut microbiota of *Pachnoda ephippiata* (Andert et al., 2010), *Pachnoda marginata* (Andert et al., 2010) and *Melolontha melolontha* scarab beetle larvae are less diverse than hindgut communities (Egert et al., 2005). Apart from axial differences, radial differences in gut microbial communities were also noted in *M. melolontha* (Egert et al., 2005). Bacteria showing CMCase activity were found in the hindgut of *Holotrichia parallela* but were typically absent from the midgut (Huang et al., 2012). Huang et al. (2012) also found *Pseudomonas* to be the most abundant taxon of cellulolytic bacteria. Archaea were also detected in the guts of *M. melolontha* (Egert et al., 2005) and *P. ephippiata* (Egert et al., 2003, Lemke et al., 2003).

The effect of environmental factors on scarab gut microbiomes has also been reported with location appearing to significantly impact gut microbiome assembly of *H. parallela* (Huang and Zhang, 2013) and *Costelytra zealandica* (Zhang and Jackson, 2008). Furthermore, Andert et al. (2010) noted that variations in gut microbes between two *Pachnoda* species were greater influenced by host phylogeny than diet. In contrast, Zhang and Jackson (2008) found distinct differences between midgut bacterial communities of *C. zealandica* fed on different diets and starved. However, no such differences were noted in the hindguts of the same individuals. Significant changes in gut bacterial communities have also been noted during larval development in *H. parallela* (Huang and Zhang, 2013) and *Onthophagus taurus* (Estes et al., 2013).

1.6 **Sequence-Based Methods in Microbial Ecology**

Modern advances in sequencing technology have resulted in the next-generation sequencing (NGS) platforms, which are alternatively named deep or massively parallel sequencing (Metzker, 2010, Chiu and Miller, 2016). These newer technologies result in millions of reads in a single run (Mardis, 2008, Chiu and Miller, 2016). This is accomplished
through preparation of libraries consisting of Deoxyribonucleic acid (DNA) fragments which have not been subjected to cloning or culture-based methods (Mardis, 2008, Morozova and Marra, 2008). The first NGS platform made available was the Roche 454 pyrosequencing instrument (Liu et al., 2012), which was quickly followed by the Illumina (MiSeq/HiSeq/NextSeq), ABI SOLiD, Life Technologies Ion Torrent, and the PacBio RS systems. For the purpose of this discussion, only technologies and platforms used in this study will be described further.

1.6.1 Amplicon-based Pyrosequencing

Pyrosequencing is a bioluminescence sequencing technique that works on measurements of pyrophosphate released during DNA synthesis (Ronaghi, 2001). Pyrosequencing was developed as an alternative to Sanger sequencing at the Royal Institute of Technology (KTH) in Stockholm (Fakruddin et al., 2012). The Roche 454 pyrosequencer uses emulsion polymerase chain reaction (PCR) to amplify single DNA isolates for library construction (Rothberg and Leamon, 2008). Emulsification PCR requires an oil-water interface to produce droplets, whereby each droplet (microreactor) contains a bead which bonds covalently to a single DNA template with attached adaptors (Mardis, 2008, Schlebusch and Illing, 2012). Amplified fragments are produced when PCR is performed across the bead’s surface (Schlebusch and Illing, 2012). The Roche 454 pyrosequencer deposits the beads into individual Picotiter plate wells (Hodkinson and Grice, 2015, Goodwin et al., 2016) where sequencing agents (DNA polymerase, Adenosine triphosphate (ATP) sulfurylase, and luciferase) are then added (Hodkinson and Grice, 2015). A single deoxyribose nucleoside triphosphates (dNTP) is added in limited concentrations to extend the primer and discontinue DNA synthesis, which is restarted by the addition of another dNTP (Metzker, 2010). This prompts pyrophosphate release during complementary strand synthesis by nucleotide incorporation (Hodkinson and Grice, 2015). Pyrophosphate release in turn produces a fluorescent signal which is recorded by a charge coupled device (CCD) camera for base calling (Metzker, 2010, Hodkinson and Grice, 2015, Chiu and Miller, 2016).
Fast sequencing time (Liu et al., 2012), longer read lengths resulting in greater identification accuracy and precision (Hodkinson and Grice, 2015), easily automated system (Ronaghi, 2001), parallel processing (Ronaghi, 2001) and overcoming bias introduced through cloning-based methods (Siqueira et al., 2012) are the general advantages of 454 pyrosequencing. Unlike earlier methods, pyrosequencing foregoes the need for gel electrophoresis and labelled reagents such as primers (Ronaghi, 2001). However, as with all sequencing technologies errors do occur, with the most notable being resulting poly-bases (homopolymers) longer than 6 base pairs (bp) (Liu et al., 2012, Hodkinson and Grice, 2015), resulting in measured insertions or deletions being introduced (Metzker, 2010, Schlebusch and Illing, 2012). Further disadvantages include the high reagent cost (Liu et al., 2012). Due to advances in newer sequencing platforms 454 life sciences are discontinuing the platform from 2016 (Hodkinson and Grice, 2015). As with all amplicon-based studies, bias may be introduced through specific primers and PCR amplification cycles (Hodkinson and Grice, 2015).

1.6.2 Environmental Shotgun Sequencing

Shotgun sequencing foregoes targeting a specific genomic locus through shearing all DNA into small fragments that are sequenced independently (Sharpton, 2014). The resultant reads are assembled into larger fragments. This technique results in both taxonomical and functional data being obtained (Eisen, 2007, Sharpton, 2014, Hodkinson and Grice, 2015, Oulas et al., 2015). As no universal marker gene is known for all domains of life, shotgun sequencing has the advantage of deciphering the entire microbial community (Hodkinson and Grice, 2015). By foregoing DNA amplification and associated biases, a clearer picture of exact relative abundances of organisms is given (Hodkinson and Grice, 2015).

Unfortunately, reference genomes are currently limited, decreasing classification efficiency (Petrosino et al., 2009, Hodkinson and Grice, 2015). Furthermore, shotgun sequencing relies on samples with large DNA concentrations (Petrosino et al., 2009, Hodkinson and Grice, 2015). High host DNA concentrations may also hinder the number of
target reads obtained (Petrosino et al., 2009, Sharpton, 2014). The cost of generating whole metagenomes is far greater than amplicon-based studies particularly when the majority of reads obtained are from the host (Sharpton, 2014).

Whole shotgun sequencing was performed using the Illumina Hiseq 2500 platform. Illumina technology uses a sequencing-by-synthesis approach performed in a flow cell (Buermans and Den Dunnen, 2014, Hodkinson and Grice, 2015). Adapter sequences are introduced during library preparation (Liu et al., 2012, Buermans and Den Dunnen, 2014, Hodkinson and Grice, 2015). During loading, double stranded DNA is denatured into single stranded DNA molecules in order to produce a copy of the original template (Liu et al., 2012, Buermans and Den Dunnen, 2014). The original template is removed and replicated (Buermans and Den Dunnen, 2014) using bridge amplification to form clusters consisting of identical fragments (Liu et al., 2012, Hodkinson and Grice, 2015). Labelled dye-terminator nucleotides are washed over and allowed to attach to the DNA fragments (Liu et al., 2012, Buermans and Den Dunnen, 2014, Hodkinson and Grice, 2015). Following imaging, the fluorescent group is removed and the terminator deactivated, allowing for the process to be repeated by the subsequent addition of nucleotides (Buermans and Den Dunnen, 2014, Hodkinson and Grice, 2015). The Illumina Hiseq 2500 platform boasts read lengths of up to 250bp with a throughput of 125-150Gb (Goodwin et al., 2016).

1.7 Aims and Objectives

1.7.1 Aims

*Pachysoma* species exhibit an unusual feeding behaviour, with species feeding on different diets. Furthermore, most dung beetles feed on wet dung. However, *Pachysoma* spp. utilise the dry alternative (as well as dry plant detritus), removing it from the environment. Unlike wet dung feeders, *Pachysoma* spp. appear unable to remove the plant fragments from the dung before consumption and instead feed on the plant biomass in the dung. The role gut microorganisms may play in the digestion of the two food-sources (dry dung and plant detritus)
as well as how these different diets affect gut microbiome assembly of this ecologically important insect species has not been previously studied. Therefore, this study aims to explore the relationship between the *Pachysoma* host and the associated gut microbiota.

### 1.7.2 Objectives

#### 1.7.2.1 Diversity Analysis of *Pachysoma* Gut Microbiomes

The technical objectives of this study included:

- The use of amplicon sequencing to characterise the gut bacterial and fungal communities of two desert insects feeding on dry plant detritus and dung pellets.
- A comparison of intra- and inter-specific bacterial and fungal gut communities, to determine any potential influence of host diet and/or host phylogeny.
- The use of shotgun sequencing to determine the composition of the entire gut microbiome of both *Pachysoma* species.

#### 1.7.2.2 Functional Analysis of *Pachysoma* Gut Microbiomes

- The use of shotgun sequencing of the whole *Pachysoma* gut microbial metagenome to determine the functional capacity of the gut community.
- Determination of correlations between the functional genes of the microbial gut communities and the different diets (dung versus plant detritus) of the two *Pachysoma* species.

### 1.8 References


© University of Pretoria


© University of Pretoria


© University of Pretoria


Chapter 2: Materials and Methods

2.1 Chemical Reagents

Chemical reagents, buffers, enzymes and the respective suppliers are listed in Table 2.1-2.3 below.

Table 2.1: List of chemical reagents.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Supplier</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid</td>
<td>Merck</td>
<td>Germany</td>
</tr>
<tr>
<td>Ammonium acetate (NH₄Ac)</td>
<td>Merck-Saarchem</td>
<td>South Africa</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>Sigma-Aldrich</td>
<td>USA</td>
</tr>
<tr>
<td>Chloroform</td>
<td>Merck</td>
<td>Germany</td>
</tr>
<tr>
<td>Disodium phosphate (Na₂HPO₄)</td>
<td>Merck-Saarchem</td>
<td>South Africa</td>
</tr>
<tr>
<td>Ethanol (EtOH)</td>
<td>Illovo sugar</td>
<td>South Africa</td>
</tr>
<tr>
<td>Ethylenediaminetetraacetic acid (EDTA)</td>
<td>Merck</td>
<td>Germany</td>
</tr>
<tr>
<td>GelRed™ (10000x)</td>
<td>Biotium</td>
<td>USA</td>
</tr>
<tr>
<td>Glycerol</td>
<td>Merck-Saarchem</td>
<td>South Africa</td>
</tr>
<tr>
<td>Hydrochloric acid (HCl)</td>
<td>Merck-Saarchem</td>
<td>South Africa</td>
</tr>
<tr>
<td>Isoamyl alcohol</td>
<td>Merck</td>
<td>Germany</td>
</tr>
<tr>
<td>Monopotassium phosphate (KH₂PO₄)</td>
<td>Merck-Saarchem</td>
<td>South Africa</td>
</tr>
<tr>
<td>SeaKem LE agarose</td>
<td>Lonza</td>
<td>USA</td>
</tr>
<tr>
<td>Sodium chloride (NaCl)</td>
<td>Sigma-Aldrich</td>
<td>USA</td>
</tr>
<tr>
<td>Tris (hydromethyl) aminomethane (Tris base)</td>
<td>Merck</td>
<td>Germany</td>
</tr>
</tbody>
</table>

Table 2.2: Enzymes used in this study.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Supplier</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysozyme</td>
<td>Sigma-Aldrich</td>
<td>USA</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Sigma-Aldrich</td>
<td>USA</td>
</tr>
<tr>
<td>DreamTaq™ DNA polymerase</td>
<td>Fermentas Life Science</td>
<td>Lithuania</td>
</tr>
</tbody>
</table>
2.2 General Methodologies

2.2.1 Collection and Storage of *Pachysoma* Samples

Adult individuals of *P. endroedyi* (Figure 2.1a) and *P. striatum* (Figure 2.1b) from single breeding populations, were collected (Figures 2.2a and b) by the Scarab Research Group from the University of Pretoria. Samples were collected in September and October 2014 (for the amplicon sequencing study; see chapter 3) and a separate set of samples were collected in August 2015 (for the shotgun sequencing study; see chapter 4) from coastal sandveld near Kommandokraal, Namaqualand, South Africa (S31°29'58.4" E18°12'29.2"; Figures 2.3-2.4) under the Cape Nature permit number 0056-AAA008-00041. Beetles were identified on site. Due to their large body size, 99% ethanol was injected into their abdomens using sterile syringes for gut preservation (Montagna et al., 2015). Insects were then stored in 99% ethanol at -80°C, until dissection. Furthermore, five cockroach (*Blaptica dubia*) samples were sourced from a pet shop for comparison of host DNA contamination reduction methods (Chapter 4). Gut dissections of cockroach samples were carried out as described for the *Pachysoma* samples (Section 2.2.3).

---

### Table 2.3: Compositions of buffers and solutions used in this study.

<table>
<thead>
<tr>
<th>Buffer/Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5M EDTA</td>
<td>37.22% [w/v] in dH2O</td>
</tr>
<tr>
<td>0.5M Tris-HCl</td>
<td>6% [w/v] Tris-base in dH2O</td>
</tr>
<tr>
<td>Ringer solution</td>
<td>0.12 g/L CaCl₂, 0.105 g/L KCl, 0.05 g/L NaHCO₃, 2.25 g/L NaCl (Sigma-Aldrich, USA)</td>
</tr>
<tr>
<td>CTAB</td>
<td>0.1M Tris-HCl [pH8.0], 1.4M NaCl, 0.02M EDTA [pH8.0]</td>
</tr>
<tr>
<td>Phosphate buffered saline (PBS)</td>
<td>8 g/liter NaCl, 0.2 g/liter KCl, 1.44 g/liter Na₂HPO₄, 0.24 g/liter KH₂PO₄ [pH 7.4]</td>
</tr>
<tr>
<td>Tris-acetate-EDTA (TAE)</td>
<td>40mM Tris (pH 7.6), 20mM acetic acid, 1mM EDTA</td>
</tr>
<tr>
<td>6x GelRed™-Loading buffer solution</td>
<td>30x GelRed™, 5ml 6x Loading buffer</td>
</tr>
<tr>
<td>6x Loading buffer</td>
<td>0.25% [w/v] Bromophenol blue, 40% [v/v] glycerol</td>
</tr>
</tbody>
</table>
Figure 2.1: *Pachysoma* samples collected for the present study: a) *P. endroedyi* and b) *P. striatum* (Courtesy of Hennie de Klerk).

Figure 2.2: Collection of *Pachysoma* samples: a) Observation of a burrow; b) a typical *Pachysoma* spp. burrow with the food-source (plant detritus) and a beetle inside (Courtesy of Hennie de Klerk).
Figure 2.3: Aerial view of the Western coast of South Africa with the Namaqualand Desert collection site indicated by the yellow marker and shown in the insert (Insert map data: Google, AfriGIS (Pty) Ltd, 2016).

Figure 2.4: Burrow being examined surrounded by typical vegetation at the collection site (Photo courtesy of H. de Klerk).
2.2.2 Insect Identification

Morphological identification of each species was carried out by Prof. C. Scholtz and experts from the Scarab Research Group of the University of Pretoria. Taxonomic keys (Scholtz and Holm, 1985, Harrison et al., 2003) were used to morphologically identify the two Pachysoma species (see Chapter 3).

2.2.3 Gut Dissection

Gut dissections were performed under a Zeiss Stemi 2000-C dissection microscope (Zeiss, Oberkochen, Germany) as previously described (Correa et al., 2012) with modifications. All equipment was sterilised before use with 10% bleach and 70% EtOH. The average body length of *P. endreodyi* ranges from 20.7-26.4mm, and the one of *P. striatum* ~19 mm (Harrison et al., 2003). The insects were placed in a wax-lined glass Petri dish with quarter strength autoclaved Ringer solution. The thorax and abdominal integument were removed using scissors before pinning the specimen to the wax layer in the Petri dish. Forceps were used to remove the membranes covering the internal organs. The rectum was pulled downwards, moving the gut gently out of the body cavity. The mid- and hindguts were separated and stored in 1.5ml eppendorf tubes at -20°C until metagenomic DNA (mDNA) extraction for the amplicon study (Chapter 3). For the shotgun sequencing study (Chapter 4), whole-guts were removed and stored at -20°C until mDNA extraction. Photographs were taken of both the mid- and hindgut using an AxioCam ERc5s (Zeiss, Oberkochen, Germany, Figure 2.5).
Figure 2.5: Dissected P. striatum showing H: hindgut, M: Midgut and P (square): pyloris (Foregut not shown).

2.2.4 Analytical Techniques

2.2.4.1 Gel Electrophoresis

Metagenomic DNA (see Sections 2.3.1 and 2.4.1) and PCR amplicons (see Section 2.3.2) were separated using 1% agarose gels prepared with 1X TAE buffer and SeaKem LE agarose. Samples were loaded with a “1X Gel Red / 3X loading buffer” mixture. After migration (100V, 35-60min), DNA was visualised under Ultraviolet (UV) light (Ultraviolet transilluminator, Spectroline, USA). Amplicon and mDNA sizes were determined by comparing their migration to the KAPA Universal Ladder (KAPA Biosystems, USA). Photographs of agarose gels were captured with the Molecular Imager Gel Doc XR+ (BioRad, USA).

2.2.4.2 Spectrophotometry

The Nanodrop 2000 spectrophotometer (Thermoscientific, Massachusetts, USA) was used to determine DNA concentrations and quality. DNA concentrations were calculated as a
value of [OD260 nm x ng/μl]. The purity of each sample was determined using the OD$_{260/280}$ and OD$_{260/230}$ ratios. A A$_{260/280}$ ratio of between 1.8 and 2 is regarded as pure (Nanodrop, 2007). The A$_{260/230}$ ratio is a secondary measure of purity and a preferred value of between 2.0 and 2.2 generally indicates a pure sample (Nanodrop, 2007).

2.3 454 Amplicon Pyrosequencing Methods – Chapter 3

2.3.1 Metagenomic DNA Extraction for Amplicon Pyrosequencing

Gut mDNA extraction was performed using a modified version of the protocols previously described by Calderón-Cortés et al. (2010) and Shi et al. (2012). Gut sections were weighed and crushed in liquid nitrogen using a sterilised epi crusher. For 10 mg of gut tissue, 100 μl of a preheated (60°C) 2% Cetyl trimethylammonium bromide (CTAB) solution was added. The mixtures were incubated for 30 min at 60°C before centrifugation for 5 min at 10000 rpm. The supernatant was transferred to a clean collection tube and enzymatic digestion of the gut samples was carried out with the addition of 2 μl lysozyme (5 mg/ml) per 100 μl CTAB solution for 30 min at 37°C under continuous shaking (120 rpm). 0.5 μl Proteinase K (20 mg/ml) per 100 μl CTAB solution was then added (Priya et al., 2012), followed by an overnight incubation at 55°C with continuous shaking. One volume phenol:chloroform:isoamyl alcohol (25:24:1) solution was added. Tubes were inverted and centrifuged at 13000 rpm at 4°C for 4 min. One volume chloroform:isoamyl alcohol (24:1) solution was added to the top aqueous phase and the mixtures were inverted before centrifugation at 13000 rpm at 4°C for 15 min. This step was repeated until no protein contamination was observed (Priya et al., 2012). DNA was precipitated with 3 M NH$_4$Ac (Lagisz et al., 2010) and ice cold 99.9% EtOH followed by overnight incubation at -20°C. Mixtures were centrifuged for 60 min at 14000 rpm at 4°C. The DNA pellet was washed twice with ice cold 70% EtOH and allowed to dry completely for 2 hours. The DNA pellet was resuspended in 50 μl filter-sterilized nanopure H$_2$O overnight at 4°C (Lagisz et al., 2010), and stored at -20°C for downstream analysis.
2.3.2 454 Amplicon Pyrosequencing

2.3.2.1 454 Pyrosequencing of the Bacterial 16S Gene and Fungal ITS Region

The gut mDNA of five individuals from each *Pachysoma* spp. was sent to Molecular Research (MR DNA, Texas, United States of America; http://mrdnalab.com) for 16S rRNA gene and ITS region pyrosequencing using the Roche 454 GS FLX Titanium platform. The primers 27F (AGRGTTTGATCMTGGCTCAG; Lane (1991)) and 338R (AGTGCTGCTCCCGTAGGAGT; Fierer et al. (2008)) were used to amplify the 16S rRNA gene region. Fungal specific fITS9 (GAACGCAGCRAAIIGYGA; Ihrmark et al. (2012)) and ITS4 (TCCTCCGCTTATTGATATGC; Gardes and Bruns (1993)) primers were used for the amplification of the ITS region.

2.3.2.2 Data Analysis

Raw pyrosequencing reads were filtered and analysed using mothur version 1.35.1 (Accessed May 2015 - January 2016) (Schloss et al., 2009, Schloss et al., 2011). Fasta, quality and flow files were extracted from the sff files using the sff.info command. For the bacterial 16S rRNA gene pyrosequencing reads, filtering of poor quality reads was done using the shhh.flows command allowing for one or two mismatches between barcodes and primers, respectively. Remaining sequences were quality filtered with the trim.seqs command allowing for maximum homopolymers of 8bp and a minimum sequence length of 100bp. Sequences were aligned to the SILVA reference database (http://www.arbsilva.de/download/arb-files/) using the align.seqs command. The screen.seqs and filter.seqs commands were used to retain only overlapping sequences. Chimeras were identified and removed using the chimera.uchime command. Sequences were classified against five databases, namely the Ribosomal Database Project (RDP), SILVA, National Center for Biotechnology Information (NCBI), The Dictyoptera gut microbiota reference Database (DictDb; data shown) and GreenGenes with a confidence threshold of 80%. OTUs were clustered for each individual beetle before removal of singletons using the remove.rare command. Samples were
subsampled to 1718 reads (Table 3.1, Chapter 3); i.e., the lowest number of reads across all samples.

ITS sequence reads were analysed as described above, with minor differences outlined previously (Bell et al., 2014). Filtering of poor quality reads was done using the trim.seqs command, allowing for one or four mismatches between barcodes and primers, respectively. Sequences were trimmed to 200bp using the chop.seqs command to ensure all sequences were the same length. Sequences were classified against the UNITE database (version 6) with a confidence threshold of 50% and subsampled to the lowest number of OTUs across all samples (107; Table 3.1, Chapter 3) for statistical analyses.

Square-root transformation of the data was performed to reduce the effect of highly dominant species and increase the effect of less abundant species (Clarke et al., 2006). To compare the composition of bacterial and fungal communities within the guts of the two *Pachysoma* spp. studied, the Bray-Curtis coefficient (Bray and Curtis, 1957) was used to build a dissimilarity matrix from the pre-treated data. To visualise (dis)similarities between the two *Pachysoma* spp. gut communities, Non-Metric Multi-Dimensional Scaling (nMDS) plots were constructed using Primer 6 software (version 6.1.5.81; Primer E Ltd, Plymouth, UK) and created by calculating distances between communities (Clarke, 1993). Kruskal’s stress value was used to evaluate the ordination of sample placement in both two- and three-dimensional nMDS plots. A stress value greater than 0.2 suggests random placement, whereas a value lower than 0.1 suggests good correlation with the sample matrices (Clarke, 1993). Differences between the bacterial gut community structures of each *Pachysoma* spp. were tested for significance using one-way global Analysis of Similarities (ANOSIM) in Primer 6 software (version 6.1.5.81; Primer E Ltd, Plymouth, UK), using 10 000 permutations (Clarke, 1993). According to Clarke (1993) an R value closer to 1 indicates greater dissimilarity between datasets than within a dataset. Venn plots were created using R statistical package 2.15.1 in vegan (www.rproject.org), to determine the intra- and inter-specific distribution of bacterial and fungal OTUs. Diversity indices and rarefaction curves were generated in mothur using the
summary.single and rarefaction.single commands, respectively. Rarefaction curves were generated for each species in Microsoft Excel. Phylogenetic comparisons of both the bacterial and fungal dataset, were done using the relative abundance of all reads in the dataset so as to ensure inclusion of rare taxa. Relative abundances (%) were calculated from the number of reads of the specific microbial taxa divided by the total number of reads for the particular *Pachysoma* individual. Nucleotide sequences for both the bacterial and fungal datasets have been uploaded to NCBI (http://www.ncbi.nlm.nih.gov/) Short Read Archive (SRA) under the accession number SRP071915.

### 2.4 Hiseq Shotgun Sequencing Methods – Chapter 4

#### 2.4.1 Metagenomic DNA Extraction

DNA was extracted following the method described in Section 2.3.1 with modifications to reduce host DNA contamination (Liu et al., 2011). Briefly, each whole-gut (160-290 mg) was placed in a 1.5ml eppendorf tube with 1ml Phosphate-buffered saline (PBS) solution. Samples were crushed using a sterilized epi-crusher. Samples were vortexed for 15s followed by three rounds of low-speed centrifugation (2900rpm at 4°C for 10 min) to remove eukaryotic DNA before extraction (Liu et al., 2011). After each round, the supernatant was transferred to a new 15ml tube and the pellet resuspended in 1ml PBS. The supernatants were then pooled and centrifuged at 2400rpm at 4°C for 10min. The supernatant was removed into a new tube. This step was repeated 4 times and the final supernatant was centrifuged at 9000xg at 4°C for 15 min. The pellet was resuspended in 1ml PBS and 100μl of a 2% preheated (60°C) CTAB solution was added for every 10mg of gut sample. DNA extraction then followed the method described in Section 2.3.1

#### 2.4.2 PCR Amplification of 16S rRNA and 18S rRNA Genes

The 18S rRNA gene was amplified using the insect specific primers 18S intfw-st12 (ATCAAGAACCAGGTTAGAG; Haring and Aspöck (2004)) and 18S rev1
(ATGGGGAACACATTGCAAGC; Haring and Aspöck (2004)) following a previous protocol (Sole et al., 2013) with minor modifications. Each PCR mixture (50μl) contained 0.2mM dNTPs, 0.5mM each primer, 0.01U DreamTaq (Fermentas, Lithuania), 1X Dream Taq Buffer and 1μl DNA template. The thermal cycle included an initial denaturation step at 95°C for 2min, followed by 30 cycles of 10s denaturation at 95°C, 20s annealing at 50°C and 1min extension at 72°C, with a final extension step at 72°C for 5min. The bacteria specific 16S rRNA gene was amplified using the primers 27F (AGAGTTTGATCCTGGCTCAG; Lane (1991)) and 1492R (TACGGYTACCTTGTTACGACTT; Lane (1991)). Each PCR mixture (final volume of 50μl) contained 0.2mM dNTPs, 0.4mg/ml BSA, 0.5mM of each primer, 0.01U DreamTaq, 1X Dream Taq Buffer and 1μl DNA template. The thermal cycle included an initial denaturation step at 94°C for 3min, followed by 25 cycles of denaturation at 94°C for 30s, annealing at 55°C for 1.30min and extension at 72°C for 2.30min with a final extension step at 72°C for 5min. PCR amplicons were visualized on 1% agarose gels.

2.4.3 Metagenome Sequencing and Data Analysis

Metagenomic DNA was extracted from the gut samples of three individuals per Pachysoma species, and pooled prior to sequencing using the Illumina Hiseq 2500 platform (Molecular Research (MR DNA), Texas, United States of America; http://mrdnalab.com). Reads were quality trimmed and assembled into contigs using the CLC Genomic Workbench (version 8.5.1; CLC Bio). Reads were trimmed by 15bp on the 5’ end and 1bp on the 3’ end allowing for no ambiguous bases and a minimum length of 75bp. For the assembly, a minimum contig length of 200bp, with a mismatch cost of 2bp and insertion and deletion costs of 3bp was accepted. Length and similarity fractions of 0.5 and 0.8, respectively, were accepted. High quality assemblies (including unmapped reads) were uploaded to the Metagenomics RAST server (MG-RAST)(Meyer et al., 2008) for taxonomic annotation. The assemblies were also annotated with Basic Local Alignment Search Tool (BLASTX) analysis using Diamond (Buchfink et al., 2015) against the non-redundant (NR) protein database with sensitive parameters (e-value 1e-3) and uploaded to the MEtaGenome ANalyzer (MEGAN)(Huson et
al., 2007, Huson et al., 2016). Contigs (or unmapped reads) predicted to originate from Metazoa (animals) and Streptophyta (plants) were extracted and removed from the original datasets, to ensure that only microbial open reading frames (ORFs)/contigs were analysed further. Prodigal was used for protein translation followed by a BLASTX and BLASTP analysis using Diamond with sensitive parameters and an e-value of 1e-3. The microbial datasets were analysed in MEGAN for both taxonomic classification and gene prediction using SEED and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Carbohydrate-Active enZYme (CAZy)(Cantarel et al., 2009) family classifications were based on Pfam (Bateman et al., 2004) domain assignments of the microbial datasets. KEGG E.C assignments were allocated using the KEGG Automatic Annotation System (KAAS)(Moriya et al., 2007).

Sequences for both datasets have been uploaded to NCBI (http://www.ncbi.nlm.nih.gov/) Short Read Archive (SRA) under the accession number SRP071915.

2.5 References


Chapter 3: Amplicon Sequencing of Bacterial and Fungal Gut Communities of *Pachysoma* Species

### 3.1 Introduction

The microbial gut communities of a wide range of insect species have been investigated (for reviews see Dillon and Dillon (2004), Douglas (2015), Broderick and Lemaitre (2012), Brune (2013), Huang et al. (2010), Engel and Moran (2013)). The gut environment is considered to be an unstable system, as microorganisms face secretion of digestive enzymes, physical disturbance, habitat shedding during insect moults and other physiochemical conditions that are typically unfavourable for colonisation (Dillon and Dillon, 2004, Engel and Moran, 2013, Douglas, 2015). However, there are significant benefits to gut colonisation, including high nutrient availability and protection from external environmental stressors (Hooper et al., 2002, Douglas, 2015).

The relationships between host and gut microbiota range across the full spectrum of interactions; i.e., from pathogenic to obligate mutualism (Dillon and Dillon, 2004). When beneficial to their host, insect-associated microbial communities may participate in a number of activities including degradation of recalcitrant materials such as lignocellulose (Park et al., 2007, Douglas, 2009, Calderón-Cortés et al., 2012, Douglas, 2013, He et al., 2013), the production of nutrients and vitamins (Douglas, 2009, Douglas, 2013, 2015), the production of components of cohesion pheromones (Dillon et al., 2002), nitrogen fixation and utilisation of nitrogenous waste products (Nardi et al., 2002, Douglas, 2009, Douglas, 2013, 2015), protection against parasites (Koch and Schmid-Hempel, 2011, Douglas, 2015), change in body colouration (Tsuchida et al., 2010) and sterol synthesis (Douglas, 2009, Douglas, 2013).

Insect gut microbiomes are known to differ between insect species, driven by variations in the gut structure, different host lifecycles, host phylogeny and diet (Colman et al., 2012, Engel and Moran, 2013, Douglas, 2015). The gut microbiome is also influenced within the
individual insect or species, varying according to host life-stage (Vasanthakumar et al., 2008, Wang et al., 2011, Arias-Cordero et al., 2012, Andongma et al., 2015, Diouf et al., 2015), and/or diet (Miyata et al., 2007, Wang et al., 2011, Bertino-Grimaldi et al., 2013, Huang et al., 2013, Montagna et al., 2015b). Host diet influences gut microbial communities as they adapt to dietary changes through the induction of enzymes and changes in community structure (Kaufman and Klug, 1991, Santo Domingo et al., 1998, Bertino-Grimaldi et al., 2013). However, a core community may persist through major dietary changes (Huang et al., 2013, Schauer et al., 2014).

Studies on insect-microbial associations have mainly focused on termites (Brune and Friedrich, 2000, Ohkuma, 2003, Warnecke et al., 2007, Husseneder, 2010, Brune, 2013, Poulsen, 2015), but also on agriculturally important species such as honeybees (Hamdi et al., 2011, Moran, 2015), and medically important insects such as mosquitoes (Wang et al., 2011, Boissière et al., 2012, Osei-Poku et al., 2012, Wang et al., 2012, Coon et al., 2014). Little attention has been given to dung beetles, which are common and abundant insects in virtually all terrestrial environments and which facilitate nutrient cycling and bioturbation (Nichols et al., 2008). The desert dung beetle genus *Pachysoma*, from the Scarabaeini tribe, of which the quintessential scarab genus, *Scarabaeus* is also a member, consists of 13 species endemic to the south-west African coast (Harrison et al., 2003, Sole et al., 2005). Members of *Pachysoma* exhibit atypical feeding behaviour. While most adult dung beetles feed, by filtration, on minute particulate fragments in wet dung (Holter and Scholtz, 2011, 2013), adult *Pachysoma* feed on various and varying dry food-sources: plant detritus, dung pellets or both. These substrates are collected on the soil surface and masticated with specially-adapted mouthparts (Figure 3.1) (Scholtz, 1989, Harrison et al., 2003, Sole et al., 2005, Holter and Scholtz, 2011).

Given that insect gut microorganisms are known to be involved in the degradation of recalcitrant materials such as lignocellulosic compounds (Douglas, 2009, Calderón-Cortés et al., 2012, Engel and Moran, 2013, Douglas, 2015), it follows that the gut microbiomes of desert
insects may play a significant role in carbon-turnover in desert ecosystems. By studying the gut microbiome diversity of *Pachysoma* spp. feeding on different plentiful and readily-available substrates, it is possible to consider the effects of host diet and/or host phylogeny on gut microbiome assembly processes. This study was designed to characterise the gut microbial (bacterial and fungal) assemblages of coprophagous (*P. striatum*) (Harrison et al., 2003) and detritivorous (*P. endroedyi*; C. Scholtz Pers. Comm.) members of the same genus from the same location and to potentially determine whether host diet and/or host phylogeny could be deterministic factors in *Pachysoma* gut microbial community assembly.

### 3.2 Results and Discussion

#### 3.2.1 The Desert Beetle Genus *Pachysoma*

The distribution of the *Pachysoma* species is restricted to the arid coastal regions of south-western Africa, principally because of the flightless nature of the genus (Harrison et al., 2003). The genus *Pachysoma* forms three distinct lineages, supporting six (lineage 1), four (lineage 2) and three (lineage 3) species, respectively (Sole et al., 2005). *Pachysoma endroedyi* is located in lineage 1 and *P. striatum* in lineage 2 (Figure 3.1). The driving forces behind the formation of these three lineages are currently unknown. However, it has been noted that all members of lineage 3 have a uniform diet (Figure 3.1) and originate from desert areas with a consistent aridity index (Harrison et al., 2003, Sole et al., 2005), whereas both the aridity index of the desert locations from which lineage 1 and 2 members originate, and their diets, fluctuate (Figure 3.1).
Figure 3.1: Cytochrome oxidase I gene Parsimony tree phylogeny of 13 *Pachysoma* spp. Branch colours indicate the diet of the *Pachysoma* spp.: dung (brown), plant detritus (green), polyphagous (blue) and unknown (black). This phylogenetic tree was adapted, with permission, from (Sole et al., 2005) and the dietary information taken from both (Harrison et al., 2003) and personal observations by Prof C. Scholtz. Two species of *Scarabaeus*, (*S. proboscideus* and *S. rugosus*) which is the sister-genus to *Pachysoma* and a typical wet-dung-feeder, were used as outgroups. Numbers to the right of the tree indicate the three *Pachysoma* lineages. The two species considered in this study, *P. endroedyi* and *P. striatum*, are indicated with stars (Adapted from Sole et al. (2005)).
The diet of *P. striatum* consists predominantly of the dry dung pellets (Scholtz, 1989, Harrison et al., 2003) of various small native mammalian herbivores and sheep. Despite observations from a decade ago stating that *P. endroedyi* was a polyphagous feeder (Harrison et al., 2003), numerous and wide-scale recent observations suggest that *P. endroedyi* is a detrivore (Prof C. Scholtz, pers. comm.), the classification adopted in this study. *Pachysoma* species have specialised anatomical and physiological features for mastication and digestion of fragments from plant detritus and dry dung (Holter and Scholtz; 2011, 2013).

The linkage between host and gut microbiome is believed to be bidirectional, in that gut microorganisms can provide nutritional assistance to the insect host (Douglas, 2011, 2013) while the host diet influences the gut microbiome assembly (Miyata et al., 2007, Wang et al., 2011, Bertino-Grimaldi et al., 2013, Huang et al., 2013, Montagna et al., 2015b). However, host phylogeny may also impact gut microbiome composition (Colman et al., 2012, Yun et al., 2014), irrespective of the diet.

### 3.2.2 Sequencing Outputs and Diversity Indices of the Bacterial 16S rRNA Gene and Fungal ITS Region of the *Pachysoma* Gut Microbiome

The gut microbiomes of five detritivorous *P. endroedyi* and five coprophagous *P. striatum* individuals were determined by 16S rRNA gene amplicon sequencing. After removal of chimeras and singletons, 39050 bacterial and 1492 fungal reads remained, with mean read lengths of 238bp and 100bp, respectively. Only 462 bacterial reads were obtained for *P. endroedyi* individual 3 (Table 3.1), which was therefore removed from further analysis. Considerable variation in the number of bacterial sequence reads was noted between individuals, ranging from 1718 to 2817 and 3911 to 10106 for *P. endroedyi* and *P. striatum*, respectively. However, Good’s coverage (>0.97 for all samples), rarefaction and chao1 diversity indices suggested that the coverage of *Pachysoma* bacterial gut communities (Figures 3.2a and b) were sufficient for a valid comparison between individuals. The fungal ITS region could not be amplified in samples from the detritivorous species *P. endroedyi,*
despite repeated attempts. The absence of fungi in the insect gut has previously been noted for individuals of various insect groups including Neuroptera and Coleoptera (using culture-dependent techniques) (Nguyen et al., 2007). In the fungal ITS sequence datasets for *P. striatum*, diversity indices and rarefaction curves showed low coverage for all but *P. striatum* individual 2, suggesting that the fungal diversity was generally underestimated (Table 3.1; Figure 3.2c).

**Table 3.1:** Values for sequence reads, OTUs, phyla and diversity indices for bacterial and fungal gut communities of *P. endroedyi* and *P. striatum* individuals.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Number of reads</th>
<th>Number of OTUs</th>
<th>Phyla</th>
<th>Singletons</th>
<th>Chao</th>
<th>Invsimpson</th>
<th>Shannon</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial 16S rRNA gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. endroedyi</em> 1</td>
<td>2120</td>
<td>213</td>
<td>10</td>
<td>105</td>
<td>244.61</td>
<td>19.01</td>
<td>3.88</td>
<td>0.97</td>
</tr>
<tr>
<td><em>P. endroedyi</em> 2</td>
<td>1718</td>
<td>258</td>
<td>11</td>
<td>133</td>
<td>282.34</td>
<td>81.00</td>
<td>4.91</td>
<td>0.97</td>
</tr>
<tr>
<td><em>P. endroedyi</em> 3</td>
<td>462</td>
<td>97</td>
<td>11</td>
<td>42</td>
<td>112.62</td>
<td>21.47</td>
<td>3.82</td>
<td>0.94</td>
</tr>
<tr>
<td><em>P. endroedyi</em> 4</td>
<td>2175</td>
<td>271</td>
<td>6</td>
<td>177</td>
<td>287.59</td>
<td>62.44</td>
<td>4.80</td>
<td>0.98</td>
</tr>
<tr>
<td><em>P. endroedyi</em> 5</td>
<td>2817</td>
<td>317</td>
<td>6</td>
<td>193</td>
<td>335.83</td>
<td>60.13</td>
<td>4.83</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Fungal ITS region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. striatum</em> 1</td>
<td>10106</td>
<td>157</td>
<td>4</td>
<td>84</td>
<td>174.53</td>
<td>8.77</td>
<td>2.82</td>
<td>1.00</td>
</tr>
<tr>
<td><em>P. striatum</em> 2</td>
<td>4901</td>
<td>158</td>
<td>4</td>
<td>87</td>
<td>194.96</td>
<td>16.64</td>
<td>3.38</td>
<td>0.99</td>
</tr>
<tr>
<td><em>P. striatum</em> 3</td>
<td>4208</td>
<td>140</td>
<td>6</td>
<td>51</td>
<td>183.05</td>
<td>8.63</td>
<td>2.88</td>
<td>0.99</td>
</tr>
<tr>
<td><em>P. striatum</em> 4</td>
<td>3911</td>
<td>119</td>
<td>4</td>
<td>71</td>
<td>125.84</td>
<td>8.12</td>
<td>2.93</td>
<td>1.00</td>
</tr>
<tr>
<td><em>P. striatum</em> 5</td>
<td>6620</td>
<td>172</td>
<td>5</td>
<td>100</td>
<td>201.29</td>
<td>13.36</td>
<td>3.32</td>
<td>0.99</td>
</tr>
</tbody>
</table>
**Figure 3.2:** Rarefactions curves showing gut microbial community richness of all *Pachysoma* individuals for bacterial 16S rRNA gene amplicon data of: a) *P. endroedyi*, b) *P. striatum*; and c) fungal ITS region amplicon data of *P. striatum*.
A total of 1009 bacterial and 294 fungal OTUs were detected at an identity threshold of 97% (Table 3.1). Numbers ranged from 213 to 317 and 119 to 172 in the *P. endroedyi* and *P. striatum* gut samples, respectively (Table 3.1). These values are comparable with results obtained for termite and cockroach gut microbiomes (Sabree and Moran, 2014). It should be noted that the fungal ITS sequence read lengths were short (only 100bp), which could explain the poor phylogenetic resolution of *P. striatum* fungal gut communities (Lim et al., 2010).

In both *Pachysoma* spp., the number of bacterial 16S rRNA sequence reads was inversely proportional to the number of bacterial OTUs; i.e., *P. striatum* gut samples had a higher average number of bacterial reads (5949 ± 2550) but a lower average number of bacterial OTUs (149 ± 20) when compared to *P. endroedyi* (2208 ± 455 reads and 265 ± 43 OTUs, respectively). Those data suggest that the gut bacterial communities of *P. striatum* are composed of a relatively low number of dominant phylotypes at high abundance (Ahn et al., 2012, Gauthier et al., 2015, Montagna et al., 2015b). Contrastingly, the *P. endroedyi* gut bacterial community may include a higher bacterial diversity (Boucias et al., 2013, Montagna et al., 2015b). This inverse relationship, and the higher Shannon diversity index of the *P. endroedyi* gut bacterial community (4.6 ± 0.5) compared with the *P. striatum* gut community (3.1 ± 0.3; Table 3.1), suggests that competition is greater in the *P. striatum* gut than in *P. endroedyi*. This difference may be a reflection of the different diets, as insects feeding on simple diets (e.g., the coprophagous diet of *P. striatum*) commonly have a lower gut bacterial diversity than those feeding on more complex diets (e.g., the detritivorous diet of *P. endroedyi*) (Colman et al., 2012, Yun et al., 2014).

### 3.2.3 Inter-specific Variations in Bacterial and Fungal *Pachysoma* Gut Communities

The gut bacterial communities of *P. endroedyi* and *P. striatum* were significantly different, sharing only 3.7% of bacterial OTUs (Figure 3.3; ANOSIM [R=1.00, p<0.008]). Both host phylogeny and host diet could be driving forces for the observed differences (Chandler
et al., 2011, Colman et al., 2012). For example, the Hymenopteran gut microbiome has previously been shown to be influenced by host phylogeny, while the gut microbiomes of detritivorous insects (e.g., certain termites, Coleoptera and Diptera) are dictated by diet (Colman et al., 2012). Gut bacterial communities of Drosophila spp. also appear to be impacted by host diet rather than host phylogeny (Chandler et al., 2011). In Coleoptera (the order in which Pachysoma is placed), gut bacterial communities are significantly different to those of other insect groups (Colman et al., 2012), indicating that host phylogeny is a significant driving force for gut microbiome assembly. However, within Coleoptera, significant similarities in bacterial assemblages of certain beetles with similar diets (e.g., those feeding on live arboreal tissue) have also been noted (Colman et al., 2012), which suggests that diet may also be a deterministic factor. It should, however, be noted that no coprophagous insects were included in this study (Colman et al., 2012), making a direct comparison with Pachysoma speculative.

Figure 3.3: nMDS ordination plot based on Bray-Curtis distance matrices of bacterial 16S rRNA gene pyrosequencing data for P. endroedyi and P. striatum individuals. A stress value of less than 0.1 represents a high quality ordination. Pachysoma endroedyi and P. striatum are represented by green and inverted brown triangles, respectively.
It is not possible to compare the gut fungal communities of the two insect species studied, given that despite numerous attempts it was not possible to PCR-amplify fungal ITS sequences from the detritivorous *P. endroedyi*. While it is unlikely that fungal species are completely absent from the gut microbiome of this species, this negative result suggests that they may represent a relatively minor fraction of the total gut microbial diversity. To fully confirm this, the sample size should be increased and *P. endroedyi* individuals from multiple breeding populations should be investigated.

It is expected that host diet would be a contributing factor in the presence (or absence) of fungi in the *Pachysoma* gut. For example, true yeasts (Saccharomycetaceae) are typically observed in the guts of litter-, plant- and wood-feeding insects (Suh et al., 2008, Scully et al., 2013, Santana et al., 2015), but not in those of predacious insects (Nguyen et al., 2007, Shao et al., 2015).

### 3.2.4 Intra-specific Variation of *Pachysoma* Gut Microbial Communities

Large intra-specific differences in *Pachysoma* gut communities were noted, with the majority of OTUs being unique to each *Pachysoma* individual (Figures 3.4a and b, Figure 3.5a) and only 11 (1.1%) and 17 (3.3%) bacterial OTUs being shared between individuals of *P. endroedyi* and *P. striatum*, respectively. Furthermore, only two non-abundant fungal OTUs (ranging from 1.6-1.7% of the community) were shared among the five *P. striatum* individuals (Figure 3.5a). Such intra-specific differences, relating to the relative abundances and diversity of bacterial members of gut communities, are not uncommon, as has been observed for honeybees (*Apis cerana* and *A. mellifera* (Ahn et al., 2012)), mosquitoes (*Aedes* spp., *Culex* spp., *Anopheles* spp., *Mansonia* spp.; (Boissière et al., 2012, Osei-Poku et al., 2012)) and the red palm weevils *Rhynchophorus ferrugineus* and *R. vulneratus* (Montagna et al., 2015b), among others. A recent study on the gut microbiomes of 218 different insect species from 21 orders (Yun et al., 2014) indicated that 46% of the total number of bacterial OTUs detected (n=9301) were only observed in single individuals. The large intra-specific variation noted in...
*Pachysoma* could be influenced by the stochastic, and transient, process of microorganisms entering the gut with the food-source (Douglas, 2015) and, for *P. striatum*, the different amounts of feeding material contained in the guts of each individual (Dillon and Dillon, 2004). Furthermore, it cannot be excluded that the ‘time of feeding’ prior to sampling may also have had an influence on intra-specific gut microbiome variability (Dillon and Dillon, 2004).

**Figure 3.4:** Venn diagrams showing distribution of bacterial OTUs between (a) *P. endroedyi* and (b) *P. striatum* individuals based on the 16S rRNA gene pyrosequencing analysis. Shared OTUs are shown in bold. Numerical labels are shown for each individual.
Figure 3.5: a) Venn diagram comparing the distribution of fungal OTUs between *P. striatum* individuals based on the ITS region pyrosequencing analysis. Shared OTUs are shown in bold. Numerical labels are given for each individual; b) Relative abundance of fungal phyla in five *P. striatum* individuals based on ITS rRNA gene region pyrosequencing analysis at a 97% identity threshold.
Of the shared bacterial OTUs, only one (assigned to the phylum Bacteroidetes) and eight (4 Firmicutes, 2 Actinobacteria, 1 Bacteroidetes and 1 Proteobacteria) were abundant (i.e., represented >2% reads) in the *P. endroedyi* and *P. striatum* gut samples, respectively. This distribution is strongly suggestive that the *Pachysoma* gut core community is very small, as has been proposed for the “minimal core” model (Hamady and Knight, 2009). Other studies have noted the presence of consistent core microbial communities within individuals of the same insect species (e.g., the bed bug *Cimex lectularius*; (Meriweather et al., 2013) and bumble bee *Bombus terrestris* (Billiet et al., 2015, Meeus et al., 2015)), or across taxonomic levels (e.g., across the ant tribe Cephalotini; (Anderson et al., 2012)). In the termite *Reticulitermes flavipes*, a substantial core bacterial microbiome (65% shared OTUs) was noted, regardless of the artificial feeding diet, suggesting that host phylogeny may play a more important role than host diet in the assembly of the gut microbiome (Huang et al., 2013). Similar results have been noted in cockroaches (Schauer et al., 2014). However, with a minimal core microbiome in both *Pachysoma* spp., phylogeny appears less important than diet. Furthermore, a minimal core gut microbiome may result from negative interactions between gut microorganisms, such as antagonism or amensalism, or indicate, as for *Drosophila* (Wong et al., 2013), the establishment of ‘non-gut-specific’ microorganisms.

It has been suggested that a ‘functional’ rather than a ‘phylogenetic’ core microbiome may be more informative in determining the assembly of gut microbiomes (Karasov et al., 2011). In studies on humans, which typically follow the minimal core model, functional gene diversity appears to be broadly similar across individuals (Turnbaugh et al., 2009, Karasov et al., 2011). Therefore, there may be a functional core community in each *Pachysoma* sp. studied, displaying shared metabolic capacities (Turnbaugh et al., 2009); i.e., exhibiting functional redundancy. As such, it has been suggested that a comparison of functional properties of hosts feeding on different diets can guide an understanding of the functional roles of different gut microbiomes (Karasov et al., 2011).
3.2.5 Phylogenetic Diversity of Bacterial and Fungal Pachysoma Gut Communities

The gut bacterial diversity of *P. endroedyi* was higher (6-11 phyla; Figure 3.6) than that of *P. striatum* (4-7 phyla; Figure 3.6). The *P. endroedyi* gut samples were dominated by Bacteroidetes (18.0-54.8%), Firmicutes (10.0-34.6%), Proteobacteria (8.7-18.1%) and Planctomycetes (2.5-25.7%), while Actinobacteria (0.1-22.5%), Elusimicrobia (0-9.3%) and Synergistetes (0-7.3%) showed highly variable abundances (Figure 3.6). The remaining 7 phyla each represented less than 2% of the community and were often detected in single insects. In *P. striatum*, Bacteroidetes (3.0-57.1%), Firmicutes (18.9-56.2%), Proteobacteria (6.4-32.1%) and Actinobacteria (5.2-21.0%) were also dominant phyla although the relative abundances varied between individuals (Figure 3.6). Three minor phyla (<2% abundance) were only detected in two *P. striatum* individuals, namely Deferribacteres, Planctomycetes and Synergistetes (Figure 3.6). All the identified bacterial phyla have previously been reported in insect microbiomes, with the phyla Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria commonly abundant in insect gut samples (Jones et al., 2013, Yun et al., 2014).
Figure 3.6: Comparison of inter-specific differences in relative abundance of bacterial phyla in the gut of two *Pachysoma* spp., *P. endroedyi* and *P. striatum*, based on 16S rRNA gene pyrosequencing analysis.

The presence of specific bacterial phyla and/or their relative abundances in insect gut samples may be linked to host diet. For example, certain insects with simpler diets (e.g., feeding on pollen and nectar (Ahn et al., 2012), fruit (Wang et al., 2014, Andongma et al., 2015), or sap (Pandey and Rajagopal, 2016)), contain gut bacterial communities which are typically dominated by heterotrophic Proteobacteria and/or Firmicutes. In contrast, Bacteroidetes (along with other phyla) were highly abundant in the gut microbiomes of insects feeding on plant materials such as wood and leaves (Köhler et al., 2012, Boucias et al., 2013, Schauer et al., 2014, Montagna et al., 2015b, Waite et al., 2015). The *P. striatum* gut bacterial communities did not display these patterns, suggesting that coprophagous diets may structure insect gut communities differently.

Fifteen and 11 bacterial genera were abundant (>2% relative abundance of reads) within the guts of *P. endroedyi* and *P. striatum*, respectively (Table 3.2). Only two of these genera
were abundant in both species (*Dysgonomonas* and unclassified Enterobacteriaceae; Table 3.2). *Dysgonomonas* was less abundant in *P. endroedyi* gut samples (2.8% ± 0) than in *P. striatum* (26.3% ± 0.2), in which it was the most abundant genus. *Dysgonomonas* have been reported to be present at high abundance in the gut system of the fungus-growing termite (*Macrotermes annandalei*) and red palm weevil larvae (*Rhynchophorus ferrugineus*) (Tagliavia et al., 2014, Zhang et al., 2014). Two species of *Dysgonomonas* have previously been characterised from the gut of termites (Yang et al., 2014, Pramono et al., 2015). Both species have been found to ferment glucose and xylan as a sole carbon source and to produce acetic acid as the major end-product (Yang et al., 2014, Pramono et al., 2015), suggesting roles in both the lignocellulosic biomass degradation pathway and in providing readily metabolisable substrates for ingestion by the host. The large difference in the abundance of this phylotype in the two *Pachysoma* species suggests a key nutritional role in *P. striatum* but not in *P. endroedyi.*
Table 3.2: Phylogenetic classification of the most abundant bacterial genera in the gut samples of *P. endroedyi* and *P. striatum*: i.e., representing >2% reads.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Family</th>
<th>Genus</th>
<th><em>P. endroedyi (%)</em></th>
<th><em>P. striatum (%)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>Propionibacteriaceae</td>
<td>Propioniobacterium 1</td>
<td>8.2 ±0.1</td>
<td>0.1 ± 0</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>Tessaracoccus</td>
<td>0</td>
<td>7.6 ± 0</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Bacteroides</td>
<td>Uncultured 1</td>
<td>0.1 ± 0</td>
<td>5.2 ± 0</td>
</tr>
<tr>
<td>Porphyromonadaceae</td>
<td>1</td>
<td>Dysgonomonas</td>
<td>2.8 ± 0</td>
<td>26.3 ± 0.2</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Porphyromonadaceae 4</td>
<td>Proteiniphilum</td>
<td>0.2 ± 0</td>
<td>3.5 ± 0</td>
</tr>
<tr>
<td></td>
<td>Rikenellaceae</td>
<td>Alisites IV</td>
<td>6.5 ± 0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>unclassified</td>
<td>3.2 ± 0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>unclassified</td>
<td>5.3 ± 0</td>
<td>0.1 ± 0</td>
</tr>
<tr>
<td>Elusimicrobia</td>
<td>Endomicrobiaceae</td>
<td>Endomicrobium</td>
<td>3.0 ± 0</td>
<td>0</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Enterococcaceae</td>
<td>Vagococcus</td>
<td>0.3 ± 0</td>
<td>4.9 ± 0</td>
</tr>
<tr>
<td></td>
<td>Family XI Incertae Sedis</td>
<td>unclassified</td>
<td>0</td>
<td>10.9 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>Lachnospiraceae</td>
<td>Uncultured 13</td>
<td>0.9 ± 0</td>
<td>2.9 ± 0</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>unclassified</td>
<td>3.3 ± 0</td>
<td>0.7 ± 0</td>
</tr>
<tr>
<td></td>
<td>Ruminococcaceae</td>
<td>Termite cockroach cluster</td>
<td>5.3 ± 0.1</td>
<td>0.1 ± 0</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>unclassified</td>
<td>3.3 ± 0</td>
<td>1.4 ± 0</td>
</tr>
<tr>
<td></td>
<td>Veillonellaceae</td>
<td>Anaeroarcus-Anaeromusa</td>
<td>0</td>
<td>11.5 ± 0.1</td>
</tr>
<tr>
<td>Planctomycetes</td>
<td>unclassified</td>
<td>unclassified</td>
<td>11.3 ± 0.1</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>unclassified</td>
<td>unclassified</td>
<td>0</td>
<td>3.1 ± 0</td>
</tr>
<tr>
<td></td>
<td>Insect cluster</td>
<td>unclassified</td>
<td>2.2 ± 0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Enterobacteriaceae</td>
<td>unclassified</td>
<td>4.2 ± 0.1</td>
<td>7.4 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>Enterobacteriaceae 1</td>
<td>unclassified</td>
<td>0</td>
<td>2.5 ± 0.1</td>
</tr>
<tr>
<td>Synergistetes</td>
<td>Synergistaceae</td>
<td>Candidatus Tammella</td>
<td>2.8 ± 0</td>
<td>0</td>
</tr>
</tbody>
</table>

Percentages are the average read relative abundances in each species (*P. endroedyi*: n=4; *P. striatum*: n=5). Colours depict the species in which the bacterial genus is abundant: *P. endroedyi* (green), *P. striatum* (brown) or both (blue). The most abundant genus of each species is shown in bold.

An unclassified Planctomycetes dominated the gut samples of *P. endroedyi* (11.3% ± 0.1 relative abundance of reads). To the best of our knowledge this is the first report of an insect gut microbiome dominated by Planctomycetes. Planctomycetes were only detected in a single *P. striatum* individual at very low abundance (0.01%). Planctomycetes have previously
been detected in the guts of the termites *Syntermes wheeleri* and *Nasutitermes* spp. (Köhler et al., 2012, Santana et al., 2015), the cockroach *Shellfordella lateralis* (Schauer et al., 2014), adult and larval beetles (*Cryptocephalus* spp., *Prionoplus reticularis* and *Pachnodas* spp. (Andert et al., 2010, Reid et al., 2011, Montagna et al., 2015a)), the tree weta *Hemideina thoracica* (Waite et al., 2015) and the mosquito *Aedes albopictus* (Zouache et al., 2012), but only in low abundances (<1-5% relative abundance).

Ascomycota was the most abundant fungal phylum (42.3-75.7%) in all *P. striatum* gut samples, which is typical for insect gut microbiomes (Suh et al., 2004, Suh et al., 2005, Scully et al., 2013, Santana et al., 2015). Basidiomycota were not ubiquitously detected, and were observed only in the gut samples of four of the five *P. striatum* individuals (1.8-15.2%; Figure 3.5b). A substantial proportion of fungal ITS sequence reads could not be classified, even at the phylum level (9.1-55.9%; Figure 3.5b). Unfortunately, relatively little is known about insect gut fungal diversity (compared to bacterial diversity (Gibson and Hunter, 2010)), with the majority of published studies being based on culture-dependent methods which are typically biased when compared with culture-independent methods (Zhang et al., 2003, Gibson and Hunter, 2010).

### 3.3 Conclusion

This is the first study to investigate the gut microbiomes of any dung beetle feeding on dry food-sources and to compare those of closely related adult dung beetle species with very different diets but from the same locality. *Pachysoma* spp. are ecologically important in arid environments where they undoubtedly participate in nutrient cycling and bioturbation (Nichols et al., 2008). We have demonstrated that, as predicted, the gut microbiomes differed significantly between two species which feed on different substrates. However, both populations showed large intra-specific variations. Thus, to further characterise the gut microbiomes of these *Pachysoma* species, the number of individuals studied should be increased and populations from different sites investigated. Such experiments would make it
possible to evaluate whether inter-specific variation was higher than intra-specific variation within a single *Pachysoma* species.

We are unable to fully assess whether host phylogeny or the host diet is the dominant driver of the *Pachysoma* gut microbiomes. Nevertheless, we provide evidence that diet probably plays a significant role, particularly noting the fact that the gut microbiomes of the detritivorous *P. endroedyi* (feeding on complex food-sources) have higher bacterial diversities than those of the coprophagous species (feeding on relatively simple food-sources) (Colman et al., 2012, Yun et al., 2014). Functional gene analysis of the microbiomes of *P. endroedyi* and *P. striatum* could potentially assist in confirming the role that host diet plays in *Pachysoma* gut microbiome assembly (He et al., 2013).

### 3.4 References


© University of Pretoria


Chapter 4: Shotgun Metagenomic Sequencing of the Gut Microbiomes of *Pachysoma endroedyi* and *Pachysoma striatum*

4.1 Introduction

A variety of methods have been used to study the insect gut microbiome including culturing, clone libraries, amplicon sequencing and most recently shotgun sequencing of the entire gut metagenome (Suh et al., 2004a, Ahn et al., 2012, Bertino-Grimaldi et al., 2013, He et al., 2013, Liu et al., 2013, Scully et al., 2013, Santana et al., 2015). Until recently, newer sequencing technologies were priced outside the range acceptable for most studies. However, as costs decrease, shotgun sequencing is becoming a favourable method for sequencing whole microbial communities in a variety of habitats (Venter et al., 2004, Adriaenssens et al., 2015, Mendes et al., 2015, Adriaenssens et al., 2016, Vikram et al., 2016). Shotgun sequencing has also been utilised for studying the gut microbiomes of well-studied taxonomic groups of insects including honeybees (Engel et al., 2012), mosquitoes (Chandler et al., 2015), termites (He et al., 2013), ticks (Carpi et al., 2011) and wood-feeding beetles (Jia et al., 2013, Scully et al., 2013).

Shotgun metagenomic sequencing allows access to the entire gut microbiome i.e. both microbial diversity and metabolic potential (Sankar et al., 2015). As a result, it is possible to study the set of functional genes and pathways that can assist a host in biological processes, with enzymes involved in digestion of recalcitrant carbohydrates (e.g., lignocellulosic compounds) important in industry and ecology alike (He et al., 2013, Scully et al., 2013)). However, a disadvantage of shotgun sequencing the insect gut microbiome is the large majority of resulting sequences being of host origin (Carpi et al., 2011, Hunter et al., 2011, Feehery et al., 2013). To overcome this challenge, the majority of studies use targeted
amplicon sequencing. Even studies which do use shotgun sequencing typically only report functional data, using amplicon sequencing for taxonomic analysis (He et al., 2013, Scully et al., 2013). However, shotgun sequencing overcomes primer bias introduced in amplicon sequencing and offers description of the entire gut microbial composition not just particular groups (e.g. bacteria or fungi) (Hodkinson and Grice, 2015).

This is the first study investigating the gut microbiome, and associated metabolic potentials, of a scarab beetle using this high throughput technology. In this chapter, shotgun sequencing was used to determine the taxonomic composition and functional capacity of the *Pachysoma* gut microbiome of two species feeding on substrates with different chemical and material compositions. Therefore, this chapter aims to answer the following main research questions:

1. What is the taxonomic composition of the *Pachysoma* gut metagenome and how does microbial community structure differ between the two *Pachysoma* spp. in comparison with the amplicon study.
2. Are there notable differences in the functional capacity of the two *Pachysoma* spp. microbiomes and their associated diets?
3. What is the capacity of microbial communities to digest refractive polycarbohydrates and how does it compare between the two different diets (plant detritus versus dry dung)?

### 4.2 Results and Discussion

#### 4.2.1 Comparison of Four Previously Published Host DNA Reduction Methods.

Shotgun sequencing randomly fragments and sequences “everything” within a DNA sample (Auburn et al., 2011). However, mDNA samples from the insect gut contain large concentrations of DNA from the insect host (Hunter et al., 2011, Feehery et al., 2013). This so-called host DNA contamination is thus problematic when studying gut microbiomes, with the majority of reads being of host origin (Carpi et al., 2011, Hunter et al., 2011, Feehery et
al., 2013). For example, it has been observed when studying insect gut microbiomes that over
80% of the shotgun reads can originate from the insect host (Carpi et al., 2011, Feehery et al.,
2013, Jia et al., 2013). Therefore in order to fully utilise shotgun sequencing to its full potential,
it is important to implement methods to reduce host DNA contamination.

Four previously published methods to reduce host DNA contamination were tested on
cockroach samples (Table 4.1). Ideally, these methods must allow unbiased access to all
microorganisms within the sample (Krishnan et al., 2014) while reducing the concentration of
host DNA extracted.
Table 4.1: Overview of methods tested for efficiency in reducing host DNA contamination during mDNA extraction.

<table>
<thead>
<tr>
<th>Method</th>
<th>No treatment (control)</th>
<th>Removal of Peritrophic Membrane</th>
<th>Filtration</th>
<th>Minimal homogenisation</th>
<th>Low-speed centrifugation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Methods used for amplicon sequencing samples</td>
<td>Removal of insect-derived PM during dissection</td>
<td>Extensive homogenisation to effectively release microorganisms from gut wall followed by filtration</td>
<td>Minimal homogenisation to reduce amount of host DNA released</td>
<td>Low-speed centrifugation to separate microbial cells from eukaryotic cells</td>
</tr>
<tr>
<td><strong>Host/source from original study</strong></td>
<td>Pachysoma spp. (this study)</td>
<td>Anoplophora glabripennis gut (wood feeding beetle)</td>
<td>Ips pini gut (bark beetle)</td>
<td>Human ileal and colonic biopsies</td>
<td>Macrotelmes annandaleigut (fungus-growing termites)</td>
</tr>
<tr>
<td><strong>Dissection</strong></td>
<td>Whole-gut dissected</td>
<td>PM removed before gut added to tube for homogenisation</td>
<td>Whole-gut dissected (as per control method)</td>
<td>Whole-gut dissected (as per control method)</td>
<td>Whole-gut dissected (as per control method)</td>
</tr>
<tr>
<td><strong>Homogenization</strong></td>
<td>Add CTAB and homogenise using a tissue grinder</td>
<td>Add CTAB and homogenise using a tissue grinder (as per control method)</td>
<td>Add CTAB; sonicate guts for 30s; macerate guts further with an epiprocessor and vortex at medium speed for 10s; filter contents through filter paper</td>
<td>Vortex at medium speed for 10s; centrifuge minimal cycle</td>
<td>Add PBS followed by multiple centrifugation cycles at 800xg</td>
</tr>
<tr>
<td><strong>mDNA extraction</strong></td>
<td>Typical phenol-chloroform method</td>
<td>Phenol-chloroform method (as per control method)</td>
<td>Phenol-chloroform method (as per control method)</td>
<td>Phenol-chloroform method (as per control method)</td>
<td>Phenol-chloroform method (as per control method)</td>
</tr>
<tr>
<td><strong>Shortcomings</strong></td>
<td>High concentrations of host DNA contamination</td>
<td>May be removing important microorganisms embedded in the PM</td>
<td>May remove a portion of fungal symbionts with host DNA</td>
<td>Homogenisation may not be efficient to release entire gut microbial community</td>
<td>May remove a portion of fungal symbionts with host DNA</td>
</tr>
</tbody>
</table>

The four tested methods yielded high quality DNA with good purity values compared to that of the no treatment (control) method (Table 4.2). PCR amplification of the insect 18S rRNA
gene was carried out with insect specific primers (detailed in Chapter 2.4.2). Only the low-speed centrifugation method was unsuccessful in amplifying the insect 18S rRNA gene (Figure 4.1a). Therefore, it appears that this method may successfully reduce the concentration of host DNA. Furthermore, the low-speed centrifugation method yielded the lowest concentration of mDNA. The 16S rRNA gene was amplified from all five samples using bacteria specific primers (See Chapter 2.4.2), suggesting that bacterial DNA was still present in the samples (Figure 4.1b).

**Table 4.2:** Nanodrop readings of mDNA yields and purity ratios for samples treated with four host DNA reduction methods.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>DNA concentration (ng/µl)</th>
<th>A$_{260/280}$</th>
<th>A$_{260/230}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No treatment</td>
<td>3327.9</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Removal of PM</td>
<td>1211.1</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Filtration</td>
<td>471.8</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Minimal homogenisation</td>
<td>60.3</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Low speed centrifugation</td>
<td>16.3</td>
<td>1.9</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Figure 4.1: Amplification of the a) 18S rRNA and b) 16S rRNA genes from cockroach gut samples subjected to all four host DNA reduction methods. L: molecular weight KAPA Universal Ladder, C: no treatment, 1: removal of PM, 2: filtration, 3: minimal homogenisation and 4: low-speed centrifugation. *Escherichia coli* cells were used as the positive control, therefore, no amplification was expected in gel (a).
4.2.2 Shotgun Sequencing of the Gut Microbiomes of *P. endroedyi* and *P. striatum*

Metagenomic DNA extracted using the low-speed centrifugation method from the gut samples of three individuals of *P. endroedyi* and *P. striatum* (Table 4.3) were pooled in equal concentrations and sent to Molecular Research (MR DNA), for shotgun sequencing using the Illumina Hiseq platform (Section 2.4.3).

Table 4.3: Nanodrop readings of mDNA yields and purity values for replicate *P. endroedyi* and *P. striatum* gut samples sent for shotgun sequencing.

<table>
<thead>
<tr>
<th>Species</th>
<th>Replicate</th>
<th>DNA concentration (ng/µl)</th>
<th>A260/280</th>
<th>A260/230</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. endroedyi</em></td>
<td>1</td>
<td>99.7</td>
<td>1.4</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>171.5</td>
<td>1.7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>77.3</td>
<td>1.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>141.7</td>
<td>1.5</td>
<td>0.7</td>
</tr>
<tr>
<td><em>P. striatum</em></td>
<td>2</td>
<td>45.9</td>
<td>1.6</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>157.1</td>
<td>2.0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

19 590 694 and 26 098 638 paired-end reads were recovered for *P. endroedyi* and *P. striatum*, respectively (Table 4.4). *De novo* assembly of high-quality reads using CLC genomic workbench resulted in 231,054 and 223,703 contigs. Both assemblies had N50 scores (the largest scaffold length where 50% of the genomic assembly is represented by scaffolds greater than this size (Kingsford et al., 2010)) of 1234 and 787 for *P. endroedyi* and *P. striatum*, respectively (Table 4.4). These N50 scores were comparable with those of another shotgun sequencing study on the insect gut microbiome of the Asian longhorned beetle, *Anoplophora glabripennis* (N50=938) (Scully et al., 2013), which along with the other metadata and metrics (Table 4.4) suggest that high quality assemblies were created. Rarefaction curves indicated successful sampling of the *P. endroedyi* gut microbiome (Figure 4.2a), suggesting that
complete microbial richness was analysed. Rarefaction curves for the *P. striatum* gut microbiome did not reach asymptote (Figure 4.2b), suggesting that the complete microbial richness may not be displayed.

**Table 4.4:** Sequencing metadata and assembly metrics for shotgun sequencing of the *P. endroedyi* and *P. striatum* gut microbiomes using CLC Genomics Workbench, MG-RAST and MEGAN.

<table>
<thead>
<tr>
<th>Parameters</th>
<th><em>P. endroedyi</em></th>
<th><em>P. striatum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. raw reads</td>
<td>19,590,694</td>
<td>26,098,638</td>
</tr>
<tr>
<td>No. reads after quality control</td>
<td>17,207,062</td>
<td>24,260,166</td>
</tr>
<tr>
<td>Average read length</td>
<td>132.63</td>
<td>131.96</td>
</tr>
<tr>
<td>Number of nucleotides</td>
<td>2,282,237,909</td>
<td>3,201,399,203</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>231,054</td>
<td>223,703</td>
</tr>
<tr>
<td>N25</td>
<td>6,346</td>
<td>14,132</td>
</tr>
<tr>
<td>N50</td>
<td>1,234</td>
<td>787</td>
</tr>
<tr>
<td>N75</td>
<td>452</td>
<td>405</td>
</tr>
<tr>
<td>Number of unassembled reads</td>
<td>3,644,008</td>
<td>2,764,456</td>
</tr>
<tr>
<td>Assembled reads</td>
<td>13,563,054</td>
<td>21,495,710</td>
</tr>
<tr>
<td>Minimum contig length (bp)</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>Maximum contig length (bp)</td>
<td>220,532</td>
<td>256,410</td>
</tr>
<tr>
<td>Average contig length</td>
<td>799</td>
<td>707</td>
</tr>
<tr>
<td>Ribosomal RNA (%)</td>
<td>0.9</td>
<td>3.9</td>
</tr>
<tr>
<td>Annotated protein (%)</td>
<td>65.6</td>
<td>32.6</td>
</tr>
<tr>
<td>Unknown protein (%)</td>
<td>31.9</td>
<td>62.7</td>
</tr>
<tr>
<td>Number of ORFs</td>
<td>267,509</td>
<td>127,539</td>
</tr>
<tr>
<td>Number of ORFs assigned to COG (%)</td>
<td>101,542 (37.9)</td>
<td>44,762 (35.1)</td>
</tr>
</tbody>
</table>
4.2.3 Community Composition of the *Pachysoma* Metagenome

Taxonomic annotation of contigs was done using both MG-RAST and MEGAN (Figure 4.3). The results of both analyses were comparable to each other with the exception of eukaryotes in *P. striatum*, where approximately twice the number of eukaryotic phylotypic signals in the gut of *P. striatum* were detected by MEGAN (22.1% contigs) compared to MG-RAST (11.2% contigs). However, in both analyses the majority of eukaryotic contigs were removed before analysis.
associated to Metazoa (animals), which will be removed from further analysis. Therefore, the MEGAN results were used further for both taxonomic and functional annotation.

As expected, Bacteria was the most dominant domain in the guts of both Pachysoma species (Figure 4.3), which is typical for insect gut systems (Belda et al., 2011, Liu et al., 2013, Shi et al., 2013, Rahman, 2016). This result also illustrates that the low-speed centrifugation method does appear to successfully reduce host DNA contamination, as has been demonstrated previously (Liu et al., 2011). This is further collaborated by the fact that few or no contigs were associated with either Pachysoma species using both MG-RAST and MEGAN (MG-RAST: P. endroedyi = 0 contigs, P. striatum = 0 contigs; MEGAN: P. endroedyi = 0 contigs, P. striatum = 1 Pachysoma sp. contig). Furthermore, Eukaryotes were detected in lower percentages than recorded in studies where host DNA reduction methods were not used (Carpi et al., 2011, Jia et al., 2013). However, the number of Eukaryotic contigs was substantially higher in the coprophagous P. striatum (22.1% contigs) than the detritivorous P. endroedyi (0.1% contigs). Archaea and viruses were also detected in low abundances (<2% contigs) in both Pachysoma species.
Shotgun sequencing revealed a greater bacterial diversity than detected using amplicon sequencing. A total 39 bacterial phyla, all of which were found in the guts of the plant detritus feeding *P. endroedyi*, while only 24 bacterial phyla were detected in the gut of the coprophagous *P. striatum*. Bacteria that could not be classified past domain were also present in both *Pachysoma* gut metagenomes. Although the diversity is greater than recorded in the
amplicon study where a total 14 bacterial phyla were detected (Chapter 3), both studies showed great differences in diversity between the two species which could be due to dietary differences as previously discussed (Chapter 3; Franzini et al. (2016)). Furthermore, in the shotgun sequencing datasets, four phyla were the dominant (>2% contigs) bacterial communities in the gut of both Pachysoma species, specifically Bacteroidetes (most dominant [50.6% of bacteria-derived contigs] in P. endroedyi), Firmicutes (most dominant [36.4% of bacteria-derived contigs] in P. striatum), Proteobacteria and Actinobacteria (Figure 4.4). These abundances may be representative of specific diets as already discussed in the amplicon study (Chapter 3). Bacteriodetes are typically highly abundant in plant feeding insects (Boucias et al., 2013, Schauer et al., 2014, Montagna et al., 2015, Waite et al., 2015), with Proteobacteria and/or Firmicutes most dominant in insects feeding on simpler diets (Ahn et al., 2012, Wang et al., 2014, Andongma et al., 2015, Pandey and Rajagopal, 2016). This corresponds with Bacteriodetes being most abundant in gut samples of the plant detritus feeding P. endroedyi and Firmicutes most abundant in P. striatum. The remaining phyla represented abundances of less than 2% of the contigs for both Pachysoma spp. This result was comparable to P. striatum amplicon sequencing results but varied from that for P. endroedyi, where seven phyla were considered abundant (>2% 16S rRNA sequence reads). The largest inconsistency was the difference in abundance of the bacterial phyla Planctomycetes and Elusimicrobia between the two sequencing technologies.
Diversity and abundance of bacterial phyla in the gut samples of *P. endroedyi* (green) and *P. striatum* (brown) using shotgun (Illumina Hiseq) analysed using MEGAN.

Metazoa (animals) was the most abundant Eukaryotic kingdom accounting for 58.9% and 99.1% of the eukaryotic-derived contigs for *P. endroedyi* and *P. striatum*, respectively. Possible causes for this high number of Metazoa-derived contigs in the *P. striatum* metagenome include, various sources contaminating the dung including the animal that excreted the faeces or as part of the original diet (Paxinos et al., 1997, Symondson, 2002, Jarman et al., 2004). The majority of Metazoa-derived contigs were from the class Insecta (84.5%) followed by Arachnida (1.0%), Chromadorea (1.0%) and Mammalia (0.9%).

A significant result, when comparing results of the 2 technologies, is the presence of fungal gut communities in both *Pachysoma* species. This is in contrast to the amplicon study where the ITS region could not be amplified for *P. endroedyi* gut samples. Furthermore, discrepancies were noted between the fungal results of the two technologies with regard to the detection of yeasts in the gut samples. True yeasts (Saccharomycetes) were detected using shotgun sequencing in low abundances in both *P. endroedyi* (2 [8.3%] fungal contigs) and *P. striatum* (1 [1.0%] fungal contig) gut samples, respectively but were absent from *P. striatum* samples in the amplicon study. Saccharomycetes are typically detected in insect guts.
using culture-dependent methods and clone libraries (Suh and Blackwell, 2004, Suh et al., 2004b, Suh et al., 2004a, Suh et al., 2005, Nguyen et al., 2007).

4.2.4 Functional Profiling of the Metagenomes of *P. endroedyi* and *P. striatum*.

Microbial gut communities are considered to fully or partially assist the host in digestion of different components within the diet (Douglas, 2009). Therefore, the functional capacity of the gut microbial communities of *P. endroedyi* and *P. striatum* were determined after removal of all Metazoa and Streptophyta derived ORFs from each respective dataset.

Only 37.9% and 35.1% microbial-derived ORFs from *P. endroedyi* and *P. striatum* gut, respectively, could be classified into 22 functional categories against the Clusters of Orthologous Groups (COG) database (Figure 4.5). Genes from eight COG categories were overrepresented in both datasets namely; carbohydrate transport and metabolism (G; 10.8-12.4%), replication, recombination and repair (L; 9.1-11.1%), amino acid transport and metabolism (E; 10.3-10.7%), cell wall/membrane/envelope biogenesis (M; 8.7-9.2%), translation, ribosomal translation and biogenesis (J; 8.1-8.2%), energy production and conversion (C; 7.8-8.2%), Inorganic ion transport and metabolism (P; 7.1-8.0%) and transcription (K; 6.6-8.3%). The enrichment of such functional genes were not significantly different between the microbiomes of the two *Pachysoma* species (p>0.05). It appears that *Pachysoma* gut communities may be suitably adapted to support their own nutritional needs through enrichment of metabolic and cellular processing genes (Rahman, 2016). Furthermore, the enrichment of such genes may be linked to the content of plant biomass in the respective diets, particularly regarding genes involved in carbohydrate transport and metabolism (G), energy production and conversion (C) and cell wall/membrane/envelope biogenesis (M) potentially indicating microorganisms need to attach to plant cell walls (Suen et al., 2011).

Functional categories absent from the annotated *Pachysoma* ORFs were nuclear structure (category Y) and the poorly categorized R (general function prediction only) and S (function unknown), which are commonly abundant in metagenomic datasets from insect gut samples.
(Belda et al., 2011, Shi et al., 2013, Rahman, 2016). The absence of such genes could indicate a weakness in the capacity of the software to efficiently bin genes into COG categories. For example, it is noted that a large percentage of genes were not categorised (\textit{P. endroedyi}: 37.9\%, \textit{P. striatum}=35.1\%). As the COG categories R and S are bins for genes where the function cannot be clearly identified, it is possible that such genes may have been classed as 'unidentified' and would therefore not be included in the accepted COG classifications (Figure 4.5).

![Figure 4.5: Functional assignments of microbial-derived ORFs from both \textit{Pachysoma} species to categories of COGs generated by MEGAN.]

### 4.2.4.1 Candidate Genes for Carbohydrate Transport and Metabolism

The focus of this study is the effect of host diet on microbiome assembly and the associated role of the respective microorganisms in digestion of that particular diet. As plant biomass is present in both diets (Holter and Scholtz, 2011, 2013), genes involved in carbohydrate transport and metabolism are of interest in this study. The diet of \textit{P. endroedyi} consists basically of undigested plant fragments (Harrison et al., 2003) that may have been subjected to previous degradation by environmental microorganisms (microbial conditioning; Swift et al. (1979)). In contrast, the coprophagous diet of \textit{P. striatum} contains smaller plant
fragments that have already been subjected to one round of digestion/mastication (Anderson and Coe, 1974) along with associated microbial conditioning. It could be expected that the composition of plant polymers would differ between the two diets due to the faecal plant matter having been previously digested. However, it is unknown whether or not faecal plant matter has been partially degraded or not. Furthermore, without knowing the species of plant in the two diets it would be speculative to suggest that different functional capacities would be needed to digest the respective plant biomass.

Genes encoding glycoside hydrolases (GHs; enzymes which hydrolyse glycosidic bonds between carbohydrates (http://www.cazy.org/; Lombard et al. (2014)) were classified into 51 and 53 families for *P. endroedyi* and *P. striatum*, respectively, based on combined Pfam domain and KEGG enzyme class (E.C.) assignments (Figure 4.6). The GH profiles were similar between the two species with the hemicellulose and cellulose GH2 (*P. endroedyi*=21.1%; *P. striatum*=15.4%) and GH3 (*P. endroedyi*=9.0%; *P. striatum*=9.8%) CAZyme families being the most abundant for the gut microbiomes of both *Pachysoma* species. Furthermore, abundances were mostly similar between the GH families of the two *Pachysoma* species, with the exception of GH1 (*P. endroedyi* = 9.1% ORFs, *P. striatum* = 1.5% ORFs; Figure 4.6). Genes encoding GHs from family one are typically cellulases and hemicellulases. Therefore, it is interesting that the coprophagous *P. striatum* has a substantially larger, albeit not statistically significant (*p>*0.05), abundance of genes in this family (7.55% ORFs). Plant fragments found in dung have previously undergone digestion/mastication by the primary animal (Anderson and Coe, 1974). Therefore, the large difference in abundance of GH1 enzymes suggests they may be necessary to degrade plant fragments previously subjected to digestion and expelled in the dung. Eight families were reported in only one of the two *Pachysoma* species, in low abundances (<6 ORFs). The associated KEGG E.C. assignments of all GH families are listed in Table 4.5.
Figure 4.6: Distribution of glycosyl hydrolase (GH) families in the gut metagenomes of *P. endroedyi* and *P. striatum* annotated using hmmmer.
Table 4.5: Glycosyl hydrolase families detected in the gut metagenomes of *P. endroedyi* and *P. striatum* and the associated KEGG E.C. assignments were applicable.

<table>
<thead>
<tr>
<th>Family</th>
<th><em>P. endroedyi</em> ORFs</th>
<th><em>P. striatum</em> ORFs</th>
<th>Pfam HMM Name</th>
<th>Pfam accession</th>
<th>KEGG ID</th>
<th>KEGG functions</th>
<th>KEGG ECs</th>
<th><em>P. endroedyi</em> ORFs</th>
<th><em>P. striatum</em> ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74</td>
<td>269</td>
<td>Glyco_hydro_1</td>
<td>PF00232.16</td>
<td>K01223</td>
<td>6-phospho-beta-glucosidase</td>
<td>[EC:3.2.1.86]</td>
<td>56</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF00703.19</td>
<td>K01190</td>
<td>beta-galactosidase</td>
<td>[EC:3.2.1.23]</td>
<td>174</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02836.15</td>
<td>K01192</td>
<td>beta-mannosidase</td>
<td>[EC:3.2.1.25]</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02837.16</td>
<td>K01195</td>
<td>beta-galactosidase</td>
<td>[EC:3.2.1.23]</td>
<td>254</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02837.16</td>
<td>K01190</td>
<td>beta-galactosidase</td>
<td>[EC:3.2.1.23]</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02837.16</td>
<td>K01192</td>
<td>beta-mannosidase</td>
<td>[EC:3.2.1.25]</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02837.16</td>
<td>K01195</td>
<td>beta-galactosidase</td>
<td>[EC:3.2.1.23]</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_2</td>
<td>PF02837.16</td>
<td>K18577</td>
<td>mannosylglycoprotein endo-beta-mannosidase</td>
<td>[EC:3.2.1.152]</td>
<td>188</td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>1034</td>
<td>456</td>
<td>Glyco_hydro_2</td>
<td>PF00933.19</td>
<td>K05349</td>
<td>beta-glucosidase</td>
<td>[EC:3.2.1.21]</td>
<td>176</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_3</td>
<td>PF00933.19</td>
<td>K05349</td>
<td>beta-glucosidase</td>
<td>[EC:3.2.1.21]</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_3</td>
<td>PF01915.20</td>
<td>K05349</td>
<td>beta-glucosidase</td>
<td>[EC:3.2.1.21]</td>
<td>153</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_3</td>
<td>PF01915.20</td>
<td>K15920</td>
<td>beta-D-glucosidase</td>
<td>[EC:3.2.1.37]</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>439</td>
<td>292</td>
<td>Glyco_hydro_4</td>
<td>PF02056.14</td>
<td>K01222</td>
<td>6-phospho-beta-glucosidase</td>
<td>[EC:3.2.1.86]</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_4</td>
<td>PF02056.14</td>
<td>K07406</td>
<td>alpha-galactosidase</td>
<td>[EC:3.2.1.22]</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_4</td>
<td>PF02056.14</td>
<td>K01232</td>
<td>maltose-6'-phosphate glucosidase</td>
<td>[EC:3.2.1.122]</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_4</td>
<td>PF02056.14</td>
<td>K01222</td>
<td>6-phospho-beta-glucosidase</td>
<td>[EC:3.2.1.86]</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>71</td>
<td>Glyco_hydro_4</td>
<td>PF11975.6</td>
<td>K07406</td>
<td>alpha-galactosidase</td>
<td>[EC:3.2.1.22]</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_4</td>
<td>PF11975.6</td>
<td>K01232</td>
<td>maltose-6'-phosphate glucosidase</td>
<td>[EC:3.2.1.122]</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cellulase</td>
<td>PF00150.16</td>
<td>K01179</td>
<td>K01210</td>
<td>endoglucanase</td>
<td>glucan 1,3-beta-glucosidase</td>
<td>[EC:3.2.1.4]</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>54</td>
<td>Cellulase</td>
<td>PF00150.16</td>
<td>K19355</td>
<td>K01210</td>
<td>mannan endo-1,4-beta-mannosidase</td>
<td>[EC:3.2.1.78]</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>10</td>
<td>Glyco_hydro_6</td>
<td>PF01871.4</td>
<td>K1179</td>
<td>No EC evidence</td>
<td>endoglucanase</td>
<td>[EC:3.2.1.4]</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>5</td>
<td>Glyco_hydro_8</td>
<td>PF01270.15</td>
<td>K01179</td>
<td>endoglucanase</td>
<td>[EC:3.2.1.4]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>8</td>
<td>Glyco_hydro_9</td>
<td>PF00759.17</td>
<td>K01179</td>
<td>endoglucanase</td>
<td>[EC:3.2.1.4]</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>110</td>
<td>83</td>
<td>GHL10</td>
<td>Glyco_hydro_10</td>
<td>PF02638.13</td>
<td>No EC evidence</td>
<td>[EC:3.2.1.4]</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>4</td>
<td>Glyco_hydro_11</td>
<td>PF0457.15</td>
<td>K01181</td>
<td>endo-1,4-beta-xylanase</td>
<td>[EC:3.2.1.8]</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14*</td>
<td>4</td>
<td>4</td>
<td>Glyco_hydro_14</td>
<td>PF01373.15</td>
<td>Beta-Amylase</td>
<td>Beta-Amylase</td>
<td>[EC:3.2.1.2]</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>9</td>
<td>Glyco_hydro_15</td>
<td>PF00723.19</td>
<td>No EC evidence</td>
<td>[EC:3.2.1.4]</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>31</td>
<td>Glyco_hydro_16</td>
<td>PF00722.19</td>
<td>No EC evidence</td>
<td>[EC:3.2.1.4]</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td>30</td>
<td>Glyco_hydro_18</td>
<td>PF00704.26</td>
<td>K06306</td>
<td>K01183</td>
<td>lysozyme</td>
<td>[EC:3.2.1.17]</td>
<td>25</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>8</td>
<td>Glyco_hydro_19</td>
<td>PF0182.17</td>
<td>K03791</td>
<td>K01183</td>
<td>putative chitinase</td>
<td>[EC:3.2.1.14]</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>146</td>
<td>109</td>
<td>Glyco_hydro_20</td>
<td>PF00728.20</td>
<td>K12373</td>
<td>K12373</td>
<td>hexosaminidase</td>
<td>[EC:3.2.1.52]</td>
<td>78</td>
</tr>
<tr>
<td>25</td>
<td>62</td>
<td>26</td>
<td>Glyco_hydro_25</td>
<td>PF01183.18</td>
<td>K07273</td>
<td>K06381</td>
<td>lysozyme</td>
<td>[EC:3.2.1.17]</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>14</td>
<td>19</td>
<td>Glyco_hydro_26</td>
<td>PF01256.13</td>
<td>K01218</td>
<td>K07214</td>
<td>mannan endo-1,4-beta-mannosidase</td>
<td>[EC:3.2.1.78]</td>
<td>7</td>
</tr>
<tr>
<td>28</td>
<td>114</td>
<td>72</td>
<td>Glyco_hydro_28</td>
<td>PF00295.15</td>
<td>K18650</td>
<td>K01811</td>
<td>alpha-D-xiloside xylohydrolase</td>
<td>[EC:3.2.1.82]</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>75</td>
<td>38</td>
<td>Glyco_hydro_30</td>
<td>PF02055.14</td>
<td>K01201</td>
<td>K15924</td>
<td>glucosylceramidase</td>
<td>[EC:3.2.1.45]</td>
<td>26</td>
</tr>
<tr>
<td>31</td>
<td>149</td>
<td>75</td>
<td>Glyco_hydro_31</td>
<td>PF01055.24</td>
<td>K01811</td>
<td>K01187</td>
<td>alpha-D-xiloside xylohydrolase</td>
<td>[EC:3.2.1.177]</td>
<td>66</td>
</tr>
<tr>
<td>32</td>
<td>118</td>
<td>53</td>
<td>Glyco_hydro_32C</td>
<td>PF08244.10</td>
<td>K03332</td>
<td>K01193</td>
<td>fructan beta-fructosidase</td>
<td>[EC:3.2.1.80]</td>
<td>4</td>
</tr>
</tbody>
</table>

© University of Pretoria
<p>| Glyco_hydro_32N | PF00251.18 | K01193 | beta-fructofuranosidase | [EC:3.2.1.26] | 6 | 4 |
| Glyco_hydro_35 | PF01301.17 | K12308 | beta-galactosidase | [EC:3.2.1.23] | 8 | 17 |
| Glyco_hydro_38C | PF01074.20 | K01191 | alpha-mannosidase | [EC:3.2.1.24] | 32 | 14 |
| Glyco_hydro_38 | PF07748.11 | K01191 | alpha-mannosidase | [EC:3.2.1.24] | 33 | 9 |
| Glyco_hydro_39 | PF01229.15 | K01191 | alpha-mannosidase | [EC:3.2.1.24] | 30 | 24 |
| Glyco_hydro_42C | PF03718.11 | K01230 | alpha-mannosidase | [EC:3.2.1.113] | 1 | 0 |
| Glyco_hydro_42 | PF03200.14 | K01228 | mannosyl-oligosaccharide alpha-1,2-mannosidase | [EC:3.2.1.106] | 1 | 0 |
| Glyco_hydro_53 | PF05988.12 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 11 | 9 |
| Glyco_hydro_55C | PF07745.11 | K01224 | alpha-glucuronidase | [EC:3.2.1.139] | 16 | 12 |
| Glyco_hydro_55 | PF03636.13 | K00691 | maltose phosphorylase | [EC:2.4.1.8] | 5 | 2 |
| Glyco_hydro_55 | PF03636.13 | K00691 | maltose phosphorylase | [EC:2.4.1.8] | 11 | 5 |
| Glyco_hydro_55 | PF03636.13 | K00691 | maltose phosphorylase | [EC:2.4.1.8] | 6 | 3 |
| Glyco_hydro_55 | PF13199.4 | K05988 | dextranase | [EC:3.2.1.11] | 0 | 6 |
| Glyco_hydro_55 | PF07477.10 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 11 | 9 |
| Glyco_hydro_55 | PF07486.10 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 16 | 12 |
| Glyco_hydro_55 | PF03648.12 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 3 | 6 |
| Glyco_hydro_55 | PF13199.4 | K05988 | dextranase | [EC:3.2.1.11] | 0 | 6 |
| Glyco_hydro_55 | PF07477.10 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 11 | 9 |
| Glyco_hydro_55 | PF07486.10 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 16 | 12 |
| Glyco_hydro_55 | PF03648.12 | K01235 | alpha-glucuronidase | [EC:3.2.1.139] | 3 | 6 |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>21</td>
<td>11</td>
<td>Glyco_hydro_70</td>
<td>PF02324.14</td>
<td>No EC evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>75</td>
<td>36</td>
<td>Glyco_hydro_76</td>
<td>PF03663.12</td>
<td>K08257</td>
<td>mannan endo-1,6-alpha-mannosidase</td>
<td>[EC:3.2.1.101]</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>75</td>
<td>36</td>
<td>Glyco_hydro_76</td>
<td>PF03663.12</td>
<td>K08257</td>
<td>uncharacterized protein alpha-1,6-mannanase</td>
<td>[EC 3.2.1.101]</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>85</td>
<td>30</td>
<td>Glyco_hydro_77</td>
<td>PF02446.15</td>
<td>K00705</td>
<td>4-alpha-glucanotransferase</td>
<td>[EC:2.4.1.25]</td>
<td>74</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>0</td>
<td>1</td>
<td>Glyco_hydro_81</td>
<td>PF03639.11</td>
<td>K01180</td>
<td>endo-1,3(4)-beta-glucanase</td>
<td>[EC:3.2.1.6]</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>85*</td>
<td>2</td>
<td>0</td>
<td>Glyco_hydro_85</td>
<td>PF03644.11</td>
<td>K15532</td>
<td>unsaturated rhamnogalacturonyl hydrolase</td>
<td>[EC:3.2.1.172]</td>
<td>46</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>206</td>
<td>150</td>
<td>Glyco_hydro_88</td>
<td>PF07470.11</td>
<td>K18581</td>
<td>unsaturated chondroitin disaccharide hydrolase</td>
<td>[EC:3.2.1.180]</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_88</td>
<td>PF03639.11</td>
<td>K00705</td>
<td>uncharacterized protein</td>
<td></td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_88</td>
<td>PF03644.11</td>
<td>K06888</td>
<td>uncharacterized protein</td>
<td></td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>344</td>
<td>145</td>
<td>Glyco_hydro_92</td>
<td>PF07971.10</td>
<td>No EC evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>218</td>
<td>209</td>
<td>Glyco_hydro_97</td>
<td>PF10566.7</td>
<td>K01187</td>
<td>alpha-glucosidase</td>
<td>[EC:3.2.1.20]</td>
<td>41</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_97</td>
<td>PF14509.4</td>
<td>K01187</td>
<td>alpha-glucosidase</td>
<td>[EC:3.2.1.20]</td>
<td>55</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_98</td>
<td>PF08307.9</td>
<td>K01187</td>
<td>alpha-glucosidase</td>
<td>[EC:3.2.1.20]</td>
<td>63</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>2</td>
<td>0</td>
<td>Glyco_hydro_98C</td>
<td>PF08030.9</td>
<td>No EC evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>0</td>
<td>3</td>
<td>Glyco_hydro_99</td>
<td>PF16317.3</td>
<td>No EC evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>8</td>
<td>3</td>
<td>Glyco_hydro_101</td>
<td>PF12905.5</td>
<td>endo-α-N-acetylglactosaminidase</td>
<td>[EC 3.2.1.97]</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>116</td>
<td>19</td>
<td>Glyco_hydro_106</td>
<td>PF17132.2</td>
<td>L-rhamnose-H+ transport protein alpha-L-rhamnosidase</td>
<td>[EC:3.2.1.40]</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108*</td>
<td>6</td>
<td>7</td>
<td>Glyco_hydro_108</td>
<td>PF05838.10</td>
<td>N-acetylmuramidase</td>
<td>[EC:3.2.1.17]</td>
<td>0</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114*</td>
<td>2</td>
<td>0</td>
<td>Glyco_hydro_114</td>
<td>PF03537.11</td>
<td>endo-α-1,4-polygalactosaminidase</td>
<td>[EC 3.2.1.109]</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>29</td>
<td>8</td>
<td>Glyco_hydro_115</td>
<td>PF15979.3</td>
<td>No EC evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>22</td>
<td>19</td>
<td>Glyco_hydro_116</td>
<td>PF2215.6</td>
<td>K17108</td>
<td>non-lysosomal glucosylceramidase</td>
<td>[EC:3.2.1.45]</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_116</td>
<td>PF04685.11</td>
<td>K17108</td>
<td>non-lysosomal glucosylceramidase</td>
<td>[EC:3.2.1.45]</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>177</td>
<td>72</td>
<td>Glyco_hydro_127</td>
<td>PF07944.10</td>
<td>K09955</td>
<td>uncharacterized protein</td>
<td></td>
<td>97</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>58</td>
<td>36</td>
<td>Glyco_hydro_130</td>
<td>PF04041.11</td>
<td>K18785</td>
<td>beta-1,4-mannooligosaccharide/beta-1,4-mannosyl-N-acetylglucosamine phosphorylase</td>
<td>[EC:2.4.1.319/ 2.4.1.320]</td>
<td>14</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glyco_hydro_130</td>
<td>PF15979.3</td>
<td>K16212</td>
<td>4-O-beta-D-mannosyl-D-glucose phosphorylase</td>
<td>[EC:2.4.1.281]</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Glycosyl hydrolase catalytic core

© University of Pretoria
Given the large differences in bacterial diversity, it is surprising that functional diversity of GH families was similar between the two *Pachysoma* microbiomes. Of particular interest are enzymes involved in the breakdown of the major components of plant cell walls, namely cellulose, hemicellulose and other plant polysaccharides. Cellulases were classified into six GH families, namely GH1, GH3, GH5, GH6 (only *P. striatum*), GH8 (only *P. striatum*) and GH9. Their corresponding KEGG E.C. assignments (assigned using KAAS) suggest that only two classes of cellulases were present in both *Pachysoma* spp., namely endoglucanases, and β-glucosidases. Endoglucanases cleave β-1,4-glucosidic bonds while β-glucosidases hydrolyses cellobiose to glucose (Terra and Ferreira, 1994, Lynd et al., 2002). No ORFs were assigned to the third class of enzymatic activity, the exoglucanases (E.C. 3.2.1.74; 3.2.1.91), which cleave cellobiose or glucose from cellulose polysaccharide chains (Terra and Ferreira, 1994, Lynd et al., 2002). Although microbial exoglucanases have been recorded from the gut of the Asian longhorned beetle (Geib et al., 2009, Scully et al., 2013), such enzymes have not been associated with higher termites (Do et al., 2014, Rahman, 2016), and appear to be restricted to populations of protozoa from lower termites (Zhou et al., 2007, Do et al., 2014, Rahman, 2016). Furthermore, insects do not appear to produce exogenous exoglucanses (Calderón-Cortés et al., 2012). It has been suggested that the presence of an elongated gut and mastication of the plant materials (features exhibited by *Pachysoma* (Holter and Scholtz, 2011, 2013)) by the insect may account for the absence of such exoglucanases (Calderón-Cortés et al., 2012). However, it should be noted that all GH families with known exoglucanase activity were present in both *Pachysoma* species, and therefore these enzymes may be present in the metagenomes although KEGG E.C. assignment was not possible.

KEGG E.C assignments showed the presence of a number of hemicellulases potentially capable of acting on a variety of substrates (Table 4.5). Xylanases were underrepresented in both *Pachysoma* metagenomes. Only two ORFs from the *P. striatum* metagenome had KEGG E.C. assignments to xylanases from GH11 and GH30, whereas no xylanases were assigned
from the *P. endroedyi* metagenome (Table 4.5). However, GH families known to include xylanases were present within both metagenomes. This result is unexpected as xylan is the most abundant hemicellulose (Kulkarni et al., 1999, Shalom and Shoham, 2003) and an abundant polymer of plant cell walls (Scheller and Ulvskov, 2010) which varies between plant species (Kulkarni et al., 1999, Beg et al., 2001, Collins et al., 2005, Scheller and Ulvskov, 2010). Furthermore, no genes encoding for α-arabinofuranosidase, which are also involved in xylan breakdown (Beg et al., 2001, Shalom and Shoham, 2003), were detected in either *Pachysoma* gut metagenome. Arabinofuranosidases are abundant in metagenomes from grass decomposing systems including insect gut samples (He et al., 2013), suggesting the absence of this plant type from *Pachysoma* diets. However, other enzymes assisting in xylan breakdown (Beg et al., 2001, Shalom and Shoham, 2003) were present, namely β-xylosidase (GH3, GH43 and GH52) and α-glucuronidase (GH67). The majority of enzymes were linked to mannose and galactose metabolism (Table 4.5), which are significant components of hemicellulases in a variety of plants (Scheller and Ulvskov, 2010) Figure 4.7 gives summary of the cellulases and hemicellulases present in the *Pachysoma* gut metagenomes.

Microbial conditioning of the food-source may be responsible for the absence of lignin degrading genes in both metagenomes. Although the lignin content of the *Pachysoma* diet is potentially lower than that of wood-feeding insects, it is not expected to be absent from either the dung or plant detritus substrates. The absence of such genes is recorded in both higher and lower termites feeding on a variety of diets which contain substantial concentrations of lignin, such as wood (Warnecke et al., 2007, He et al., 2013, Rahman, 2016). Microbial conditioning may be responsible for the absence of such genes in fungus-growing termites (Hyodo et al., 2000, Ohkuma et al., 2001, Hyodo et al., 2003). Furthermore, disruption of the physical integrity of the plant biomass may reduce lignin concentrations (Hyodo et al., 1999).
Figure 4.7: Summary of the plant cell wall degrading enzymes present in the *P. endroedyi* and *P. striatum* gut metagenomes. Only hemicellulases present in at least one of the metagenomes are listed. Enzymes are highlighted according to the metagenome in which they were detected: *P. striatum* (brown) or both *Pachysoma* metagenomes (yellow).
KEGG E.C assignments (after removal of possible false positive identifications) could only be made for 43.9% and 55.1% of the ORFs assigned to a GH family for the *P. endroedyi* and *P. striatum* gut microbial metagenomes, respectively. This result, coupled with the low percentage of ORFs that could be classified into COG categories (*P. endroedyi* = 37.9% and *P. striatum* = 35.1%), suggests that novel enzymes may be present within the *Pachysoma* metagenomes.

### 4.2.4.2 Other Nutritional Pathways in the *Pachysoma* Gut Metagenome

Due to the similarities of plant degrading potential between the two *Pachysoma* gut microbiomes, other nutritional pathways were compared to determine variations in the functional capacity of the two *Pachysoma* microbiomes dependent on diet. Comparisons were made between genes from the abundant COG categories E (amino acid transport and metabolism) in the context of nitrogen metabolism, C (energy production and conversion) and G (carbohydrate transport and metabolism) by assessing the respective KEGG pathways. Furthermore, vitamin and sterol metabolism were also considered, due to the importance of these nutritional components to insects.

#### 4.2.4.2.1 Nitrogen Metabolism

Nitrogen is an essential nutrient to insects with those feeding on nitrogen rich diets possibly meeting their nutritional needs through their diet, whereas those feeding on nutrient poor diets, such as wood, commonly rely on microbial symbionts to fix nitrogen (Douglas, 2009). Nitrogen content of plant detritus varies (Kemp et al., 2003, Hättenschwiler et al., 2008), whereas dung is considered a nitrogen rich source (Liao et al., 2004), although the nitrogenous content of naturally dried dung and plant detritus fed on by *Pachysoma* is not well understood. However, it appears that the nitrogen content of both *Pachysoma* diets is sufficient due to the paucity of nitrogen fixation genes. Furthermore, only a small percentage of ORFs (*P. endroedyi* = 2.1%; *P. striatum* = 2.5%) were assigned to KEGG nitrogen metabolism pathways. The majority of genes were involved in the assimilation of ammonium and synthesis of the α-
amino acid L-glutamate (Figure 4.8) corresponding with the insects utilizing fixed nitrogen from the diet. Similar pathways were also dominant in a wood feeding beetle (Scully et al., 2013).

**Figure 4.8:** Adapted KEGG nitrogen metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes.

Microbial debris excreted with the dung (Mlambo et al., 2015) or new microbial decomposers may serve as a source of nitrogen in the dung (He et al., 2013). As both the dung and plant detritus are expected to undergo microbial conditioning before consumption by *Pachysoma*, it is possible that microbial cells are a major source of nitrogen for both
*Pachysoma* species and would explain the lack of nitrogen fixation genes in the *P. striatum* metagenome. This is consistent with the presence of genes encoding full or partial peptidoglycan metabolism pathways (a component of microbial cell walls) in both metagenomes (Figure 4.9). Furthermore, a large percentage of genes encoded to the COG category E amino acid transport and metabolism (10.3-10.7%). Full or partial pathways were present in both species for metabolism of all the essential amino acids. Little difference was noted in the genes encoding amino acid metabolism between the two species despite the availability of amino acids from dung (Barbehenn et al., 1999, Nation, 2008), although the amino acid composition of the exact diet fed on by *P. endroedyi* and *P. striatum* is unknown. This suggests that *P. endroedyi* may be able to metabolise amino acids from dung, enabling them to potentially feed on this food-source when plant detritus is unavailable.
Figure 4.9: Adapted KEGG peptidoglycan biosynthesis reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes.
4.2.4.2.3 Energy Metabolism

Genes encoding oxidative phosphorylation (\(P. \text{endroedyi} = 19.2\%\) and \(P. \text{striatum} = 25.6\%\) energy metabolism ORFs) and methane metabolism (\(P. \text{endroedyi} = 29.1\%\) and \(P. \text{striatum} = 22.8\%\) energy metabolism ORFs) pathways were abundant within both \(Pachysoma\) spp. metagenomes (Figure 4.10). The high abundance of genes involved in methane metabolism is not proportional to the abundance of methanogens within the gut. However, upon further inspection few genes were involved in synthesis of methane with the methane molecular marker gene, \textit{mcrA} (K00399) (detected from “\textit{Candidatus Methanomethylophilus alvus}”, a known methanogen from the human gut; (Borrel et al., 2012)) only detected in the \(P. \text{endroedyi}\) metagenome. This corresponds with the diets of \(Pachysoma\), especially dung, being substantial methane sources (Keppler et al., 2006, Willén, 2011)

\[\text{Figure 4.10: Abundance and diversity of genes encoding energy metabolism pathways (KEGG) for } P. \text{endroedyi (green) and } P. \text{striatum (brown) gut microbial metagenomes.}\]

4.2.4.2.2 Carbohydrate Metabolism

Both \(Pachysoma\) metagenomes had an abundance of genes encoding the two major polymer hydrolysis pathways: glycolysis (\(P. \text{endroedyi} = 11.0\%\) and \(P. \text{striatum} = 11.6\%\)
carbohydrate metabolism ORFs) and pentose phosphate (\(P. \text{endoodyi} = 8.9\%\) and \(P. \text{striatum} = 7.5\%\) carbohydrate metabolism ORFs) pathways (Figure 4.11). This is similar to results noted in other plant feeding insects such as termites (Rahman, 2016). Genes encoding proteins for the starch and sucrose metabolism pathway were more abundant within \(P. \text{endoodyi}\) than \(P. \text{striatum}\) which corresponds with the larger concentration of plant materials in detritus than dung. This further suggests that \(P. \text{striatum}\) may not feed exclusively on plant biomass but also on the brown liquid component of the dung.

![Figure 4.11: Abundance and diversity of genes encoding carbohydrate metabolism pathways (KEGG) for \(P. \text{endoodyi}\) (green) and \(P. \text{striatum}\) (brown) gut microbial metagenomes.](image)

### 4.2.4.2.4 Vitamin, Sterol and Fatty Acid Synthesis

Water soluble (Vitamin B complex and C) and fat soluble (Vitamin A and E) vitamins are considered essential for all insect species (Barbehenn et al., 1999). Vitamin B acts as co-enzymes with pyridoxine (Vitamin B₆), folate (Vitamin B₉), riboflavin and FAD (Vitamin B₂), biotin (Vitamin B₇), thiamine (Vitamin B₁), nicotinic acid (Vitamin B₃), cobalamin (Vitamin B₁₂) and pantothenate (Vitamin B₅) essential to insects (Barbehenn et al., 1999). As noted with
plant degrading enzymes, profiles of genes involved in vitamin metabolism were similar for both species, although *P. striatum* typically had a higher diversity of genes in low abundances.

A number of genes involved in the metabolism of the essential B vitamins, with the exception of pyridoxine, were detected in the *Pachysoma* gut metagenome of both species. This suggests that microorganisms are, at least in part, responsible for the metabolism of the vitamin B complex. However, only a few genes involved in the pyridoxine metabolism pathway were detected (Figure 4.12), which could be due to the insect being able to synthesize vitamin B6 itself. The absence of such genes is unexpected, with microbial genes for synthesis of pyridoxine detected from the microbiomes of other plant feeding insect species (Scully et al., 2013).

**Figure 4.12:** Adapted KEGG pyridoxine (vitamin B6) metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes.
Genes involved in the metabolism of other vitamins were also detected. Vitamin C (L-ascorbic acid) metabolism pathways were not complete in *Pachysoma*. Vitamin C is only required by herbivorous insects (Barbehenn et al., 1999), presumably those feeding on live plants. However, the detection of several genes involved in the synthesis of L-Ascorbate-6P (Figure 4.13) suggests that microbial communities may partly assist in vitamin C synthesis. Vitamin C is readily available in plant substrates and therefore would not be expected to be synthesized by the insect. Similarly, vitamin E is not required by all insects (Barbehenn et al., 1999), which is confirmed by the absence of genes involved in α-tocopherol (vitamin E) synthesis pathways suggesting that if this vitamin is essential to *Pachysoma*, it is derived from the diet or synthesized by the insect. Another suggestion is that vitamin K$_1$, of which synthesis pathways were present in *Pachysoma*, may be substituted for vitamin E (Barbehenn et al., 1999).

Figure 4.13: Adapted KEGG Ascorbate (Vitamin C) and Aldarate metabolism reference pathway (http://www.genome.jp) depicting which enzymes were present in the *P. endroedyi* (green), *P. striatum* (brown) and both (yellow) metagenomes.
Animals are unable to synthesise sterols, which are essential for insect nutrition. Therefore, the majority of insects derive sterols from their diet. This appears to be the case in *Pachysoma* with no synthesis pathways for the fungal-derived sterol ergesterol or synthesis of the necessary insect sterols 7-dehydrocholesterol and cholesterol (Douglas, 2009). These results further highlight the similarities in the functional capacities of the two *Pachysoma* microbiomes, once again suggesting that individuals of the two species may be able to feed on both diets.

### 4.3 Conclusion

This is the first shotgun metagenomic study on a dry detritus/dung feeding insect. Here it has been possible to compare the gut microbial community structure, and functional capacity thereof, of two *Pachysoma* species feeding on different diets. The bacterial diversity was shown to be greater than previously expected in *P. endroedyi* gut samples. However, large differences in bacterial diversity were noted, as found for the amplicon study (Chapter 3; Franzini et al. (2016)). In comparison, the functional capacity of both *Pachysoma* microbiomes was similar, with the exception of methane synthesis. Therefore, it would appear that *Pachysoma* species may have a core functional microbiome and that individuals of each *Pachysoma* species may be able to feed on and successfully utilize the other diet (i.e. dung feeding species feed on plant detritus and vice versa) when needed. This corresponds with *P. endroedyi* originally being classified as polyphagous through feeding observations (Harrison et al., 2003). Furthermore, there is evidence that novel GH enzymes, particularly cellulases and hemicellulases, may be present in *Pachysoma* metagenomes and this should be determined further by specialized studies.

### 4.4 References


Chapter 5: Concluding Remarks

5.1 Studying the *Pachysoma* gut microbiome

Insect-microbiota relationships have been studied for several years. Such studies have increased knowledge in how microorganisms interact with the host and how these microorganisms may be used for industrial purposes (e.g. biofuel research (Sun and Scharf, 2010)), pest control (Douglas, 2007) and other agricultural activities (e.g. understanding and improving honeybee health (Evans and Schwarz, 2011)) and understanding and treating disease (e.g. malaria (Dong et al., 2009)). However, the majority of gut microbiome studies focus on only a few specific insect taxa including termites (Brauman, 2000, Brune and Friedrich, 2000, Ohkuma, 2003, Brune, 2013), mosquitoes (Wang et al., 2011, Boissière et al., 2012, Osei-Poku et al., 2012, Wang et al., 2012, Coffey et al., 2014), flies (Broderick and Lemaitre, 2012, Wong et al., 2013, Wang et al., 2014, Wong et al., 2015) and honey bees (Ahn et al., 2012, Engel et al., 2012, Sabree et al., 2012). However, few studies have investigated the gut microbiota of the ecologically important adult dung beetles. These beetles are essential for nutrient cycling and removal of decaying dung which may decrease pest and disease prevalence (Nichols et al., 2008).

The lack of studies on the gut microbiota of dung beetles may be attributed to the continued emphasis on insects of medical, economic and agricultural import. A majority of studies also focus on insects feeding primarily on lignocellulosic materials as enzymes produced by lignocellulose degrading microorganisms may be useful in industry. As dung beetles typically feed on wet dung they do not fall into the above categories. Although plant fragments are found in dung (Anderson and Coe, 1974), beetles feeding on wet dung typically remove any plant biomass, exhibiting soft saprophagy (Anderson and Coe, 1974, Holter et al., 2002). In contrast, *Pachysoma* species feed on plant detritus or dry dung (Harrison et al., 2003), where plant biomass cannot easily be removed by the insect before consumption,
making them good candidates for studying microbiomes involved in lignocellulosic degradation. Evidence suggests that *Pachysoma* species feeding on both plant detritus and dung have adapted to feed on plant material (Holter and Scholtz, 2011, 2013), with researchers hypothesising that they are able to successfully digest lignocellulosic diets.

Insect host phylogeny has an important impact on gut microbiome assembly (Colman et al., 2012). Therefore, host phylogeny should always be considered when studying the impact of environmental variables (diet, location, etc.) on gut microbiome assembly. However, host phylogeny may have a greater impact on gut microbiome assembly in some insect groups compared to others (Colman et al., 2012). Host phylogeny is expected to have a less pronounced effect on gut microbiome assembly in insect species from the same genus. Therefore, by studying insect genera with species feeding on different diets, such as *Pachysoma*, we may be able to better understand the effect of host diet on gut microbial communities where the influence of host phylogeny is reduced.

In order to research the relationship between *Pachysoma* species and their respective gut microbiota, this study aimed to determine the taxonomical diversity and functional capacity of the *Pachysoma* gut microbiome. This research acted as the primary study to determine the basic characteristics and potential of the associated microorganisms. This was achieved by comparing the gut microbiomes of plant detritus (*P. endroedyi*) and dry dung (*P. striatum*) feeding insects to answer the following three main research questions:

1. What is the gut microbial diversity of the two *Pachysoma* species and how does it differ between the two species?
2. What is the functional capacity of the *Pachysoma* gut microbiome to assist the host with digestion of the respective food-source?
3. Is host diet or phylogeny the main driver of microbiome assembly?

Next-generation sequencing provided answers to the first two research questions, the implications of which will be discussed further below. However, this study was unable to
conclusively answer the third research question which is discussed further as a limitation (Section 5.2). Suggestions for future studies to better understand the role of host diet and phylogeny in *Pachysoma* gut microbiome assembly will be discussed in Section 5.3.

Both the amplicon and shotgun sequencing studies showed distinct differences in gut bacterial diversity between the two *Pachysoma* species with the amplicon study indicating a small core microbiome. However, the functional capacities of the two *Pachysoma* gut microbial communities were similar. This is in line with the suggestion that a functional core microbiome may be present in place of a phylogenetic core (Turnbaugh et al., 2009, Karasov et al., 2011). Furthermore, *Pachysoma* species and individuals appear to associate with different microbial communities which assist the host in a similar fashion nutritionally. Furthermore, *P. striatum* does appear to feed on plant biomass found within the dung, unlike dung beetles feeding on wet faeces. All the above would allow the different *Pachysoma* species to feed on both diet types in the situation where the respective specialised diet is not available. For desert dwellers, such an adaption would be invaluable. Although competition for specific food-sources is reduced in desert ecosystems (Wharton, 2002), the specialised diet may be unavailable. Such observations are in line with previous reports of polyphagous feeding behaviour in *P. endroedyi* (Harrison et al., 2003). Since then, unpublished data has suggested a specialised feeding behaviour in *P. endroedyi* (C. Scholtz Pers. Comm.). However, this study may suggest that previous observations were in fact indicating the ability of individuals to feed on both food-sources when necessary.

### 5.2 Limitations of this study

The insect gut is a complex environment where numerous factors affect microbiome assembly (Engel and Moran, 2013, Douglas, 2015). Although certain variables were controlled during this study (location, season and life-stage), other factors are likely to have an impact on the respective gut microbiota, although the intensity of such is unknown. In particular host diet and phylogeny are both probable role players in microbiome assembly (Colman et al.,
2012, Yun et al., 2014), even though the effect of host phylogeny is expected to have a lesser impact than across higher insect taxa. In this study, these variables are intertwined, with each species feeding on a different diet. Although some evidence was found to suggest that diet is the main role player, further studies would need to confirm these findings (Section 5.3).

A large variety of studies have focused on the insect gut microbiome. However, the majority of studies have been dedicated to particular insect taxa. This specialised focus may be of detriment to research on understudied insect groups. This is of particular note in this study where no literature was found on *Pachysoma* gut microbiomes or specialised dung feeding insects, particularly adult scarab beetles. This means that comparisons need to be made with insects in different families/orders feeding on different substrates to that of *Pachysoma*. Due to the large number of factors causing variations in gut microbiome assembly (Engel and Moran, 2013, Douglas, 2015), comparisons across insect groups is very speculative and therefore was used with caution throughout this study. Furthermore, studies on gut microbiomes of scarab beetles typically focus on larvae (Egert et al., 2003, Lemke et al., 2003, Egert et al., 2005). Although, there is evidence that both *Pachysoma* species feed on plant material within their specific diets, only studies on adult insects were compared with as the microbiomes of larvae have been found to differ considerably from adults in holometabolous insects (Engel and Moran, 2013). This study will hopefully serve as a basis for future studies considering the gut microbiomes of *Pachysoma* species, dry dung feeders, scarab beetles and future diet-microbiome related studies.

A further limiting factor, which affects all insect-microbiome studies, is linked to methodological constraints. The complexity of mDNA extracted from insect guts complicates downstream analyses. Traditional mDNA extraction methods typically yield high concentrations of host DNA (Carpi et al., 2011, Feehery et al., 2013, Jia et al., 2013). In order to overcome this challenge, target-specific primers are commonly used. This has been successful for amplifying the bacterial 16S rRNA gene in insect-microbiome studies as
described in Chapter 3 (Franzini et al., 2016). However, as insects and fungi are closely related (Gibson and Hunter, 2010), primers targeting the fungal ITS region and 18S rRNA gene are typically non-specific, amplifying insect DNA. For this reason the majority of insect-fungal studies utilise culture-dependent methods (Gibson and Hunter, 2010). In this study, primers targeting the fungal ITS region with minimal insect coverage were used (Ihrmark et al., 2012). However, amplification of the ITS region was not successful for the plant detritus feeding *P. endroedyi* (Chapter 3; Franzini et al. (2016)). In comparison, shotgun sequencing of *Pachysoma* gut samples, where host DNA contamination was reduced, detected fungal, archaeal, bacterial and viral communities from both *Pachysoma* species. At current, few studies have utilised shotgun sequencing to study insect gut microbiomes (e.g. He et al. (2013), Jia et al. (2013), Scully et al. (2013), Shi et al. (2013)). When studying the overall microbial community structure in a broad context as in this study, shotgun sequencing is a valuable tool for uncovering phylogenetic and functional information in the absence of bias associated with amplicon sequencing (Hodkinson and Grice, 2015). However, methods which reduce host DNA contamination may skew eukaryotic microbial (i.e. fungi) results (Chapter 4). Therefore, studies focusing on fungal or other eukaryotic microorganisms would do better to utilise alternative methods.

### 5.3 Future study prospects

Certain suggestions can be made regarding sampling of *Pachysoma* individuals for future studies. Firstly, it is suggested that the individual beetles be collected while foraging or feeding so that the exact food-source can be collected with the respective beetle. This would allow for a parallel analysis of the microorganisms associated with both the exact food-source and the insect gut. Secondly, a large number of samples should be collected and analysed due to the substantial intra-specific variation recorded in this study (Chapter 3; Franzini et al. (2016)).
In order to truly understand the effect diet has on *Pachysoma* gut microbiome assembly it is necessary to know the chemical and material composition of the food on which the individual insects feed. The composition and nutritional quality of wet dung fed on by scarab beetles is fairly well understood (Holter, 2016). However, less information is available on the nutritional composition of dry dung, particularly faeces fed on by adult *Pachysoma* species. In depth detail would need to be known, extending to the plant species which make up the plant detritus and are found in the dung. As the chemical and material composition of plant polymers differs according to plant species (Hättenschwiler et al., 2008), it is important to identify the composition of the respective *Pachysoma* food-source. Some studies have observed the feeding behaviour of the coprophagous *P. striatum*, which appear to feed on rabbit dung (Scholtz, 1989). However, newer observations of several *Pachysoma* species should be made. Furthermore, it is believed that *P. striatum* likely feed on dung from the sheep grazing on the field where the samples are collected (C. Scholtz Pers. Comm.).

Lab-based studies have shown great promise in determining the effect of diet on gut microbiome assembly. However, it is not possible to lab-rear all insect species. *Pachysoma* species are particularly difficult to study in a laboratory environment (C. Scholtz Pers. Comm., (Harrison et al., 2003)). One such possible reason for this is their need to burrow before feeding (Scholtz, 1989, Harrison et al., 2003, Holter et al., 2009). As their burrows are quite deep (Scholtz, 1989), it would be ineffective to keep the insect in cages where variables such as environmental conditions can be controlled. However, attempts to lab-rear *Pachysoma hippocrates* noted the ability of the beetles to forage and burrow successfully although they did not breed (Harrison et al., 2003). However, it may be possible to answer the necessary research questions using specialised field-based studies. The suggested alternative would be to use shotgun sequencing to analyse the gut metagenomes of all 13 *Pachysoma* species. Of the 13 *Pachysoma* species, four are specialised plant detritus feeders, six are specialised dung feeders, one is polyphagous and the dietary specialisation of the remaining two is
unknown (C. Scholtz Pers. Comm., (Harrison et al., 2003)). Furthermore, although all 13 species are found in South Africa or Namibia (Harrison et al., 2003), species are collected from different sample sites (C. Scholtz Pers. Comm.). Therefore, by studying all 13 species, cluster analysis can be used to determine the impact of host diet, phylogeny and location. This should also give insight into the functional capacity of *Pachysoma* and the hypothesis presented in this study that the different species are functionally equipped to feed on both substrates.

Another suggested line of study would be comparing the gut microbiomes of other dry dung feeding scarab beetles, particularly focusing on genera where species feed on different materials, as in *Pachysoma*. One suggested taxa is the South American *Anomiopsoides* genera, which have been found to also include species feeding on dry dung or plant material (Ocampo, 2005). By comparing the gut microbiomes of these species with those of *Pachysoma* species it would be possible to determine differences or similarities between host feeding preferences, phylogeny and/or locality across insect taxa of the same family. However, caution would need to be taken in sampling the food-source along with the beetle and determining the composition of each individual food-source.

This study has succeeded in showing the potential that future studies in *Pachysoma* gut microbiomes would have in further understanding the relationship of host diet and microbiome assembly. There is a lot of research potential in better understanding the ecological impact of the gut microbiomes of these unusual dung beetles.

### 5.4 References


